Nancy-Université INPL

INSTITUT NATIONAL POLYTECHNIQUE DE LORRAINE

ECOLE DOCTORALE : *Ressources Procédés, Produits et Environnement*

Laboratoire Lorrain de Recherche en Informatique et ses applications

Laboratoire Agronomie et Environnement

THÈSE

Présentée et soutenue publiquement le 12/11/2010
pour l'obtention du grade de Docteur de l'INPL
(Spécialité : Sciences agronomiques)

par

Zainab ASSAGHIR

Analyse formelle de concepts et fusion d'informations : application à l'estimation et au contrôle d'incertitude des indicateurs agri-environnementaux

Directeur de thèse :	Amedeo Napoli	Directeur de Recherche, LORIA, Nancy
Co-directeur de thèse :	Philippe Girardin	Directeur de Recherche, INRA, Colmar
Composition du jury :		
Président du jury :	Thierry Denoeux	Professeur, UTC, Compiègne
Rapporteurs :	Didier Dubois	Directeur de Recherche, IRIT, Toulouse
	David Makowski	Chargé de Recherche, Inra Grignon
Examinateurs :	Florence Le Ber	ICPEF, ENGEES, Strasbourg
	Sylvain Plantureux	Professeur, ENSAIA, Nancy

Remerciements

Je tiens à saluer ici les personnes qui, de près ou de loin, ont contribué à la concrétisation de ce travail de thèse de doctorat.

Tout d'abord, je tiens a remercier mes deux directeurs Amedeo Napoli et Philippe Girardin. En particulier, Amedeo qui a su me laisser la liberté nécessaire à l'accomplissement de mes travaux tout en gardant un œil critique. Nos continuelles oppositions, contradictions et confrontations ont sûrement été la clé de notre travail commun. Plus qu'un encadrant ou un collègue, je crois avoir trouvé en lui un ami qui m'a aidée aussi bien dans le travail que dans la vie lorsque j'en avais besoin.

Je tiens à exprimer ma profonde gratitude à Thierry Denoeux, qui m'a fait l'honneur de présider le jury de thèse de doctorat, pour l'intérêt et le soutien chaleureux dont il a toujours fait preuve.

Je remercie Didier Dubois de m'avoir fait l'honneur de rapporter ma thèse. J'éprouve un profond respect pour son travail et son parcours, ainsi que pour ses qualités humaines. Le regard critique, juste et avisé qu'il a porté sur mes travaux ne peut que m'encourager à être encore plus perspicace et engagée dans mes recherches. Ses commentaires et ses questions, tant sur la forme du mémoire que sur son fond, ont contribué à améliorer de manière significative le document.

Je remercie aussi David Makowski d'avoir accepté de rapporter cette thèse ainsi que d'être membre de son comité de pilotage. Merci à Florence Le Ber et Sylvain Plantureux d'avoir accepté de faire partie de mon jury de thèse.

Un grand merci à Henri Prade dont la persévérance et la rigueur m'ont permis de mener à bien ce travail. Je te remercie pour tout ce que tu m'as apportée, pour ton écoute et tes conseils. Tes compétences scientifiques et ta rigueur m'ont été d'un profit inestimable. Sois assuré de mes meilleurs sentiments.

Je remercie Serge Guillaume pour ses encouragements et pour les discussions que nous avons eues pendant mes travaux de thèse, lors de rencontres scientifiques, à un moment critique où je manquais de repères. De même, je remercie Sébastien Destercke pour ses conseils techniques et pratiques, ainsi que Brigitte Charnomordic et Olivier Strauss qui n'ont pas manqué de me donner des conseils.

Un merci particulier à Kamel Smaili, pour son amitié, ses encouragements, ses conseils ainsi que les discussions sur de nombreux sujets allant des sous-ensembles flous au traitement des langues. Ta patience et ta gentillesse sont inaltérables. Te côtoyer fut un plaisir.

Je remercie les membres de l'équipe Agriculture Durable : Nathalie, Chantal, Béatrice, Christophe, Bernard, Christian Rousset, M. Chapot et M. Delphin, pour l'acceuil ainsi que pour le soutien et l'encouragement tout au long de ces quatre ans. Je remercie également les membres de l'équipe Orpailleur du Loria : Malika, Marie-Dominique, Dave, Emmanuelle et Yannick pour l'acceuil, les encouragements et les conseils.

Merci à ceux qui ont été tour à tour mes collègues de bureau, pour leur accueil et pour les nombreuses discussions que nous avons eues.

Merci aux quelques amis que j'ai eus la chance d'avoir à mes côtés : Yesmine, Florent, Julien, Lazaros et Vishwesh.

Merci à Nada et Nizar, je vous fait part de mon plaisir à vous avoir rencontré.

Merci à Anisah pour l'encouragement quotidien et la compagnie des soirs et des week end au laboratoire.

Merci à Julie, pour l'aide précieuse et les avis que tu m'as apportées aux moments importants.

Un merci à Chedy pour le soutien, l'aide, et les pages de remarques qui ont bien amélioré la présentation de mon travail, ainsi les belles photos qu'il a prises le jour de ma soutenance. Ta sympathique compagnie depuis ton arrivée au laboratoire fut un réel plaisir.

Je tiens à remercier Mehdi pour l'intérêt qu'il a montré pour mon travail, pour l'aide, le soutien et pour tous les moments passés ensemble au travail et en dehors.

Un Merci chaleureux à Jean pour l'aide, le soutien, l'encouragement, les nombreuses relectures de ce document et pour les moments inoubliables d'humour au travail et en dehors. Sois assuré de ma sincère amitié.

Merci à Heba, Rola, Hassan, Jamal et Tarek qui étaient tous les jours à côté et ont aussi tout organisé pour le pot de ma soutenance.

Enfin, tous mes remerciements à ma famille pour leur confiance et leur soutien et un merci spécial du fond du coeur à Georges qui a toujours été présent lorsque j'en ai eu besoin et à qui je dédie cette thèse.

Table des matières

Table des figures

Liste des tableaux

Introduction

1 Contexte de travail

Dans le domaine de l'analyse de risques environnementaux, les chercheurs de l'INRA de Colmar mettent en place des modèles agronomiques, appelés indicateurs agri-environnementaux, afin de calculer les risques des pratiques agricoles utilisées par l'agriculteur aux champs. Ces indicateurs sont dédiés à l'aide à la décision et ils sont utilisés pour lutter contre la pollution de l'environnement. A titre d'exemple, il existe un indicateur "*pesticide*" qui évalue le risque de l'utilisation de pesticides sur l'environnement et qui permet d'établir un diagnostic sur le choix des pratiques agricoles et en particulier le pesticide utilisé. Ces indicateurs sont calculés à partir de plusieurs variables dont une partie provient de sources d'informations hétérogènes car de nature bien différentes : expert du domaine (agronome), base de données, manuels d'agronomie, fiches techniques...

Pour calculer la valeur d'un indicateur, les experts doivent choisir une valeur parmi les valeurs proposées par les sources et ensuite faire le calcul. Le résultat de calcul de l'indicateur peut être considéré comme entaché d'imperfections car les informations fournies par les sources sont souvent de nature imparfaite, par exemple à cause de l'imprécision ou de manque de fiabilité des données disponibles. C'est la que réside le problème à l'origine de cette thèse : comment contrôler la propagation de l'imprécision dans le calcul d'un indicateur à partir de données imparfaites. Le problème recouvre beaucoup plus d'aspects qu'il n'y paraît, allant de la représentation des données par des possibilités à l'analyse formelle de concepts en passant par la fusion d'informations.

Plusieurs problématiques sont liées au traitement de l'imperfection : l'inférence à partir des données imparfaites, la fusion d'informations et la prise de décision. L'inférence consiste à tirer des conclusions à partir de variables d'entrée grâce à un modèle, et concerne par exemple le problème de la propagation de l'incertitude à travers un modèle déterministe ou stochastique. L'inférence est monotone dans le cas des modèles déterministes. Cela signifie que plus l'incertitude est petite sur les variables d'entrée, plus elle l'est sur les variables de sortie du modèle. Mais ce n'est pas forcément le cas avec les modèles stochastiques pour lesquels une information précise sur les variables d'entrée ne préjuge pas de la grandeur de l'incertitude sur les variables de sortie du modèle. La fusion d'informations, quant à elle, consiste à résumer un ensemble d'éléments d'informations fournies par plusieurs sources souvent hétérogènes en une information exploitable et facile à interpréter, tout en tenant compte de l'incohérence et des conflits entre les informations à fusionner ainsi que des relations entre les sources d'informations. Enfin, la prise de décision permet de déterminer l'action optimale à entreprendre dans une situation donnée. La décision est un acte qui a un impact sur le monde une fois la décision prise.

Dans ce travail, nous nous intéressons à la prise en compte des imperfections dans le calcul des indicateurs agri-environnementaux, en particulier lorsque le calcul dépend de variables dont la valeur est fournie par plusieurs sources d'informations non nécessairement cohérentes entre elles. Dans la littérature, plusieurs approches ont été proposées pour représenter et manipuler

des données imparfaites. Chaque approche possède ses avantages et ses limites. Le contexte de l'étude, la nature de l'information et l'objectif attendu jouent un rôle important dans le choix d'une approche adaptée à la prise en compte des données imparfaites.

L'imperfection qui est considérée dans cette thèse est l'imprécision : les données sont imprécises car elles sont insuffisante pour permettre de comprendre une situation donnée.

Il existe différentes méthodes de fusion d'informations qui dépendent du formalisme de représentation des données imparfaites choisi et des méta-information sur les sources (fiabilité, indépendance). Ainsi, la fusion d'informations dans le cadre de la théorie des possibilités s'appuie sur des opérateurs de fusion qui s'appliquent à l'ensemble de toutes les sources et qui calcule un résultat unique et global. Tout le problème est donc de pouvoir se servir de résultat pour prendre effectivement une décision.

Dans notre contexte qui porte sur l'analyse de risque environnemental, les sources d'informations fournissent des informations numériques (valeurs et intervalles de valeurs) et le risque– c'est-à-dire la variable de sortie qu'il faut évaluer– est une moyenne arithmétique dépendant des sources d'informations en entrée. Ici, la théorie des possibilités, qui a été retenue pour ce travail, offre des outils bien adaptés pour représenter et manipuler ces informations, et notamment une certaine variété des opérateurs de fusion d'informations imparfaites.

2 Principales contributions

Dans ce mémoire de thèse, nous avons essayé de traiter le problème d'aide à la décision et le calcul d'indicateurs en présence d'informations imparfaites. Pour cela, nous avons considéré ce problème comme un problème de fusion d'informations imparfaites et nous avons mis en oeuvre un cadre de calcul d'indicateurs qui fait appel à la fois à la théorie des possibilités et l'Analyse Formelle de Concepts (AFC). La théorie des possibilités permet de prendre en compte et contrôler l'imperfection des données dans le calcul. L'Analyse formelle de concepts et son extension les structures de patrons fournissent un cadre global pour la fusion d'informations. L'AFC permet de construire un treillis de concepts formels à partir d'une table binaire (les attributs peuvent être à valeurs dans des intervalles pour les structures de patrons). Un concept se présente sous la forme d'un couple composé d'une intension et d'une extension. L'intension est décrite par un ensemble d'attributs et fournit la description maximale commune de l'ensemble des objets présents dans l'extension du concept. Donc, nous utilisons la théorie des possibilités pour représenter les informations, ensuite nous utilisons l'AFC pour combiner ces informations. Nous montrons comment un opérateur de fusion peut être considéré comme un *infimum* dans un demi-treillis lorsqu'il vérifie les propriétés de commutativité, d'associativité et d'idempotence. Nous considérons ainsi le cas des opérateurs non-associatifs et présentons une méthode de prétraitement des données permettant d'appliquer malgré tout de tels opérateurs de fusion puis de construire le treillis de concepts.

Le treillis de concepts obtenu fournit une classification des informations, des sources et des résultats de la fusion pour un opérateur de fusion donné. Les extensions représentent des sous-ensembles maximaux de sources. Les intensions correspondent aux résultats de la fusion des valeurs fournies par les sources de l'extension. De plus, le treillis fournit une base de choix pour l'utilisateur. Il présente à la fois le résultat global (c-à-d le résultat de la fusion de l'ensemble de toutes les sources) et les résultats partiels (c-à-d les résultats de la fusion d'un sous-ensemble de sources) de la fusion. Il permet d'identifier des sous-ensembles maximaux de sources avec leurs résultats de fusion. Il classifie les informations, sous la forme d'une hiérarchie de concepts, suivant leur précision et leur fiabilité. De plus, il montre les liens entre les informations et les

sources.

Les résultats obtenus à partir du treillis sont utilisés pour calculer des indicateurs environnementaux qui jouent un rôle dans l'aide à la décision environnementale. En effet, les résultats obtenus servent à évaluer les pratiques agricoles des agriculteurs. La prise en compte de l'imperfection des indicateurs permet aux experts de donner plus de flexibilité aux agriculteurs dans le choix de leurs pratiques. Les experts sont plus proches de la réalité (de terrain) et peuvent donner des conseils plus précis en fonction de certaine sources d'informations, sachant que les sources d'informations ne sont pas cohérentes entre elles à cause de la diversité des pratiques dans le monde entier.

3 Organisation du manuscrit

Le chapitre 1 rappelle les méthodes de traitement des données imparfaites existant dans la littérature. Nous nous concentrons sur la théorie des possibilités qui permet de prendre en compte l'imprécision qui est une forme de l'imperfection des données. Nous présentons les notions de base permettant de résoudre le problème du traitement de l'incertitude dans le contexte environnemental. Ainsi, nous nous intéressons au problème de la fusion des informations imparfaites dans le cadre de la théorie des possibilités et présentons les opérateurs de fusion généralement utilisés avec leurs propriétés.

Le chapitre 2 présente la construction de treillis de concepts à partir de données binaires dans le cadre de l'analyse formelle de concepts (AFC). Ce chapitre contient un état de l'art sur les extensions de l'AFC aux données numériques et symboliques ainsi qu'aux contextes flous et incertains. Nous détaillons *les structures de patrons* permettant de construire un treillis directement à partir de données numériques. Enfin, nous présentons des relations existant entre les treillis de concepts et la fusion d'informations. L'essentiel des résultats de cette partie a fait l'objet des articles [11, 7, 116, 6, 119, 122].

Le chapitre 3 présente la première partie de la contribution : la fusion d'informations par treillis de concepts à partir des informations représentées par des intervalles. Dans ce chapitre, nous proposons d'introduire les opérateurs de fusion pour construire le treillis de concepts et de considérer les concepts comme un choix donné à l'utilisateur, un concept contenant un ensemble de sources avec le résultat de fusion qui lui est associé. Nous utilisons les opérateurs conjonctifs, disjonctifs et de "compromis" (fondé sur les sous-ensembles maximaux cohérents (SMC)). Nous proposons également un pré-traitement des données à effectuer pour les opérateurs de fusion ne vérifiant pas les conditions d'un opérateur infimum d'un demi-treillis (essentiellement la non-associativité). Les résultats de cette partie ont fait l'objet des articles [4, 8, 10, 9].

La taille des treillis de résultats de la fusion disjonctive est trop importante lorsque nous traitons des données incohérentes. Nous proposons alors une méthode pour réduire le nombre de concepts dans le treillis. Cette méthode est fondée sur le choix d'une distance entre les valeurs et d'un seuil de similarité. Ce dernier permet l'enveloppe convexe des résultats de la fusion des sources lorsqu'ils sont similaires. Cette approche permet de ne conserver que les concepts "utiles" ayant des valeurs similaires et exploitables pour l'utilisateur. L'essentiel des résultats de cette partie ont fait l'objet des articles [118, 117].

Dans le cadre de la fusion des informations, nous généralisons dans le chapitre 4 l'approche proposée dans le chapitre précédent en représentant l'information par des distributions de possibilités, en particulier par des intervalles flous triangulaires et trapézoïdaux. Nous détaillons les caractéristiques d'un opérateur de fusion permettant de construire un treillis dans le cadre des distributions de possibilités. Le treillis résultant fournit une classification des informations, de

leurs résultats de fusion ainsi qu'un intervalle de degré de confiance. Nous proposons de même une méthode de pré-traitement des données à effectuer lorsque l'opérateur ne permet pas de construire le treillis directement. Les résultats de cette partie ont fait l'objet des articles [9, 12].

Le chapitre 5 présente une méthode d'évaluation environnementale utilisée par les experts pour valider les pratiques agricoles utilisées par les agriculteurs aux champs. Cette méthode s'appuie sur la définition des *indicateurs agri-environnementaux*. Nous nous concentrons sur l'indicateur "pesticide" et ses modules. Nous abordons la validation des méthodes présentées dans le chapitre 3 en présentant une application concernant un risque environnemental lié à l'utilisation des pesticides. Le calcul des indicateurs à partir des données imparfaites a fait l'objet de plusieurs articles [5, 4, 8, 118, 117, 12, 119].

Le lecteur pourra se reporter à l'annexe A pour les notions de la théorie de possibilités. L'annexe C est consacrée aux rappels sur les relations d'ordre, la définition d'un treillis et la connexion de Galois. Pour comprendre les méthodes de propagation de l'incertitude utilisées dans ce travail, le lecteur se reporte à l'annexe B. Dans cette annexe, nous rappelons les méthodes classiques d'analyse d'intervalles ainsi que le calcul de la moyenne floue pondérée ayant des intervalles en entrée. Enfin, l'annexe D représente les résultats d'un exemple de calcul d'indicateurs pour plusieurs matières actives afin de calculer un indicateur pesticide pour un programme de traitement.

1

Représentation de l'imperfection

Sommaire

Ce chapitre présente quelques travaux du domaine de la représentation et de la fusion des informations imparfaites dans sa première partie. Dans la deuxième partie de ce chapitre, un rappel de l'analyse d'intervalles est présenté ainsi que le problème du calcul de la variation de la moyenne pondérée floue suivant la variation de ses composants. Nous utiliserons l'analyse d'intervalles pour propager les informations à travers des modèles mathématiques dans le chapitre 5, en particulier la fonction de la moyenne floue pondérée et sa variation en fonction de la variation de ses arguments puisque l'indicateur phytosanitaire est calculé à partir d'une moyenne pondérée.

1.1 Introduction

La formalisation de l'incertain est apparue au XVII^{e} siècle avec les travaux de Huygens, Pascal et J. Bernoulli. Ces travaux traitaient les problèmes des jeux et en déduisaient une théorie de la chance, mais ne parlaient jamais de la notion de probabilité. Bernoulli propose dès 1680 une théorie mathématique des probabilités et de leurs combinaisons. En particulier, les probabilités n'étaient pas toujours additives. Dans une partie de son travail, Bernoulli a proposé la loi des grands nombres qui permet d'estimer *a posteriori* une probabilité inconnue *a priori* à partir de l'observation de fréquences d'occurrence. Plus tard, la théorie proposée par Bernoulli a été simplifiée dans les travaux de Moivre en 1711 où l'on trouve la première règle d'additivité et une

représentation des probabilités entre 0 et 1. Au XVIII^{e} siècle, Lambert, poursuivant les travaux de Bernoulli, a traité de la théorie des erreurs et a proposé une méthode connue sous le nom de "Maximum de vraisemblance". Plus tard, Bayes a calculé la probabilité connaissant le nombre de succès dans un échantillon et l'a exprimée en terme de probabilité initiale (*a priori*), finale (après l'expérience) et vraisemblance (probabilité de l'expérience sachant la probabilité initiale) alors que Bernoulli estime le nombre de succès à partir de la connaissance de probabilité. Les travaux de Bayes ont été repris par Laplace qui a proposé le principe d'équiprobabilité si l'on a aucune raison de penser autrement.

Le XIX^{e} siècle a vu le développement de la loi gaussienne en utilisant le maximum de vraisemblance dans le problème de l'estimation des erreurs d'observations. Mais la notion de probabilité est restée liée à la notion de la fréquence jusqu'au XX^{e} siècle avant les travaux de [163, 88]. Ces derniers ont cherché à justifier l'axiome d'additivité. Les théories du XIX^{e} siècle permettent seulement de résoudre les problèmes liés aux probabilités physiques (ou objectives).

L'émergence de l'informatique, la naissance de l'intelligence artificielle, la représentation des connaissances et le raisonnement entachés d'incertitude, d'imprécision et de contradiction ont conduit les chercheurs à revenir sur la notion subjective de la probabilité à laquelle un grand nombre des chercheurs étaient intéressés du point de vue philosophique par rapport aux probabilités objectives [163, 88]. Ainsi, deux grandes classes de théories ont vu le jour : celle des probabilités additives et celle des probabilités non-additives. La première classe s'appuie sur des postulats pour arriver aux probabilités et à leurs propriétés [53, 112]. La deuxième classe remet l'accent sur la notion d'additivité en s'appuyant sur plusieurs travaux [168]. Par exemple, Dempster généralise les règles de Lambert, qui ne peuvent traiter que des arguments portant sur une seule conclusion, au cas où plusieurs hypothèses sont à envisager. De même, Shackle propose dans le domaine d'économie, des modèles utilisant des notions proches à celles de la théorie des possibilités.

A partir des années 60, de nouvelles théories de l'incertain qui ne sont pas liées aux probabilités sont apparues. Zadeh a inventé la théorie des sous-ensembles flous [188], Shortliffe et Buchanan [37] ont développé le formalisme des "facteurs de certitude" pour un système expert à base de règles à application médicale (*MYCIN*). Shafer a exposé la théorie des fonctions de croyance [167]. Ensuite, Zadeh a introduit la théorie des possibilités, en relation avec la théorie de sous-ensembles flous [189], développée ensuite par Dubois et Prade [74]. Walley a introduit la théorie des probabilités imprécises [174]. Ces nouvelles théories connaissent aujourd'hui un grand développement dans le traitement de l'incertain et en fusion d'informations imparfaites.

1.2 Les imperfections de l'information

Une information est une collection de symboles ou de signes produits soit par l'observation de phénomènes naturels ou artificiels, soit par l'activité cognitive humaine. Une information est destinée à comprendre le monde qui nous entoure, à aider à la prise de décision, ou à communiquer avec des individus [82].

On distingue deux sortes d'informations : objectives et subjectives. Les informations objectives sont issues de l'observation directe de phénomènes, comme les mesures de capteurs, tandis que celles dites subjectives sont les informations exprimées par des individus et conçues sans le recours à l'observation directe du réel.

Une information peut prendre deux formes : numérique et symbolique. Les informations numériques (généralement objectives) peuvent prendre différentes formes : nombres, intervalles de nombres, etc. Les informations symboliques (subjectives) sont exprimées en langage naturel

et souvent codées par des représentations logiques ou graphiques. Néanmoins, une information subjective peut être numérique, et l'information objective peut être symbolique (un capteur qui fournit des couleurs).

On distingue ainsi deux types d'information : l'information contingente et l'information générique. L'information contingente concerne une situation particulière, la réponse à une question sur l'état courant du monde, comme par exemple une observation, ou un témoignage. L'information générique se réfère à une classe de situations, par exemple un modèle statistique issu d'un ensemble représentatif d'observations. Cette distinction est importante dès qu'on aborde les problèmes d'inférence ou de révision d'informations incertaines. De plus, une problématique comme l'apprentissage ou l'induction s'intéresse à l'élaboration de connaissances génériques à partir d'informations contingentes. Inversement la prédiction peut être vue comme l'utilisation d'une connaissance générique sur la fréquence d'un événement pour estimer un degré de confiance en l'occurrence contingente de cet événement dans une situation particulière.

Les données disponibles pour un système d'informations sont souvent imparfaites. Une information est parfaite si elle est précise et certaine. L'imperfection est due à l'incertitude, l'imprécision et l'inconsistance. Ces dernières notions représentent les aspects majeurs des données imparfaites [35, 169, 74, 80, 82, 141] qui vont être définis juste après.

Généralement, l'imprécision et l'inconsistance sont liées au contenu de l'information. L'incertitude résulte du manque d'information à propos d'une telle situation, l'inconsistance entre les informations, l'incomplétude de l'information et la variabilité. L'imprécision et l'inconsistance concernent le contenu de l'information tandis que l'incertitude concerne la validité de l'information.

1.2.1 Information imprécise

Une information est dite imprécise si elle est insuffisante pour permettre de comprendre une situation donnée. L'information imprécise est liée à l'idée de l'information incomplète.

En effet, la question à laquelle on cherche à répondre est la suivante : *quelle est la valeur de la variable X ?* ou *est-ce que la valeur v de X satisfait une certaine propriété ?*

Considérons par exemple l'âge d'une personne et le sous-ensemble $S = \{1, 2, \ldots, 100\}$ (en nombre d'années). Le terme *jeune* est imprécis, quand il s'agit de connaître l'âge de la personne. Une information imprécise prend la forme d'une disjonction de valeurs mutuellement exclusives [82]. Par exemple, considérons l'information suivante : l'âge de Georges est entre 20 et 25 ans, ce qui se traduit par l'âge de Georges (noté X) appartient au sous-ensemble $\{20, 21, 22, 23, 24, 25\}$, c-à-d $X = 20$ ou $X = 21$ ou $X = 22$ ou $X = 23$ ou $X = 24$ ou $X = 25$ (car la variable n'a qu'une seule valeur). En logique classique, l'imprécision apparaît comme une disjonction. Affirmer $p \vee q$ c'est dire qu'une des propositions $p \wedge q$, $p \wedge \neg q$, $\neg p \wedge q$ est vraie.

1.2.2 Information incertaine

Une information est dite incertaine si l'on ne sait pas si elle est vraie ou fausse. Une information élémentaire est une proposition ou l'affirmation qu'un événement s'est produit. Elle est modélisée par un sous-ensemble de valeurs possibles avec un marqueur d'incertitude. Ce marqueur peut être numérique ou linguistique. Les marqueurs sont toujours des nombres (probabilité) ou des modalités symboliques (certain, possible, probable). Par exemple : il est certain que Georges arrive en retard ou la probabilité que Georges arrive en retard est de 0.7.

1.2.3 Information inconsistante

L'inconsistance apparaît quand plusieurs informations concernant la même situation sont en conflit. Par exemple : "Georges est célibataire" et "La femme de Georges s'appelle Maria". Le conflit dans les données peut mener à une incohérence dans la conclusion.

1.3 Représentation des informations imparfaites

La théorie des probabilités est la plus ancienne des théories de l'incertain, la mieux développée mathématiquement et la plus établie. Elle était la seule à interpréter la notion de hasard et d'incertitude. Cette mesure d'incertitude peut varier suivant les circonstances ou l'observateur. Un individu peut avoir des difficultés à décrire de manière précise son état de connaissance. Il peut paraître plus naturel de fournir un ensemble des valeurs possibles. Cette caractérisation de l'incertitude non pas par un nombre unique mais par plusieurs valeurs a ouvert la voie à d'autres théories de gestion et de manipulation de l'imperfection durant les trente dernières années [74, 143, 174, 157, 56, 57, 167].

On trouve ainsi une hiérarchie de représentations à base de familles convexes de probabilités, incluant la théorie des probabilités imprécises [40], la théorie de fonctions de croyance de Shafer [167], les ensembles aléatoires [143] les boîtes de probabilités (ou p-boxes), la théorie des possibilités [73, 83] et plus récemment les nuages [150]. Des liens de passage existent entre l'une et l'autre de ces théories. Une étude détaillée de ces théories est explorée par S. Destercke [55], qui offre un panorama des théories de l'incertain en mettant l'accent sur des aspects qui permettent d'unifier ces cadres et de chercher le meilleur cadre à appliquer pour une situation donnée.

1.4 Théorie des probabilités

La notion de probabilité est liée à celle de l'expérience aléatoire. Une expérience est dite aléatoire si l'on ne peut pas prédire avec certitude le résultat. Formellement, une probabilité p est une fonction positive sur un espace U à valeurs dans $[0,1]$, $p : U \rightarrow [0,1]$, telle que $\sum_{\omega \in U} p(\omega) = 1$.

Un sous-ensemble $A \subseteq U$ est appelé événement. On définit la mesure de probabilité $P(A) = \sum_{\omega \in A} p(\omega)$. Cette mesure évalue le degré de réalisation de l'événement A. En considérant deux événements A et B, la mesure de probabilité vérifient les axiomes suivants :

– Axiome d'additivité :

$$\forall A, B \subseteq U \quad , \quad P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{1.1}$$

– Axiome de dualité :

$$\forall A \subseteq U \quad , \quad P(A) = 1 - P(\overline{A}) \tag{1.2}$$

La fonction de répartition est définie à partir de la distribution de densité de probabilités p, et on a $\forall x \in \mathbb{R}, F(x) = \int_{-\infty}^{x} p(x)dx$.

Dans plusieurs domaines, la théorie des probabilités est la plus utilisée surtout quand des distributions et des grands échantillons de données sont disponibles.

En absence de données statistiques, la connaissance probabiliste dont dispose un individu, est représentée par un ensemble de propositions logiques avec leurs probabilités plutôt que par une mesure de probabilité sur un ensemble exhaustif d'états mutuellement disjoints. La première engendre en général une famille de probabilités et non une distribution unique.

La théorie des probabilités semble pauvre pour la manipulation et la représentation de la méconnaissance (l'ignorance partielle ou totale). Elle ne permet pas de représenter d'une manière rigoureuse et propre un état d'ignorance ou de connaissance partielle. Pour illustrer cette limite de la théorie des probabilités, considérons un exemple d'une pièce et un jeu de $Pile$ et $Face$ dans deux situations.

- Le premier cas : la pièce est montrée (elle n'est pas fausse), et la probabilité d'avoir $Pile$ ou $Face$ est égale ($P(Pile) = P(Face) = \frac{1}{2}$).
- Le deuxième cas : la pièce n'est pas montrée (c-à-d elle peut être "$Face - Face$", "$Pile - Pile$", ou "$Pile - Face$"). Partant de la théorie des probabilités, on est obligé de considérer l'équiprobabilité (c-à-d $P(Pile) = P(Face) = \frac{1}{2}$). Toute autre représentation ne serait pas cohérente et même dans un jeu où l'une des parties est favorisée par rapport à l'autre.

Donc, on ne peut pas modéliser de deux façons différentes les deux situations de savoir et de ne pas savoir.

Les limites de la théorie des probabilités apparaissent aussi quand il s'agit de modéliser un phénomène ou un modèle n'ayant qu'un simple échantillon de données, ou dans le cadre de la représentation des informations subjectives (les opinions d'experts). L'approche probabiliste est mal adaptée à la modélisation de l'imprécision des opinions d'experts et limite le choix des méthodes de fusion quand il s'agit de synthétiser une information interprétable à partir d'un ensemble d'informations provenant de plusieurs experts [166].

1.5 Théorie des possibilités

La théorie des possibilités a été proposée par L.Zadeh en 1987 [189] en s'appuyant sur la théorie des sous-ensembles flous qu'il a inventé en 1965 [188]. Cette dernière permet de représenter des classes d'objets dont les critères d'appartenance sont graduels [187]. La théorie des possibilités est ensuite développée par Dubois et Prade [74] pour traiter dans un seul cadre l'incertitude et l'imprécision inhérentes à certaines données.

L'outil de base de la théorie des possibilités est la distribution de possibilités qui est équivalente à un degré d'appartenance (ou fonction d'appartenance) de la théorie de sous-ensembles flous. La distribution de possibilité (notée π) est une fonction d'un ensemble d'événements élémentaires (noté U) à valeurs dans $[0, 1]$.

La hauteur h d'une distribution est la plus grande valeur que la distribution π peut atteindre :

$$h = \max_{u \in U} \pi(u) \tag{1.3}$$

Une distribution de possibilités π est dite normalisée si sa hauteur est égale à 1. Ce qui signifie qu'un des états est totalement possible, et qui se traduit par

$$\max_{u \in U} \pi(u) = 1 \tag{1.4}$$

Etant donnée une valeur $\alpha \in]0, 1]$, on définit les deux coupes de niveaux α (ou α-coupe) stricte et régulière.

L'α-coupe stricte, de la distribution de possibilités π, est définie par :

$$\pi_\alpha = \{u \in U | \pi(u) > \alpha\} \tag{1.5}$$

L'α-coupe régulière de la distribution de possibilités π est définie par

$$\pi_{\bar{\alpha}} = \{u \in U | \pi(u) \geq \alpha\} \tag{1.6}$$

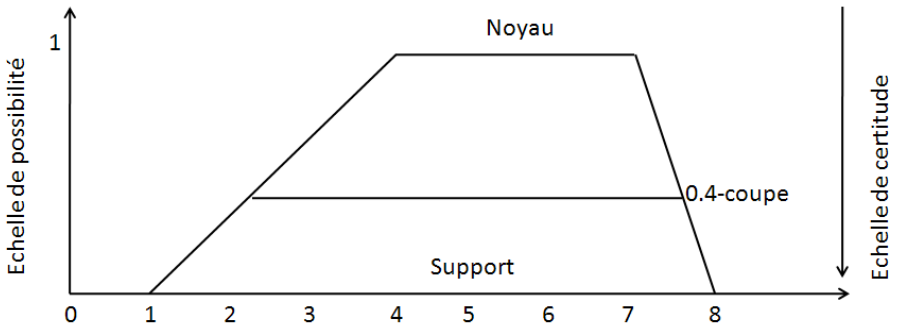


FIGURE 1.1 – La distribution de possibilités construite à partir de l'information d'expert

Le support d'une distribution de possibilités π représente l'α-coupe stricte pour $\alpha \to 0$. Le noyau représente l'α-coupe régulière pour $\alpha = 1$.

Une distribution peut s'interpréter comme un ensemble d'intervalles emboîtés $I_1 \subseteq I_2 \subseteq \ldots \subseteq I_n \subseteq U$ ayant des degré de confiance $\alpha_1 \leq \alpha_2 \leq \ldots \alpha_n \leq 1$.

La théorie des possibilités permet de traiter les imperfections de l'information. En effet, l'imprécision est modélisée par des distributions de possibilité et l'incertitude est mesurée en utilisant les mesures de la théorie des possibilités.

Exemple

Considérons une source fournissant l'information suivante pour la variable X : "*la variable* X *a une valeur entre* 1 *et* 8*, mais les valeurs entre* 4 *et* 6 *semblent être plus vraisemblables*". Cette information peut être représentée en utilisant une distribution de possibilité ayant pour noyau l'intervalle $[4, 6]$ (dans lequel l'expert est certain de trouver la valeur de la variable X) et ayant l'intervalle $[1, 8]$ pour support. Sans autre information, les deux intervalles (support et noyau), sont reliés par une fonction affine définissant une distribution de possibilité trapézoïdale vue comme un emboîtement d'intervalles hiérarchisés par leur degré de possibilité. La distribution de cet exemple est donnée sur la Figure 1.1 avec son support, son noyau, l'α-coupe de niveau 0.4.

Etant donné un ensemble de distributions des possibilités, on dit qu'elles sont conflictuelles (non cohérentes ou incohérentes) si l'intersection de leurs supports est vide. Elles sont à la fois partiellement conflictuelles et partiellement cohérentes si l'intersection de leurs supports n'est pas vide et l'intersection de leurs noyaux est vide. Enfin, elles sont dites cohérentes si leurs noyaux s'intersectent.

1.6 Théorie des possibilités et fusion d'informations

Un des avantages de la théorie des possibilités est le nombre d'opérateurs disponibles pour la fusion des informations imparfaites [1]. Le problème de la fusion est devenu important dans plusieurs domaines où la décision joue un rôle, tel que la fusion de données issues de plusieurs capteurs, le traitement d'images, l'analyse de risques et la fusion de bases de données.

Plusieurs définitions ont été proposées pour la fusion d'informations depuis 1987 [26]. Nous reprenons ici la définition proposée par Bloch et al.[25] :

1. En français, le terme fusion est généralement adopté. En revanche, en anglais, plusieurs termes sont en concurrence dans la littérature, comme par exemple, "merging" [49] et "combining" [76]

La fusion d'informations consiste à combiner des informations hétérogènes issues de plusieurs sources afin d'améliorer la prise de décision.

Depuis des années, des méthodes de fusion (Une méthode, ou un opérateur, ou opération de fusion est le processus utilisé pour combiner les informations fournies par les sources) sont développées et adaptées pour plusieurs applications dans des diverses domaines :

- En robotique, quand il s'agit d'agréger des mesures issues de plusieurs capteurs [1].
- Pour synthétiser les opinions d'un groupe d'experts qui fournissent chacun une estimation de la valeur d'un paramètre mal connu [52].
- Dans les systèmes d'information quand on interroge plusieurs bases de données et qu'il faut fournir une seule réponse à l'utilisateur [48, 50, 51].
- Dans le domaine de l'imagerie médicale [27, 24]

Plusieurs cadres de représentation de données imparfaites ont présenté des opérateurs de fusion, comme, la méthode de consensus utilisée par Cooke [52] pour la fusion des opinions d'experts.

Dans le cadre possibiliste, beaucoup de travaux traitent le problème de la fusion d'informations qu'elles soient numériques ou symboliques [97, 52, 48, 50, 51, 22, 79, 58, 75, 70, 81]. Cependant il n'existe pas de méthode unique pour la fusion même si le cadre de la représentation de l'imperfection est choisi [75]. Dans ce travail, nous nous intéressons au cadre de la théorie des possibilités où les informations fournies par les sources sont modélisées par des distributions de possibilités. Dubois et Prade [75, 81] soulignent l'intérêt de l'utilisation de la théorie de sous-ensembles flous et de la théorie des possibilités pour le problème de la fusion et ont proposé plusieurs méthodes pour la fusion d'informations numériques représentées dans le cadre de la théorie des possibilités.

Plus récemment, un opérateur de fusion fondé sur la notion de sous-ensemble maximal cohérent (SMC) est proposé dans plusieurs cadres de traitement de l'incertain notamment la théorie des possibilités [58, 55]. Cette méthode permet d'obtenir une vue globale lorsqu'aucune information concernant les sources n'est disponible [69, 58, 55].

La notion des sous-ensembles maximaux cohérents utilisée dans la fusion d'informations permet de garder le maximum d'information et ne néglige aucune source(voir Section 1.7.3, plus de détails dans le Chapitre 3 pour la fusion des SMC des informations représentées par des intervalles (voir Section 3.3) et pour les distributions (voir Section 4.3 du Chapitre 4). Elle permet ainsi de traiter le problème d'inconsistance dans les informations.

Tous ces travaux appliquent l'opérateur de fusion sur l'ensemble de toutes les sources et proposent ce résultat à l'analyste.

Une classification des opérateurs de fusion est proposée par Bloch et al. [26]. Dans ce dernier travail, un opérateur de fusion peut suivre un des trois comportements : conjonctif, disjonctif et de compromis. L'opérateur conjonctif utilise des normes triangulaires (t-normes comme le minimum et le produit) et s'applique dans le cas où les sources sont considérées comme fiables. L'opérateur disjonctif utilise des t-conormes (comme le maximum et la somme probabiliste) et correspondent au cas où au moins une source est fiable. Les opérateurs de compromis permettent de passer continûment du mode conjonctif au mode disjonctif.

1.7 Fondamentaux des modes de fusion possibiliste

Un des intérêts des théories des sous-ensembles flous et des possibilités est qu'elles offrent une grande variété d'opérateurs de combinaison. Le choix d'un opérateur peut se faire selon plusieurs critères [26], comme par exemple, le comportement de l'opérateur de fusion. Étant donné N

informations fournies par N sources, chacune représentée par une distribution de possibilités π_i pour $i = 1, \ldots, N$, la fusion d'informations peut suivre trois comportements principaux [26, 75, 81] : conjonctif, disjonctif et de compromis.

1.7.1 Le comportement conjonctif

Le comportement conjonctif est le pendant d'une intersection ensembliste. Le résultat $\phi(\pi_1, \ldots, \pi_N)$ d'un opérateur conjonctif est toujours compris dans tous les éléments d'information fournies par les sources. L'opérateur conjonctif réduit donc l'incertitude globale et fournit un résultat plus précis que chacune des informations provenant des sources prises séparément. Il suppose que toutes les sources sont fiables, et peut fournir un résultat très peu fiable, voire vide, en cas d'inconsistance des informations fournies par les sources. L'opérateur conjonctif s'écrit pour une variable m :

$$\hat{\pi} = \bigwedge_{i=1,\ldots,N} \pi_i$$

où $\bigwedge$ est une norme triangulaire (appelée aussi t-norme)[2]. Ce qui se traduit dans le cas où chacune des sources fournit un intervalle I_i pour la variable m par l'équation :

$$\hat{I} = \bigcap_{i=1,\ldots,N} I_i$$

1.7.2 Le comportement disjonctif

Le comportement disjonctif est le pendant d'une union ensembliste. Le résultat $\phi(\pi_1, \ldots, \pi_n)$ d'un tel opérateur disjonctif contient toujours toutes les informations données par les sources. Un opérateur disjonctif augmente donc l'incertitude globale, et fournit un résultat moins précis que chacune des sources prise séparément. Il fait la supposition qu'au moins une des sources est fiable. Le résultat d'une telle opération est généralement très fiable, mais peut être très imprécis, ce qui réduit son utilité. L'opérateur disjonctif, pour une variable m, s'écrit :

$$\hat{\pi} = \bigvee_{i=1,\ldots,N} \pi_i$$

où $\bigvee$ est une conorme triangulaire (appelée aussi t-conorme) qui sont les opérateurs duaux[3]. ce qui se traduit dans le cas des intervalles, pour une variable m, par l'équation :

$$\hat{I} = \bigcup_{i=1,\ldots,N} I_i$$

1.7.3 Le comportement de compromis

Le résultat d'un comportement de compromis se situe entre les résultats de la disjonction et de la conjonction. De tels comportements sont généralement utilisés quand les informations fournies par les sources sont partiellement inconsistantes. L'objectif d'un tel comportement est d'obtenir un résultat qui ait un bon équilibre entre informativité et fiabilité. Nous distinguons deux types de compromis :

2. Une t-norme est une application de $[0,1] \times [0,1]$ dans $[0,1]$ commutative, associative, non-décroissante en chaque membre et ayant 1 comme élément neutre. L'opérateur minimum et le produit sont les plus utilisés.

3. La plus utilisée des t-conormes est l'opérateur maximum.

– adaptatif : un comportement de compromis est appelé adaptatif si le résultat dépend du contexte. Le but est de passer d'un comportement conjonctif à un comportement disjonctif au fur et à mesure que l'inconsistance entre les informations augmente. On retrouve alors la disjonction (respectivement conjonction) en cas d'inconsistance totale (respectivement consistance totale) entre les sources. Entre ces deux situations, le comportement est de compromis.
Dubois et Prade ont proposé une règle de fusion adaptative [75], dont le but est de prendre en compte le conflit entre les sources. Le principe est de déterminer deux valeurs p et q encadrant l'ensemble des sources fiables. Ces valeurs correspondent respectivement aux nombres de sources ayant une intersection non vide de leurs supports et aux nombres de sources ayant une intersection non vide de leurs noyaux respectivement. La règle de Dubois et Prade s'écrit :

$$\pi_{AD}(x) = \max\left(\frac{\pi_p(x)}{h(p)}, \min(\pi_q(x), 1 - h(p))\right) \tag{1.7}$$

avec

$$\begin{array}{rcl} h(T) & = & \sup_x(\min_{i\in T}\pi_i) \\ h(p) & = & \max(h(T), |T| = p) \end{array}$$

T est un sous-ensemble de sources et $|T|$ sa cardinalité. $h(T)$ est la plus grande intersection entre les sources appartenant à T. $h(p)$ est la plus grande valeur d'intersection non nulle du plus grand nombre de sources possibles. Ce terme traduit aussi la concordance entre les sources. Plus tard, Oussalah et al. [156] ont proposé une règle progressive adaptative pour la combinaison des données imparfaites qui possède deux caractéristiques de la règle de Dubois et Prade déjà citée. En effet, cette règle permet de prendre en compte la distance séparant la zone de consensus de chaque point du support et permet une élimination progressive et contrôlable d'une information en contradiction avec les autres. Elle est en outre robuste par rapport à la forme des distributions de possibilités. Une généralisation de la règle de Dubois et Prade est de plus donnée dans [155]. Ainsi, les méthodes utilisant les sous-ensembles maximaux cohérents, notion originaire de la logique [164], introduites dans plusieurs cadres de traitement de l'incertitude [55] sont de bons représentants des opérateurs de compromis. En résumé, la méthode fondée sur la recherche des sous-ensembles maximaux cohérents consiste d'une manière générale à utiliser un opérateur conjonctif sur les sous-ensembles non-conflictuels de sources, et à appliquer ensuite un opérateur disjonctif sur les résultats de ces derniers. Soit $E = \{E_1, \ldots, E_N\}$ l'ensemble des éléments d'information fournis par les sources pour une variable m. Un sous-ensemble $\mathcal{E} \subseteq E$ est dit cohérent si $\bigcap_{i=1,\ldots,|\mathcal{E}|} E_i \neq \emptyset$ avec $E_i \in \mathcal{E}$ et $\mathcal{E}$ est maximal. Ainsi, on obtient p ($p \leq N$) sous-ensembles maximaux cohérents et la méthode s'écrit :

$$\hat{\pi} = \bigvee_{j=1,\ldots,p} \bigwedge_{i=1,\ldots,|\mathcal{E}_j|} E_i$$

On retrouve la conjonction en cas d'absence de conflit et la disjonction si les opinions des sources sont conflictuelles deux à deux.

– non-adaptif : un comportement de compromis est non-adaptatif quand il se comporte toujours de la même manière, quelque soit le contexte. Les moyennes arithmétiques (ou combinaisons convexes) pondérées constituent un exemple typique de tels opérateurs, et sont de loin les opérateurs de fusion les plus utilisés en pratique.

$$\pi_{WA} = \sum_{i=1,\ldots,N} \lambda_i \pi_i$$

où λ_i sont des pondérations considérées pour mesurer la fiabilité des sources. Yager a introduit l'utilisation de la moyenne pondérée ordonnée (en anglais : Ordered Weighted Average OWA) [180] et a proposé une extension de cet opérateur dans [182].

1.8 Propriétés mathématiques des méthodes de fusion

Plusieurs propriétés des opérateurs de fusion dans le cadre de la théorie des possibilités sont étudiées [154, 55]. Formellement, nous considérons N sources délivrant des informations représentées par $\pi_1, \ldots, \pi_N$ pour chacune des sources. Le résultat de la fusion de p sources est donc donné par $\phi(a_1, \ldots, a_p)$.

1. **Commutativité** [173] : Un opérateur ϕ est commutatif si $\phi(\pi_i, \pi_j) = \phi(\pi_j, \pi_i)$ pour tout $i \neq j$ dans $\{1, \ldots, N\}$. Cette propriété est importante lorsqu'on n'est pas capable de distinguer les sources entre elles.
2. **Associativité** [173] : Un opérateur ϕ est associatif si $\phi(\pi_i, \phi(\pi_j, \pi_k)) = \phi(\phi(\pi_i, \pi_j), a_k)$ pour tout i, j et k de l'ensemble $\{1, \ldots, N\}$. Cette propriété permet de fusionner les informations les unes après les autres, c-à-d de fusionner d'abord π_1 et π_2, puis d'ajouter π_3, etc. L'associativité de l'opérateur de fusion est importante notamment lorsque la fusion n'est pas opérée en une seule fois, c-à-d quand on combine des informations en en attendant d'autres qui ne sont pas encore fournies par d'autres sources. Néanmoins, l'associativité est également pratique dans le calcul.
3. **Idempotence** : Un opérateur de fusion ϕ est idempotent si $\phi(\pi_i, \pi_i) = \pi_i$ pour toute valeur i entre 1 et N. Si toutes les sources fournissent la même information, l'idempotence garantit qu'il n'y aura pas d'effet de renforcement entre les sources.
4. **Cohérence totale** : Cette propriété est vérifiée si le noyau du résultat de la fusion $\phi(\pi_1, \ldots, \pi_N)$ n'est pas vide. Ce qui est équivalent à éliminer le conflit entre les sources.
5. **Préservation maximale** [154] : Un opérateur ϕ satisfait cette propriété si ϕ ne prend en compte que les informations fournies par les sources, c-à-d qu'un élément considéré impossible par les sources doit être également considéré impossible par le résultat de la fusion.
6. **Plausibilité maximale faible** : Un opérateur ϕ vérifie cette propriété si un élément considérée plausible par toutes les sources est également considéré plausible par l'information résultant de la fusion. Cette propriété est vraie si toutes les sources sont en conflit (aucun des éléments n'est considéré comme plausible par toutes les sources).
7. **Plausibilité maximale stricte** [154] : Un opérateur ϕ vérifie cette propriété si un élément considéré plausible par au moins une des sources est également considéré plausible par l'information résultant de la fusion. Cette propriété n'est pas souhaitable quand il s'agit de réduire l'incertitude à partir du processus de fusion puisqu'il tend à rendre le résultat de la fusion disjonctif et consistant avec toutes les informations données par les sources.
8. **Pertinence** [154] : Un opérateur ϕ satisfait cette propriété si la fusion conjonctive de n'importe quelle combinaison de sources de 1 à N avec une source π_i,$i = 1, \ldots, N$ n'est pas vide. Ce qui se traduit par la cohérence partielle de l'information résultante avec chaque information délivrée par les sources.
9. **Insensibilité avec l'ignorance totale** : Un opérateur ϕ satisfait cette propriété si tout ajout d'une nouvelle information ne change pas le résultat obtenu à partir des sources existantes. Formellement, si π_{N+1} représente l'ignorance totale de la source ajoutée alors

$\phi(\pi_1, \pi_2, \ldots, \pi_N, \pi_{N+1}) = \phi(\pi_1, \pi_2, \ldots, \pi_N)$. Si π_{N+1} est moins informative que toutes les sources existantes alors l'opérateur ϕ vérifie cette propriété.

10. **Convexité** [173] : Cette propriété est vérifiée par un opérateur de fusion si l'information résultante est convexe dans les situations où $\pi_1, \ldots, \pi_N$ sont convexes. En général, avoir un résultat de fusion convexe facilite le calcul mais il n'existe aucune raison pour admettre toujours cette propriété. Il est toujours possible de considérer l'enveloppe convexe du résultat de la fusion s'il n'est pas convexe, mais on perd de l'information dans le processus en gagnant en efficacité de calcul.
11. **Robustesse/continuité** [173] : Un opérateur ϕ satisfait cette propriété si tout changement de l'opinion d'une source conduit à un changement du résultat de la fusion.

Les opérateurs conjonctifs vérifient les propriétés de commutativité, d'associativité, d'idempotence, de convexité, la propriété de pertinence ainsi que les propriétés de la plausibilité maximale. Les propriétés de cohérence totale et de plausibilité maximale stricte ne sont pas vérifiées puisque les opérateurs conjonctifs doivent être choisis dans les situations où les sources sont fiables et cohérentes.

Les opérateurs disjonctifs vérifient la propriété de commutativité, d'associativité, d'idempotence, la propriété de cohérence totale, la propriété de plausibilité maximale, la pertinence et la robustesse. Néanmoins, les opérateurs disjonctifs ne vérifient pas la propriété de convexité en général. De plus, en ajoutant une nouvelle information représentée par une ignorance totale, l'information résultant de la fusion disjonctive change. Par conséquent l'opérateur disjonctif n'obéit pas à la propriété d'insensibilité avec l'ignorance totale.

Les opérateurs de compromis adaptatifs auxquels nous nous intéressons pour la méthode de fusion fondée sur les SMC sont commutatifs, idempotents mais ils ne sont pas associatifs. Le résultat de la fusion des opérateurs adaptatifs n'est pas convexe en général puisque l'opérateur de fusion est une disjonction. De plus, l'opérateur ne satisfait pas la propriété de robustesse. Cependant la propriété de cohérence totale est vérifiée, l'ajout d'une information représentée par une ignorance totale est aussi vérifié. Les propriétés de plausibilité maximale stricte et faible sont satisfaites. Généralement quand une source fournit une information, l'opérateur de fusion des SMC ne la considère pas forcément.

La Table 1.1 résume les propriétés des opérateurs de fusion conjonctif, disjonctif et de compromis adaptatifs de la Section 1.7.

		Mode de fusion		
		Conjonctif (au sens de min)	Disjonctif	Adaptatifs
Propriétés	Commutativité	×	×	×
	Associativité	×	×	
	Idempotence	×	×	×
	Cohérence totale		×	×
	Préservation maximale	×		×
	Plausibilité maximale faible	×	×	×
	Plausibilité maximale stricte		×	
	Pertinence	×	×	×
	Insensibilité à l'ignorance totale	×		×
	Convexité	×		
	Robustesse		×	

TABLE 1.1 – Propriétés des modes de fusion

2

Analyse formelle de concepts

Sommaire

De nombreuses méthodes ont été développées et regroupées sous le nom de la "fouille de données". L'idée est de fouiller dans une masse de données pour en extraire de nouvelles informations sur les liaisons existant entre des sous-ensembles de données. Ceci peut servir à réutiliser les données pour d'autres besoins et à analyser des données pour découvrir des phénomènes inconnus. Dans plusieurs domaines, les observations sont disponibles mais l'utilisateur ne sait pas comment les utiliser.

L'analyse formelle de concepts est une méthode ensembliste permettant de générer et de structurer un ensemble de concepts à partir des relations entre des objets et des propriétés. Nous rappelons ici l'analyse formelle de concepts et ses extensions dans un premier temps. Ensuite, nous montrons quelques travaux s'appuyant sur ce formalisme pour la fusion d'informations. Nous réutiliserons l'AFC et les techniques de construction de treillis dans les Chapitres 3 et 4.

2.1 Analyse Formelle de Concepts

L'Analyse Formelle de Concepts (AFC) (appelée aussi Analyse de Concepts Formels (AFC)) [177, 96] est un formalisme mathématique pour l'analyse de données, la représentation de connaissances et la visualisation de connaissances. L'idée de base de l'AFC est d'extraire des concepts regroupant des objets et leurs propriétés/attributs à partir de données et de construire une hiérarchie à partir de ces concepts.

2.1.1 Contexte formel

Définition 1 (Contexte formel) *Un **contexte formel** est un triplet (G, M, R) où G est un ensemble d'objets, M est un ensemble de propriétés/attributs et R est une relation binaire entre G et M. Un couple $(g, m) \in R$ (également noté gRm) signifie que l'objet $g \in G$ possède la propriétés $m \in M$.*

Un contexte formel peut être représenté sous la forme d'un tableau à deux dimensions où les lignes correspondent aux objets et les colonnes aux attributs. Les cases du tableau sont remplies suivant la présence/absence de la propriété, autrement dit si le i$^{\text{ème}}$ objet g est en relation R avec le j$^{\text{ème}}$ alors la case à l'intersection de la ligne i et de la colonne j contient "$\times$", sinon la case est vide. La table 2.1 [54] donne un exemple de contexte formel représentant les planètes du système solaire. Dans cet exemple, Mercure possède les attributs : petite taille, proche du soleil et n'est pas un satellite.

TABLE 2.1 – Un contexte formel représentant les planètes du système solaire.

Attribut / Objet	Taille			Distance au soleil		Satellite	
	petite	moyenne	grande	proche	loin	oui	non
Mercure	×			×			×
Vénus	×			×			×
Terre	×			×		×	
Mars	×			×		×	
Jupiter			×		×	×	
Saturne			×		×	×	
Uranus		×			×	×	
Neptune		×			×	×	
Pluton	×				×	×	

2.1.2 Connexion de Galois dans un contexte formel

Définition 2 *Soit (G, M, R) un contexte formel. Pour tout $A \subseteq G$ et $B \subseteq M$, on définit :*

$$A' = \{m \in M | \forall g \in A, gRm\}$$

$$B' = \{g \in G | \forall m \in B, gRm\}$$

Intuitivement, A' est l'ensemble des attributs communs à tous les objets de A et B' est l'ensemble des objets possédant tous les attributs de B. Les applications $' : 2^G \to 2^M$ et $' : 2^M \to 2^G$ sont appelées opérateurs de dérivation entre l'ensemble de parties de G noté par 2^G et l'ensemble de parties de M noté par 2^M. La composition de ces opérateurs produit deux opérateurs $'' : 2^G \to 2^G$ et $'' : 2^M \to 2^M$. Le premier opérateur permet d'associer à un ensemble d'objets A l'ensemble maximal d'objets dans G ayant les attributs communs aux objets de A. Cet ensemble est noté A''. De façon duale, le second opérateur permet d'associer à un ensemble d'attributs B l'ensemble maximal d'attributs dans M communs aux objets ayant les attributs dans B. Cet ensemble est noté B''. Les opérateurs $'' : 2^G \to 2^G$ et $'' : 2^M \to 2^M$ définissent deux fermetures, respectivement sur l'ensemble des parties de G et sur l'ensemble des parties de M. Les ensembles A'' et B'' sont des fermés pour ces deux opérateurs respectifs. L'ensemble des fermés de 2^G muni de l'inclusion est un treillis complet. De la même façon, l'ensemble des fermés de 2^M muni de l'inclusion est un treillis complet. Les opérateurs de dérivation $' : 2^G \to 2^M$ et $' : 2^M \to 2^G$ forment une bijection entre les ensembles de fermés de 2^G et 2^M et définissent un isomorphisme entre les deux treillis respectifs : à chaque fermé A dans 2^G correspond un unique fermé B dans 2^M et vice versa. De cette façon, les opérateurs de dérivation $'' : 2^G \to 2^G$ et $'' : 2^M \to 2^M$ forment une connexion de Galois entre $(2^G, \subseteq)$ et $(2^M, \subseteq)$.

2.1.3 Concept formel

Soit (G, M, R) un contexte formel. Un **concept formel** est un couple (A, B) tel que $A \subseteq G$, $B \subseteq M$, $A' = B$ et $B' = A$. A et B sont respectivement appelés *extension* (extent) et *intension* (intent) du concept formel (A, B).

Dans un contexte formel, un concept correspond à un rectangle maximal de la table formée par la relation binaire du contexte : tout objet de l'extension a tous les attributs de l'intension. Il est important de noter que cette notion de rectangle maximal est indépendante de l'ordre des lignes et des colonnes. Ces ensembles maximaux d'objets et d'attributs correspondent à des fermés dans 2^G et 2^M respectivement. Un sous-ensemble B de M est l'intension d'un concept formel dans (G, M, R) si et seulement si $B'' = B$ (B est fermé pour $''$) et, de façon duale, un sous ensemble A de G est l'extension d'un concept formel dans l'ensemble de concepts du contexte (G, M, R) si et seulement si $A'' = A$ (A est fermé pour $''$).

Les concepts du treillis du contexte (G, M, R) sont ordonnés par une relation d'ordre hiérarchique entre concepts (appelée aussi relation de subsomption) notée $\leq$ et définie comme suit.

Définition 3 (Relation de "subsomption") *Soient (A_1, B_1) et (A_2, B_2) deux concepts formels de l'ensemble des concepts du contexte (G, M, R). $(A_1, B_1) \leq (A_2, B_2)$ si et seulement si $A_1 \subseteq A_2$ (ou de façon duale $B_2 \subseteq B_1$). (A_2, B_2) est dit **super-concept** de (A_1, B_1) et (A_1, B_1) est dit **sous-concept** de (A_2, B_2). La relation "$\leq$" est dite relation de subsomption.*

La relation "$\leq$" s'appuie sur deux inclusions duales, entre ensembles d'objets et entre ensembles d'attributs et peut ainsi être interprétée comme une relation de généralisation/spécialisation entre les concepts formels. Un concept est plus général qu'un autre concept s'il contient plus d'objets dans son extension. En contre partie, les attributs partagés par ces objets sont réduits. De façon duale, un concept est plus spécifique qu'un autre s'il contient moins d'objets dans son extension. Ces objets ont plus d'attributs en commun.

2.1.4 Treillis de concepts

Définition 4 (Treillis de concepts) *La relation "$\leq$" permet d'organiser les concepts formels en un treillis complet appelé* ***treillis de concepts*** *ou encore* ***treillis de Galois*** *[23]. La borne supérieure et la borne inférieure dans $\underline{\mathfrak{C}}(G, M, R)$ sont données par :*

$$\bigwedge_{j \in J} (A_j, B_j) = \left(\bigcap_{j \in J} A_j, \left(\bigcup_{j \in J} B_j \right)'' \right)$$

$$\bigvee_{j \in J} (A_j, B_j) = \left(\left(\bigcup_{j \in J} A_j \right)'', \bigcap_{j \in J} B_j \right)$$

Le diagramme de Hasse représentant le treillis de concepts correspondant au contexte formel donné dans la table 2.1 est donné dans la figure 2.1 (visualisé grâce au logiciel ConExp [4]). La notation des concepts est dite réduite (dite aussi étiquetage réduit) : elle s'appuie sur l'héritage à la fois des attributs et des objets entre les concepts du treillis, autrement dit un objet et un attribut n'apparaissent qu'une seule fois sur le diagramme. Les attributs sont placés au plus haut dans le treillis : à chaque fois qu'un nœud N est étiqueté par un attribut m, tous les descendants de N dans le treillis héritent l'attribut m. De façon duale, les objets sont placés au plus bas dans le treillis : à chaque fois qu'un nœud N est étiqueté par un objet g, g est hérité "vers le haut" et tous les ancêtres de N le partagent. Ainsi l'extension A d'un concept (A, B) est obtenue en considérant tous les objets qui apparaissent sur les descendants du nœud N dans le treillis et son intension B est obtenue en considérant tous les attributs qui apparaissent sur les ancêtres du nœud N dans le treillis. Le treillis de concepts est une représentation équivalente des données contenues dans un contexte formel qui met en avant les groupements possibles entre objets et attributs ainsi que les relations d'inclusion entre ces groupements. De plus, la représentation graphique du treillis de concepts, sous la forme d'un diagramme de Hasse, facilite la compréhension et l'interprétation de la relation entre les objets et les attributs d'une part, et entre objets ou attributs d'autre part. L'avantage de cette représentation est qu'à partir d'un treillis de concepts il est toujours possible de retrouver le contexte formel correspondant et inversement.

2.1.5 Algorithmes de construction de treillis de concepts

Plusieurs algorithmes ont été proposés pour la construction de treillis : **Close by one** [124], **Chein** [46], **Norris** [153], **NextClosure** [93, 95], **Bordat** [34]), et ont fait l'objectif de plusieurs comparaisons [127, 152]. Dans leur article, Kuznetsov et Obiedkov [127] analysent plusieurs algorithmes de construction de treillis de concepts. Ils présentent une étude de leurs complexités théoriques et une comparaison expérimentale sur des jeux de données artificiels. Les auteurs font des recommandations en fonction de la nature du contexte.

De plus, il existe plusieurs outils qui permettent d'éditer des contextes formels, et de construire le treillis de concepts associé : ConExp et Galicia sont deux outils libres couramment utilisés. Une liste plus complète d'autres logiciels peut être consultée sur `www.upriss.org.uk/fca`.

2.2 AFC et Recherche d'Information

L'AFC est un formalisme qui s'appuie sur une théorie mathématique, la théorie des treillis [23] en l'adaptant afin de permettre d'analyser les données sous la forme d'un contexte formel et de

4. http ://sourceforge.net/projects/conexp/

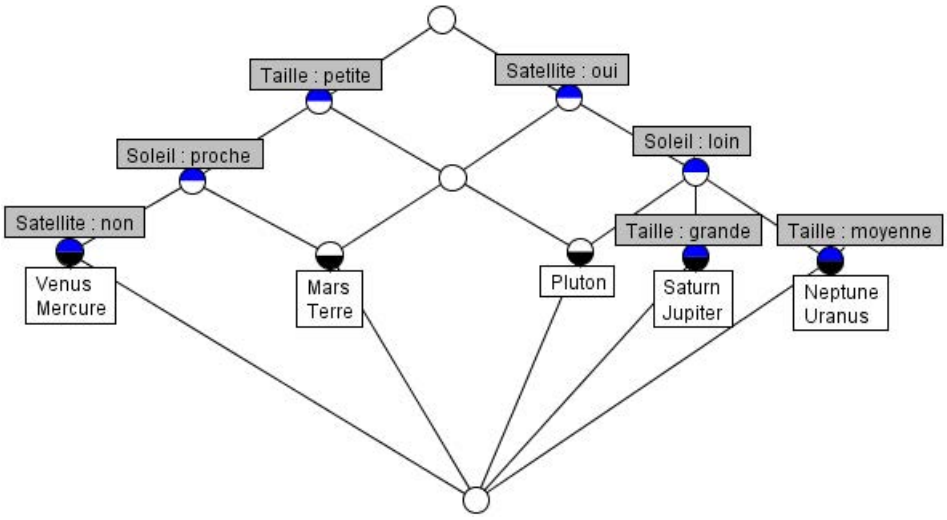


FIGURE 2.1 – Le treillis de concepts correspondant au contexte formel (G, M, R) donné dans la table 2.1

structurer ces données sous forme d'un treillis. La structure de treillis permet de visualiser les données grâce à la notion de "concept". Le treillis permet d'obtenir des classes structurées et interprétées automatiquement en faisant ressortir des relations entre les objets, mais aussi entre les attributs, et également entre les objets et les attributs. L'AFC est utilisée dans plusieurs domaines d'application d'analyse et d'exploitation de données tels que la classification, la recherche d'information [42], la selection de services Web [15], la construction d'ontologies [21], la découverte de ressources [138, 139, 59], l'extraction de connaissances [128], l'apprentissage machine [125], l'ingénierie des logiciels [171, 103], la linguistique [162], etc.

Les treillis de concepts ont deux principales caractéristiques : la structuration conceptuelle des données et l'ordre hiérarchique entre les concepts.

2.2.1 La structuration conceptuelle

Dans un treillis de concepts les données sont structurées sous forme de concepts. Un concept peut être vu comme une classe d'objets (l'extension du concept) caractérisée par un ensemble de propriétés (l'intension du concept).

2.2.2 Les liens généralisation/spécialisation

Dans un treillis de concepts les concepts sont ordonnés selon deux critères duaux liés à leurs extensions et à leurs intensions (voir définition 3). Les concepts les plus généraux sont situés en haut du treillis alors que les concepts les plus spécifiques sont situés en bas du treillis. Les liens entre les concepts peuvent être interprétés comme des généralisations ou des spécialisations entres les classes représentées par les concepts. En effet, un parcours ascendant des concepts d'un treillis se traduit à chaque étape par la diminution progressive du nombre d'attributs dans les intensions des concepts et l'augmentation progressive du nombre d'objets dans leurs extensions.

TABLE 2.2 – Mesures réelles relatives aux planètes du système solaire.

	Diamètre (km)	Distance au Soleil (10^6km)	Satellite
Mercure	4 878	58	0
Venus	12 400	108	0
Terre	12 756	150	1
Mars	6 800	228	2
Jupiter	142 800	778	16
Saturne	120 800	1 427	19
Uranus	47 600	2 870	5
Neptune	44 600	4 500	8
Pluton	2.320	9 950	1

Cela correspond au passage d'une classe plus spécifique, qui contient peu d'objets qui vérifient plusieurs critères, à une classe plus générale, qui contient plus d'objets qui ne vérifient qu'une partie des critères de la classe spécifique. De façon duale, un parcours descendant des concepts d'un treillis correspond au passage d'une classe générale à une classe plus spécifique.

2.3 AFC et données complexes

L'AFC est un formalisme défini pour analyser les données du monde réel. Cependant, ces données ne se présentent pas forcément sous la forme d'un contexte binaire, la structure de base des données analysées par l'AFC. Par exemple, dans le cas de la recherche documentaire, une façon plus précise de représenter la relation entre les documents et leurs termes d'indexation est de considérer les fréquences d'apparition des termes dans les documents au lieu de la simple information binaire indiquant la présence ou l'absence d'un terme dans un document. De même, dans le cas des données relatives au système solaire, il est possible de représenter des valeurs de mesures réelles des diamètres des planètes ou de leurs distances au soleil au lieu des simples attributs qualificatifs indiquant qu'une planète est grande ou petite et qu'elle est proche ou éloignée du soleil. Les valeurs réelles des mesures relatives aux planètes du système solaire représenté par le contexte formel donné dans la table 2.1 peuvent être représentées avec plus de précision sous la forme d'un contexte non binaire, en particulier numérique. Dans ce cas, les cases du tableau contiennent les valeurs de la mesure des attributs pour chaque planète. Le contexte multivalué obtenu est donné dans la table 2.2. L'application des résultats de l'AFC à un contexte numérique nécessite la transformation de celui-ci en un contexte binaire. Cette étape de transformation s'appelle **échelonnage conceptuel** [96] ou discrétisation.

L'**échelonnage conceptuel** (*conceptual scaling*) consiste à transformer chaque attribut multivalué en un ensemble d'attributs binaires qui forment un contexte binaire appelé **échelle conceptuelle** (*conceptual scale*) de l'attribut multivalué. Une échelle conceptuelle est un contexte formel dont les objets sont des valeurs et les attributs sont des attributs d'échelle. Ce contexte permet de structurer le domaine de valeurs de cet attribut sous la forme d'un treillis de concepts qui définit une hiérarchie entre les attributs d'échelle.

Reprenons l'exemple du contexte numérique multivalué des planètes du système solaire donné dans la table 2.2. Une échelle conceptuelle possible pour l'attribut "*Distance au soleil*" est donnée

dans la table 2.3 [41]. Les attributs binaires de l'échelle dans cet exemple sont "≥ 10", "≥ 100", "≥ 500" et "≥ 1000". De la même manière on peut définir des échelles pour les attributs multivalués "Diamètre" et "Satellite(s)". Les échelles conceptuelles qui correspondent aux deux attributs sont données dans les tables 2.4 et 2.5, respectivement.

TABLE 2.3 – Échelle conceptuelle de l'attribut "*Distance au soleil*" et le treillis correspondant.

	$ds \geq 10$	$ds \geq 100$	$ds \geq 500$	$ds \geq 1000$
58	×			
108	×	×		
150	×	×		
228	×	×		
778	×	×	×	
1 427	×	×	×	×
2 870	×	×	×	×
4 500	×	×	×	×
9 950	×	×	×	×

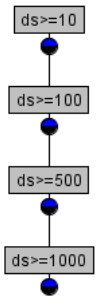


TABLE 2.4 – Échelle conceptuelle de l'attribut "*Diamètre*" et le treillis correspondant.

	$d \leq 10$	$10 < d \leq 20$	$20 < d \leq 50$	$d \geq 50$
4 878	×			
12 400		×		
12 756		×		
6 800	×			
142 800				×
120 800				×
47 600			×	
44 600			×	
2.320	×			

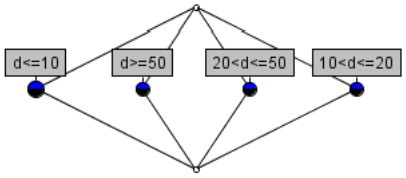
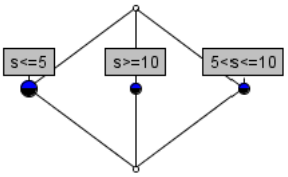


TABLE 2.5 – Échelle conceptuelle de l'attribut "*Satellite*" et le treillis correspondant.

	$s \leq 5$	$5 < s \leq 10$	$s \geq 10$
0	×		
0	×		
1	×		
2	×		
16			×
19			×
5	×		
8		×	
1	×		



La transformation n'est pas unique et ne suit pas une règle générale de transformation. Elle dépend d'un choix particulier des seuils de découpage de domaine. Elle dépend ainsi d'une interprétation particulière de l'attribut multivalué et des valeurs qu'il prend dans le contexte

multivalué. Par conséquent, chaque échelonnage fournit un treillis et les treillis résultants n'ont pas la même interprétation. Cependant, Ganter et Wille ont défini une quinzaine d'échelonnages conceptuels typiques pour des attributs multivalués de formats particuliers [96]. Considérant ces définitions, l'échelle conceptuelle de l'attribut multivalué "*Distance au soleil*" est obtenue en effectuant un échelonnage **ordinal** tandis que celles de "Diamètre" et "Satellite(s)" sont obtenues en effectuant un échelonnage **nominal** [96].

Le contexte binaire (monovalué) est obtenu à partir du contexte numérique (multivalué), en fonction des échelles conceptuelles de chaque attribut. L'ensemble des objets est conservé et l'ensemble des attributs est l'union disjointe des attributs des échelles conceptuelles. Une fois que le contexte monovalué est obtenu, on peut construire le treillis de concepts correspondant par l'une des méthodes de construction de treillis détaillé dans la section 2.1.4.

Le contexte monovalué obtenu à partir du contexte multivalué des planètes du système solaire est donné dans la table 2.6. Le treillis de concepts correspondant à ce contexte est donné dans la figure 2.2.

TABLE 2.6 – Contexte binaire résultant de l'échelonnage conceptuel du contexte multivalué des planètes du système solaire.

	Diamètre (km)				Distance au Soleil (10^6km)				Satellite		
	$d \leq 10$	$10 < d \leq 20$	$20 < d \leq 50$	$d > 50$	$ds \geq 10$	$ds \geq 100$	$ds \geq 500$	$ds \geq 1000$	$s \leq 5$	$5 < s \leq 10$	$s > 10$
Mercure	×				×				×		
Venus		×			×	×			×		
Terre		×			×	×			×		
Mars	×				×	×			×		
Jupiter				×	×	×	×				×
Saturne				×	×	×	×	×			×
Uranus			×		×	×	×	×	×		
Neptune			×		×	×	×	×		×	
Pluton	×				×	×	×	×	×		

2.4 Extensions de l'AFC

Le choix d'un échelonnage particulier est un compromis entre perte d'information, explosion du nombre de concepts et calculabilité du treillis. Ces problèmes sont souvent aggravés par la complexité des données et leur grande quantité dans le cadre des applications réelles. Pour pallier ces limites, plusieurs travaux de recherche ont été menés dans le but d'étendre les définitions de l'AFC pour couvrir les données complexes. Cependant, la diversité des données étudiées et les caractéristiques particulières de chaque domaine d'application étudié dans le cadre de ces travaux ont abouti à la définition de différentes extensions. Chacune de ces extensions est adaptée à un format particulier de données et à une interprétation particulière des attributs complexes multivalués et de leurs valeurs dans les contextes "hétérogènes". Le point commun de

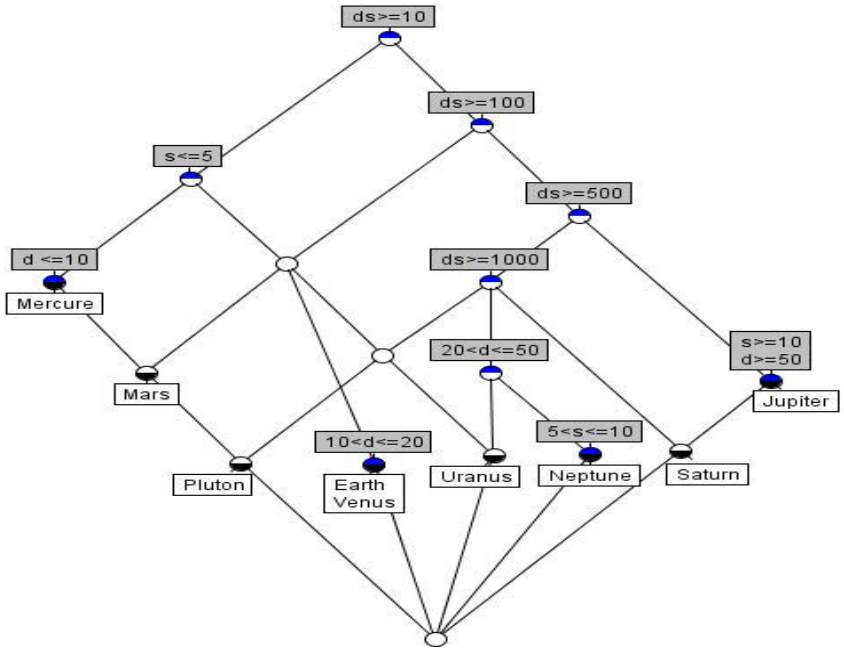


FIGURE 2.2 – Le treillis de concepts correspondant au contexte formel donné dans le tableau 2.6.

ces approches est qu'elles redéfinissent la connexion de Galois entre les objets et les attributs en vue de l'adapter à la particularité du domaine étudié.

2.4.1 Les structures de patrons

Les *structures de patrons* ont été introduite par Ganter et Kuznetsov [94] pour construire un treillis de concepts à partir de données complexes sans les avoir transformées sous forme binaire et sans perte d'information. La connexion de Galois classique met en évidence une relation entre les éléments du treillis de $(2^G, \subseteq)$ et les éléments du treillis d'attributs $(2^M, \subseteq)$ et *vice versa*. Ces treillis sont partiellement ordonnés. Pour construire un treillis à partir d'un contexte hétérogène [5], il faut donc trouver un ordre partiel sur les descriptions des objets dans le contexte hétérogène. Pour cela, Ganter et Kuznetsov ont introduit les structures de patrons en formalisant l'ensemble des objets de G et les descriptions (appelées patrons) d de l'inf-demi-treillis des descriptions ordonnées. Revenons au cas classique de l'AFC, en considérant le treillis des attributs $(2^M, \subseteq)$. Soient P et Q deux sous-ensembles d'attributs, on a $P \subseteq Q \Leftrightarrow P \cap Q = P$, par exemple avec $P = \{a\}$ et $Q = \{a, b\}$, $\{a\} \subseteq \{a, b\} \Leftrightarrow \{a\} \cap \{a, b\} = \{a\}$. L'intersection ensembliste possède les propriétés d'un infimum dans un demi-treillis, d'où l'idée d'ordonner les patrons suivant la relation :

$$c \sqsubseteq d \Leftrightarrow c \sqcap d = c \tag{2.1}$$

Formellement, on considère le triplet $(G, (D, \sqcap), \delta)$ où G est l'ensemble des objets, $(D, \sqcap)$ est l'inf-demi-treillis des descriptions, et δ est la fonction qui associe à chaque objet sa description

5. Un contexte est dit hétérogène ou complexe s'il n'est pas binaire. Il est appelé aussi "contexte multivalué" [96].

dans D. Le triplet $(G, (D, \sqcap), \delta)$ est appelé structure de patrons et la construction du treillis suit l'AFC classique. Pour établir la connexion de Galois entre $(2^G, \subseteq)$ et $(D, \sqsubseteq)$, les opérateurs de dérivation sont les suivants : Pour $A \subseteq G$

$$A^{\square} = \bigsqcap_{g \in A} \delta(g) \tag{2.2}$$

Pour $d \subseteq (D, \sqcap)$

$$d^{\square} = \{g \in G | d \sqsubseteq \delta(g)\} \tag{2.3}$$

Le premier opérateur retourne pour un sous-ensemble A d'objets de G l'*infimum* de leurs descriptions. De manière duale, le second opérateur retourne pour toute description d dans l'inf-demi-treillis le sous-ensemble d'objets ayant une description qui subsume la description d.

Les concepts de $(G, (D, \sqcap), \delta)$ sont des couples (A, d) tels que $A^{\square} = d$ et $d^{\square} = A$. Ils sont ordonnées par $(A_1, d_1) \leq (A_2, d_2) \Leftrightarrow (A_1 \subseteq A_2)(\Leftrightarrow d_2 \sqsubseteq d_1)$ pour former le treillis de concepts de $(G, (D, \sqcap), \delta)$.

Dans d'autres travaux comme [120, 121, 117, 118], les auteurs appellent $\sqcap$ opérateur de "similarité" dans le cadre des données numériques. Le premier opérateur $(.)^{\square}$ retourne la similarité entre les objets. De manière duale, le second opérateur $(.)^{\square}$ donne pour une description le plus grand ensemble d'objets dont la similarité est représentée par cette description. Considérons l'exemple du contexte numérique représentant les mesures réelles des caractéristiques des planètes du système solaire de la table 2.7 et l'infimum considéré dans [120, 121, 117, 118] pour les données numériques. Ces travaux définissent l'infimum de deux intervalles est le plus petit intervalle contenant les deux intervalles (voir l'équation 2.4. La relation d'ordre représente la relation d'inclusion des intervalles (voir l'équation 2.5).

$$[a, b] \sqcap [c, d] = [\min(a, c), \max(b, d)] \tag{2.4}$$

$$[a, b] \sqsubseteq [c, d] \Leftrightarrow [a, b] \supseteq [c, d] \tag{2.5}$$

TABLE 2.7 – Mesures réelles des caractéristiques des planètes du système solaire

	Diamètre (km)	Distance au Soleil (10^6km)	Masse (10^{23}kg)	Satellite
Mercure	4 879	58	3.30	
Vénus	12 104	108	48.69	
Terre	12 756	150	59.74	1
Mars	6 794	228	6.42	2
Jupiter	142 984	778	18 988	16
Saturn	120 536	1 427	5 685	19
Uranus	51 118	2 870	866.25	5
Neptune	49 528	4 500	1027.8	8
Pluton	2 302	5 950	0.13	1

Si on considère la variable "Distance au Soleil" (notée DS) et le sous-ensemble de planètes $\{Mercure, Vénus\}$, alors :

Le premier retourne la similarité d'un ensemble d'objets, i.e. l'infimum de leurs descriptions :

$$\begin{aligned}\{Mercure, Vénus\}^{\square} &= \bigsqcap_{g \in \{Mercure, Vénus\}} \delta(g) \\ &= \delta(Mercure) \sqcap \delta(Vénus) \\ &= [58, 58] \sqcap [108, 108] \\ &= [58, 108]\end{aligned}$$

Le second retourne l'ensemble maximal d'objets dont la similarité est représentée par une description en argument :

$$\begin{aligned}[58, 108]^{\square} &= \{g \in G \mid [58, 108] \sqsubseteq \delta(g)\} \\ &= \{g \in G \mid [58, 108] \supseteq \delta(g)\} \\ &= \{Mercure, Vénus\}\end{aligned}$$

Ainsi le couple $(\{Mercure, Vénus\}, [58, 108])$ est tel que $A^{\square} = d$ et $d^{\square} = A$. Par conséquent c'est un concept pour la structure de patrons représentée par le contexte numérique des mesures réelles des données de planètes du système solaire.

Traitement de plusieurs variables

Dans de nombreuses applications, les descriptions d'objets ne sont pas aussi simples et s'appuient plutôt sur des vecteurs de nombres (ou vecteur d'intervalles). Par exemple, dans la table 2.7, l'objet *Mercure* est décrit par le vecteur $\langle [4879, 4879], [58, 58], [3.3, 3.3], [0, 0] \rangle$. D'où la définition d'un nouveau type de patrons : les vecteurs d'intervalles [121].

Définition 5 (Vecteur d'intervalles) *[121]. Un vecteur d'intervalles est un vecteur à p dimensions, composé uniquement de p intervalles.*

L'infimum d'un vecteur d'intervalles est défini comme le vecteur d'intervalles des infima des intervalles pour chacune des dimensions. Chaque variable est considérée comme une dimension, l'infimum est appliqué sur chacune des variables. Les variables sont supposées indépendantes et le test de la relation de subsomption se fait par dimension [121].

Pour e et f deux vecteurs d'intervalles, avec $e = \langle [a_i, b_i] \rangle_{i \in [1,p]}$ et $f = \langle [c_i, d_i] \rangle_{i \in [1,p]}$, leur infimum est défini comme suit :

$$\begin{aligned}e \sqcap f &= \langle [a_i, b_i] \rangle_{i \in [1,p]} \sqcap \langle [c_i, d_i] \rangle_{i \in [1,p]} \\ &= \langle [a_i, b_i] \sqcap [c_i, d_i] \rangle_{i \in [1,p]} \end{aligned} \tag{2.6}$$

Les vecteurs d'intervalles sont aussi partiellement ordonnés au sein d'un demi-treillis par :

$$\begin{aligned}e \sqsubseteq f &\Leftrightarrow \langle [a_i, b_i] \rangle_{i \in [1,p]} \sqsubseteq \langle [c_i, d_i] \rangle_{i \in [1,p]} \\ &\Leftrightarrow \langle [a_i, b_i] \sqsubseteq [c_i, d_i] \rangle_{i \in [1,p]} \end{aligned} \tag{2.7}$$

Considérons la table 2.7, on a :

$$\begin{aligned}\{Mercure, Vénus\}^{\square} &= \bigsqcap_{g \in \{Mercure, Vénus\}} \delta(g) \\ &= \delta(Mercure) \sqcap \delta(Vénus) \\ &= \langle [4879, 12104], [58, 108], [3.3, 48.69], [0, 0] \rangle\end{aligned}$$

$$\begin{aligned}\langle[4879,12104],[58,108],[3.3,48.69],[0,0]\rangle^{\square} &= \{g \in G \mid \langle[4879,12104],[58,108],[3.3,48.69],[0,0]\rangle \supseteq \delta(g)\} \\ &= \{Mercure, Vénus\}\end{aligned}$$

Algorithmes de construction de treillis

Les algorithmes classiques de l'AFC sont adaptés pour construire le treillis résultant des structures de patrons [94]. Pour calculer les concepts formels d'une structure de patrons, les auteurs utilisent plusieurs algorithmes classiques de l'AFC *Norris, CloseByOne* et *NextClosure* légèrement modifiés pour les besoins [121].

2.4.2 Analyse de concepts logiques

L'**Analyse de Concepts Logiques** (ACL) [87] consiste à étendre les résultats de l'AFC aux contextes logiques. Un contexte logique est un contexte multivalué dans lequel les attributs sont des descriptions qui prennent comme valeurs des formules logiques décrivant les objets du contexte.

Définition 6 (Contexte logique) *Un **contexte logique** est un triplet $(G, \mathcal{L}, \delta)$ où G est un ensemble fini d'objets, $\mathcal{L}$ est un langage, et δ est une fonction de G dans $\mathcal{L}$ qui associe à chaque objet une formule décrivant les propriétés de l'objet (ou une description logique).*

Un exemple de contexte logique est donné dans la table 2.8 (gauche) [87].

Les opérateurs de dérivation entre 2^G et $\mathcal{L}$ sont définis comme suit.

$$\sigma : 2^G \to \mathcal{L}, \;\; \sigma_{(A)} = \sqcup_{g \in A} \delta(g)$$

$$\tau : \mathcal{L} \to 2^G, \;\; \tau(f) = \{g \in G \mid \delta(g) \sqsubseteq f\}$$

$\sqcup$ est l'opérateur de disjonction (correspond à l'union ($\cap$) en AFC et à l'opérateur *infimum* en structure de patrons). $\sqsubseteq$ est la relation de subsomption sur les formules logiques (correspond à l'inclusion en AFC et à $\sqsubseteq$ en structure de patrons). $\sigma(A)$ est l'expression par une formule logique des propriétés communes à tous les objets dans A et $\tau(f)$ est l'ensemble de tous les objets de G dont la description est subsumée par la formule f. Les deux opérateurs forment une connexion de Galois entre 2^G et $\mathcal{L}$. Ils sont à l'origine de la définition de concept logique et du treillis de concepts logiques correspondant à un contexte logique donné. Un **concept logique** est une paire (A, f) telle que $\sigma(A) = f$ et $\tau(f) = A$. A et f sont respectivement l'extension et l'intension du concept. Les concepts logiques peuvent être ordonnés selon l'inclusion entre leurs extensions ou de manière équivalente selon la subsomption entre leurs intensions. L'ensemble des concepts logiques d'un contexte logique ordonnés de cette façon forme un **treillis de concepts logiques**. Le treillis de concepts logiques correspondant au contexte logique donné dans la table 2.8 (gauche) est donné dans la même figure à droite. Les nœuds du diagramme de Hasse de ce treillis sont étiquetés selon l'étiquetage réduit des concepts logiques [87].

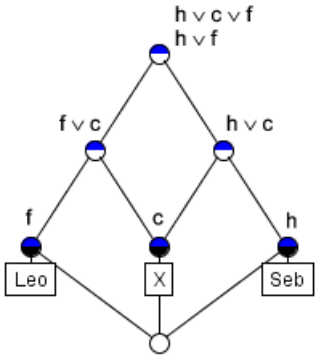
2.4.3 Analyse Formelle Généralisée

Parallèlement à l'approche de Ferré [87], Chaudron et Maille [44] proposent l'Analyse Formelle Généralisée (*Generalized Formal Concept Analysis*), une extension pour les contextes formels dans le cas de données symboliques.

Définition 7 (Contexte généralisé) *Un **contexte généralisé** est un triplet $(G, (\mathcal{L}, \leq, \sqcap, \sqcup), \delta)$ où G est un ensemble fini d'objets, $(\mathcal{L}, \leq, \sqcap, \sqcup)$ est un treillis, et d est une fonction de G dans $\mathcal{L}$.*

TABLE 2.8 – Un exemple de contexte logique (gauche) et le treillis de concepts logiques correspondant (droite). Les lettres h, f et c sont les abréviations respectives de *homme*, *femme* et *chauve*.

objet	description
Seb	h
Leo	f
X	$c \wedge (h \vee f)$



Les opérateurs de dérivation entre 2^G et $\mathcal{L}$ sont définis comme suit.

$$\phi : 2^G \to \mathcal{L}, \quad \phi_(A) \quad = \quad \sqcap_{g \in A} \delta(g) \tag{2.8}$$

$$\psi : \mathcal{L} \to 2^G, \quad \psi(B) \quad = \quad \{g \in G \mid B \leq \delta(g)\} \tag{2.9}$$

Le premier opérateur retourne la plus grande borne inférieure des représentations des objets de A. De manière duale, le second opérateur retourne pour toute représentation B les objets dont la représentation n'est pas plus générale que B. La relation d'ordre est définie comme suit : $(A_1, B_1) << (A_2, B_2) \Leftrightarrow A_1 \subseteq B_1$. Les paires (A, B) telle que $A = \psi(B)$ et $B = \phi(A)$ sont appelées concepts formels. L'AFC est un cas particulier où on a $\mathcal{L} = 2^M$, $\sqcap = \cap$, $\sqcup = \cup$ et $\leq = \subseteq$.

2.5 Différentes extensions floues de l'AFC

2.5.1 Extensions aux contextes flous

Les extensions floues de l'AFC consistent à étendre les résultats de l'AFC aux contextes flous. Un contexte flou (appelé aussi L-contexte) est décrit par (L, G, M, R) où la relation R est définie comme une application associant à tout couple objet×attribut un degré de vérité dans L pour lequel l'objet est en relation avec l'attribut ($R : G \times M \to L$). De nombreux travaux ont été menés dans le cadre de l'AFC floue [20, 183, 19, 98]. Dans [183], un objet est en relation avec un attribut s'il a un degré d'appartenance supérieur ou égal à $\alpha \in [0, 1]$. Pour la définition des inclusion, union, et intersection entre les parties de l'ensemble d'attributs, les auteurs ont repris les opérateurs classiques appliqués aux sous-ensembles flous [188].

Dans [18, 19, 98], la définition d'un contexte flou est plus générale que celle définie dans [183]. Ces approches utilisent généralement un treillis résidué $< (L, \wedge, \vee, *, \to) >$ comme structure algébrique pour le calcul des degrés. Un treillis résidué [175] est un treillis complet où $(L, *)$ est un monoïde commutatif [6] et la paire $(*, \to)$ vérifie le principe de résiduation [7]. Une telle algèbre permet de conserver les propriétés de fermeture requises par une connexion de Galois. Cette

6. Un monoïde est un ensemble muni d'une loi de composition interne associative admettant un élément neutre
7. $\forall a, b, c$ on a $a * b \geq c \Leftrightarrow a \geq b \to c$

algèbre a été étendue dans [136] de manière à considérer différentes paires de $(*, \rightarrow)$. Georgescu et Popescu [98] ont traité le cas où le monoïde $(L, *)$ est une structure non commutative. Les auteurs dans [63] montrent qu'une algèbre résiduée n'est pas absolument requise pour la conservation des propriétés inhérentes à une connexion de Galois. Les définitions des opérateurs de dérivation comportent des implications. La généralisation de ces opérateurs est ainsi fondée sur des implications floues [38]. Le choix de l'opérateur $\rightarrow$ est guidé par le respect des propriétés de fermeture de la connexion de Galois. Par exemple, les auteurs de [39] utilisent des S-implications de la forme $\neg p \vee q$ et les auteurs dans [18, 98, 136] utilisent des implications résiduées. Un concept formel dans le cas booléen classique de l'AFC est un rectangle maximal de croix. Cette propriété est préservée pour des connexions de Galois floues basées sur des paires résiduées $(*, \rightarrow)$ [19].

2.5.2 Différentes significations des degrés

Classiquement, l'AFC s'applique à une relation binaire représentée par un contexte formel. Un objet satisfait une propriété ou ne la satisfait pas. Plusieurs approches ont proposé des extensions des contextes formels en contextes flous et des extensions de la connexion de Galois pour construire le treillis complet de concepts. De même que pour les extensions de la section précédente, ces approches insistent sur l'aspect mathématique. Cependant, peu d'approches discutent de la signification des degrés introduits dans le contexte formel. Des recherches récentes [60, 61, 63, 67, 84] ont soulevé différentes questions sémantiques liées à la généralisation d'un contexte formel. Dans [60, 61], les auteurs ont traité la notion de gradualité pour les concepts formels flous et l'impact de l'incertitude sur la notion de concept formel dans le cadre de la théorie des possibilités. Ils ont également introduit la notion de typicalité pour les objets et la notion d'importance pour les propriétés. Ces deux notions n'ont pas été introduites dans l'AFC.

Un première interprétation correspond au cas où les degrés représentent à quel niveau un objet satisfait une propriété. Cette interprétation est appelé interprétation à échelle unipolaire positive et $R(m)$, qui désigne l'ensemble des objets possédant la propriété m, représente le support du sous-ensemble flou des objets satisfaisant la propriété m. Considérer la convention opposée où $R(m)$ serait le noyau du sous-ensemble flou des objets satisfaisant la propriété m revient à introduire des degrés dans les cases vides. Pour la prise en compte des contextes avec des niveaux de satisfaction imprécisément connus, Djouadi et al. [63] ont proposé une connexion de Galois ainsi qu'une généralisation de l'implication de Gödel[8] pour construire le treillis complet.

La seconde interprétation consiste à remplacer les cases ayant une croix et les cases vides par des degrés d'appartenance de $L = [0, 1]$, l'idée de base de la logique floue, et à considérer une valeur pivot (par exemple $\alpha = 0.5$) qui permet de séparer le cas où l'objet satisfait plus la propriété qu'il ne la satisfait pas. Ce type d'interprétation est appelé interprétation à échelle bipolaire.

Les interprétations positives permettent d'avoir un contexte flou à partir d'un contexte à attributs quantitatifs [179]. D'autres aspects de degrés sont traités comme l'incertitude. Dans ce cas, les propriétés peuvent rester non floues mais la certitude de la satisfaction de la propriété par un objet n'est pas complète. Elle peut être représentée par une distribution de probabilités ou de possibilités au sens de [60, 61]. Plus précisément, dans une table décrivant un contexte incertain, les cases sont remplies par des paires (α, β) où α est le degré de nécessité que l'objet aît la propriété et β est le degré de nécessité que l'objet n'aît pas la propriété avec $\min(\alpha, \beta) = 0$. Les situations $(1, 0)$ et $(0, 1)$ correspondent aux situations complètement informées où on sait avec certitude que l'objet a la propriété ou qu'il ne l'a pas. La situation $(0, 0)$ reflète l'ignorance

8. Une implication de Gödel est définie par : $I(p \rightarrow q) = q$ si $p > q$

totale et celle où $0 < \max(\alpha, \beta) < 1$ reflète l'ignorance partielle. Des extensions des contextes formels introduisant la notion d'incomplétude sont proposées dans[67, 84, 62]. Dans ces travaux, les extensions sont fondées sur la définition des opérateurs de dérivation et des connexions de Galois en utilisant les mesures définies dans la théorie des possibilités : la possibilité (Π) et son dual la nécessité (N), la suffisance (Δ) et son dual la certitude potentielle (∇). La suffisance est la mesure utilisée dans le cadre de l'AFC classique. Les opérateurs de nécessité et de possibilité sont appliqués à des contextes flous. L'intérêt de ces travaux est d'introduire une nouvelle connexion de Galois, inspirée par une lecture possibiliste de l'AFC concernant les données binaires [67, 84, 62] et pour les contextes hétérogènes [11]. Une des connexions de Galois met en évidence la décomposition du contexte en plusieurs sous-contextes indépendants. L'idée de ces travaux est de représenter la valeur d'une case dans un contexte (elle peut être binaire, numérique, ou symbolique) par une distribution de possibilités et ensuite d'appliquer les opérateurs de dérivation pour construire le treillis de concepts dans le cas où les opérateurs forment une connexion de Galois.

2.5.3 Les quatre opérateurs

Cas de contextes binaires

La relation entre objets et propriétés est définie par gRm où l'objet g possède la propriétés m. On note respectivement $R^{-1}(m)$ et $R(g)$ l'ensemble d'objets ayant la propriété m et l'ensemble des propriétés que possède l'objet g. Les opérateurs de dérivation correspondant aux mesures de la théorie des possibilités (voir Section 1.5 du Chapitre 1) sont donnés dans la suite pour tout $A \subseteq G$ et $B \subseteq M$.

- **Mesure de possibilité** [9]

$$A^{\diamond} = \{m \in M | R^{-1}(m) \cap A \neq \emptyset\} = \cup_{g \in X} R(g) \tag{2.10}$$

et son dual

$$B^{\diamond} = \{g \in G | R(x) \cap B \neq \emptyset\} = \cup_{m \in M} R^{-1}(m) \tag{2.11}$$

Le premier opérateur retourne l'ensemble des propriétés possibles appartenant à A, c-à-d les propriétés de M qu'un objet dans le sous-ensemble A satisfait. D'une manière duale, le second opérateur retourne pour un ensemble de propriétés le sous-ensemble d'objets qui possèdent au moins une propriété dans B. En utilisant les propriétés de la mesure de possibilité on a $(A_1 \cup A_2)^{\diamond} = A_1^{\diamond} \cup A_2^{\diamond}$.

- **Mesure de nécessité**

$$A^{\square} = \{m \in M | R^{-1}(m) \subseteq A\} = \cap_{g \notin A} \overline{R(g)} \tag{2.12}$$

et son dual

$$B^{\square} = \{g \in G | R(g) \subseteq B\} = \cap_{m \notin M} \overline{R^{-1}(m)} \tag{2.13}$$

Le premier opérateur représente les propriétés satisfaites par des objets de A, i.e. seuls les objets de A ont ces propriétés. Le second opérateur retourne les objets qui ne possèdent

9. Les notations utilisées dans cette partie sont utilisées dans [11]. Les notations $\diamond, \square, \triangle$ et $\bigtriangledown$ sont associées respectivement aux mesures de possibilité, nécessité, suffisance et certitude potentielle. Les auteurs dans [67, 84, 62] utilisent respectivement pour les quatre mesures les notations $(.)^{\Pi}, (.)^{N}, (.)^{\Delta}$ et $(.)^{\nabla}$.

que les propriétés de B. De plus, en utilisant les propriétés de la mesure de nécessité [78], on a : $A^{\square} = \overline{(\overline{A})^{\diamond}} = M \setminus (\overline{A})^{\diamond}$, et $(A_1 \cap A_2)^{\square} = A_1^{\square} \cap A_2^{\square}$.

– **Mesure de suffisance** [10]

$$A^{\Delta} = \{m \in M | R^{-1}(m) \supseteq A\} = \cap_{g \in A} R(g) \tag{2.14}$$

$$B^{\Delta} = \{g \in G | R(g) \supseteq B\} = \cap_{m \in B} R^{-1}(m) \tag{2.15}$$

Pour tout $A \subseteq G$, $R^{\Delta}(A)$ représente l'ensemble de toutes les propriétés communes aux objets dans A et d'une manière duale, $R^{\Delta}(B)$ retourne l'ensemble des objets ayant toutes les propriétés dans B. De plus, $R^{\Delta}(A_1 \cup A_2) = R^{\Delta}(A_1) \cap R^{\Delta}(A_2)$.

– **Mesure de certitude**

$$A^{\nabla} = \{m \in M | R^{-1}(m) \cup A \neq G\} = \cup_{g \notin A} \overline{R(g)} \tag{2.16}$$

$$B^{\nabla} = \{g \in G | R(g) \cup B \neq G\} = \cup_{g \notin B} \overline{R(g)} \tag{2.17}$$

Notons que $R^{\nabla}(A) = \overline{(\overline{A})^{\Delta}} = M \setminus (\overline{A})^{\Delta}$. Autrement dit, dans le contexte R, pour toute propriété dans A^{∇}, il existe au moins un objet qui n'est pas dans A et qui n'a pas cette propriété. De plus, on a $(A_1 \cap A_2)^{\nabla} = A_1^{\nabla} \cup A_2^{\nabla}$.

Exemple illustratif

Considérons la Table 2.9 (exemple utilisé par Dubois et al. dans [84]) où on a huit objets et neuf propriétés. La Table 2.10 montre un exemple de plusieurs sous-ensembles d'objets avec les quatre opérateurs de dérivation et la Table 2.11 montre les quatre opérateurs duaux appliqués à des sous-ensembles de propriétés.

R	*a*	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*
g_1							×		
g_2							×	×	
g_3							×	×	
g_4							×	×	×
g_5	×	×		×		×			
g_6	×	×	×	×		×			
g_7	×		×	×	×				
g_8	×		×	×		×			

TABLE 2.9 – Le contexte formel donné dans [84]

10. Pour garder la cohérence avec la mesure de suffisance de la théorie des possibilités Δ, nous gardons le symbole Δ pour les opérateurs de dérivation A^{Δ} et B^{Δ} sachant que ces opérateurs sont à la base de l'AFC classique notés pour $A \subseteq G$ et $B \subseteq M$ par A' et B' dans la Définition 2.

A	$(.)^{\diamond}$	$(.)^{\square}$	$(.)^{\Delta}$	$(.)^{\nabla}$
$\{g_1, g_2\}$	$\{g, h\}$	$\{g\}$	$\{g\}$	M
$\{g_1, g_2, g_3, g_4\}$	$\{g, h, i\}$	$\{g, h, i\}$	$\{g\}$	$\{b, c, e, f, g, h, i\}$
$\{g_5, g_6, g_7, g_8\}$	$\{a, b, c, d, e, f\}$	$\{a, b, c, d, e, f\}$	$\{a, d\}$	$\{a, b, c, d, e, f, h, i\}$

TABLE 2.10 – Exemple des sous-ensembles d'objets avec les quatre opérateurs de dérivation

B	$(.)^{\diamond}$	$(.)^{\square}$	$(.)^{\Delta}$	$(.)^{\nabla}$
$\{g\}$	$\{g_1, g_2, g_3, g_4\}$	$\{g_1\}$	$\{g_1, g_2, g_3, g_4\}$	G
$\{g, h, i\}$	$\{g_1, g_2, g_3, g_4\}$	$\{g_1, g_2, g_3, g_4\}$	$\emptyset$	G
$\{a, b, c, d, e, f\}$	$\{g_5, g_6, g_7, g_8\}$	$\{g_5, g_6, g_7, g_8\}$	$\emptyset$	G

TABLE 2.11 – Exemple des sous-ensembles de propriétés avec les quatre opérateurs de dérivation

Connexion de Galois

Dans l'AFC classique, les opérateurs utilisant la mesure de suffisance forment une connexion de Galois [96]. Une paire (A, B) telle que $A^{\Delta} = B$ et $B^{\Delta} = A$ est un concept formel dont l'extension est A et l'intension est représentée dans B. Ainsi, la paire (A, B) est un concept formel pour ∇ si et seulement si la paire $(\overline{A}, \overline{B})$ est un concept formel pour Δ, on note ainsi $A^{\nabla} = B$ et $B^{\nabla} = A$ si $\overline{A}^{\Delta} = \overline{B}$ et $\overline{B}^{\Delta} = \overline{A}$ [67, 84]. D'autre part, (A, B) est un concept formel pour $\diamond$ si et seulement si la paire $(\overline{A}, \overline{B})$ est un concept formel pour $\square$, on note ainsi $A^{\diamond} = B$ et $B^{\diamond} = A$ si $\overline{A}^{\square} = \overline{B}$ et $\overline{B}^{\square} = \overline{A}$ [67, 84, 62]. Revenons à l'exemple de la Table 2.9, la paire $(\{g_1, g_2, g_3, g_4\}, \{g\})$ est un concept formel pour l'opérateur de suffisance Δ. Ainsi, les paires $(\{g_1, g_2, g_3, g_4\}, \{g, h, i\})$ et $(\{g_5, g_6, g_7, g_8\}, \{a, b, c, d, e, f\})$ sont des paires (A, B) telles que $A^{\square} = B$ et $B^{\square} = A$, ainsi que $\overline{A}^{\square} = \overline{B}$ et $\overline{B}^{\square} = \overline{A}$. D'où la connexion de Galois formée à partir des deux opérateurs de possibilité et de nécessité. Dans ce cas, le contexte peut être décomposé en plusieurs sous-contextes dont les objets respectivement les attributs, sont des ensembles disjoints.

Si les paires (A, B) telles que $A^{\square} = B$ et $B^{\square} = A$ n'existent plus dans le contexte, alors le contexte n'est plus décomposable. Par exemple, si on modifie la relation R en R' (donnée dans la Table 2.12), les paires mentionnées n'existent plus puisque $R'^{\square}(\{g_1, g_2, g_3, g_4\}) = \{g, h\}$ et $R'^{\square}(\{g, h\}) = \{g_1, g_2, g_3\}$.

R	a	b	c	d	e	f	g	h	i
g_1							×		
g_2							×	×	
g_3							×	×	
g_4				×			×	×	×
g_5	×	×		×		×			
g_6	×	×	×	×		×			
g_7	×		×	×	×				
g_8	×		×	×		×			

TABLE 2.12 – Le contexte formel donné dans [84]

Cas de contextes hétérogènes

Dans le cas des contextes complexes et hétérogènes, les données ne sont plus binaires. Elles peuvent être numériques ou symboliques. La théorie des possibilités offre un cadre intéressant pour la représentation des données hétérogènes à l'aide des distributions de possibilités. D'autre part, nous avons vu dans la section 2.4.1 une méthode pour construire le treillis de concepts sans transformer les données. Récemment, dans [11] les auteurs ont mis en évidence une lecture possibiliste des structures de patrons en utilisant les mesures de possibilité, de nécessité, de suffisance et de certitude potentielle de la théorie des possibilités, dont l'opérateur utilisant la mesure de suffisance et son dual sont les opérateurs de dérivation proposés par Ganter et al. [94]. Ce point de vue est analogue à celui proposé dans le cadre des données binaires. Son intérêt est de représenter l'incertitude en utilisant la théorie des possibilités, et de décomposer le contexte en sous-contextes indépendants grâce à la dualité des mesures de la théorie des possibilités. La relation R n'est plus binaire, les cases du contexte contiennent des valeurs, des intervalles, des symboles, etc. Dans [11], les auteurs ont défini les quatre opérateurs de dérivation en s'inspirant de la formule donnée pour une structure de patrons $(G, (D, \sqcap), \delta)$. Ils ont étudié deux cas : numérique et qualitatif.

Dans le cas des données numériques, la description d'un objet pour une propriété est un intervalle ou un intervalle flou (informations modélisées respectivement par des distributions de possibilités dans $\{0, 1\}$ et $[0, 1]$). Par exemple, si on considère le cas des intervalles : $\sqcap$ représente l'intersection et la relation d'ordre $\sqsubseteq$ désigne l'inclusion classique des intervalles ($\subseteq$). Par conséquent, l'opérateur de dérivation utilisant la mesure de suffisance retourne pour tout sous-ensemble d'objets leur description commune. D'une manière analogue au cadre des contextes binaires, les opérateurs de dérivation se définissent en remplaçant les opérations $\sqcup$ par $\cup$ et $\sqsupseteq$ par $\supseteq$. L'opérateur de dérivation utilisant la mesure de possibilité appliqué à un sous-ensemble d'objets est similaire à l'opérateur de dérivation proposé dans [126]. Mais ce dernier utilise la plus petite enveloppe convexe de l'union des descriptions. D'autre part, l'opérateur de dérivation d'une description, dans un contexte numérique avec des intervalles, utilisant la mesure de nécessité retourne le même sous-ensemble d'objets que celui calculé à partir de l'opérateur de dérivation d'une description dans [126].

Dans le cas des données qualitatives, les informations sont représentées en utilisant la logique. Pour chaque objet, une base de connaissances est utilisée. La construction des quatre opérateurs de dérivation et de leurs duaux est presque similaire à celle des données binaires.

2.5.4 Autres extensions

Il existe encore d'autres extensions, notamment l'extension de l'AFC aux objets symboliques [28, 161]. Trois types de données ont été considérés par cette extension : les données complexes qui se présentent sous la forme d'un contexte multivalué où la relation objet×attribut peut prendre un ensemble de valeurs au lieu d'une seule, les données intervalles qui se présentent sous la forme de contexte multivalué où la relation objet×attribut prend un intervalle de valeurs et les données histogrammes qui se présentent sous la forme de contexte multivalué où la relation objet×attribut prend pour valeurs un histogramme. G. Polaillon [161] a défini de nouvelles connexions de Galois qui permettent de dériver à partir de contextes, dont la relation R est représentée par des valeurs ou des intervalles de valeurs, deux types de treillis : un treillis d'union et un treillis d'intersection. Une approche est proposée dans [111] pour les données complexes et il est similaire à l'approche de Polaillon. D'autres applications de l'AFC et adaptations de la connexion de Galois pour les contextes complexes ayant des valeurs, intervalles de valeurs ou

termes linguistiques, ont été étudiées dans [140, 137, 7, 6]. Dans ces travaux une notion de similarité est définie pour la notion de partage des valeurs des attributs par des objets. L'utilité de ces approches est d'éviter la transformation des données. une comparaison entre ces approches et les structures de patrons est présentée dans [118, 117]. Ainsi, une application des structures de patrons est proposée dans [121] pour les données numériques et a été comparée avec l'approche étudiée dans [137].

2.6 AFC et fusion d'informations

En s'appuyant sur la structure de treillis, des approches ont été développées pour la fusion des informations symboliques [134, 45, 160]. Ces approches ne sont pas liées aux théories de l'incertain. En effet, la relation d'ordre du treillis correspond à un enrichissement de l'information. Une information est automatiquement positionnée par rapport aux autres informations présentes dans le treillis. Les approches [134, 45, 160, 159] utilisent un treillis de concepts avec ses opérateurs pour combiner les informations, notamment les informations symboliques.

T. Pham a proposé, dans ses travaux de thèse [159], les "treillis d'informations", qui permet de définir les notions de contradiction, de compatibilité et de redondance entre informations. Il a proposé également différentes solutions comme le *consensus* et l'*agrégation* pour la fusion des sources ayant le même treillis d'informations. L'agrégation fournit une information plus détaillée que celles provenant des sources, tandis que le consensus fournit l'information qui est un accord général parmi les informations issues de sources. Le consensus calcule l'information commune des informations fournies par les sources et l'agrégation calcule l'information pouvant être obtenue en considérant que les informations sont complémentaires.

Chaudron et Maille [134, 45] ont proposé un modèle appelé "le modèle de Cubes", qui permet de représenter et manipuler l'information. Chaque information est représentée par un cube (une conjonction finie de littéraux) et leur méthode de fusion est définie selon deux critères : l'endogénéité et l'intégrité. Le premier critère d'endogénéité signifie que l'information résultante représente une des informations issues des sources. Le second critère d'intégrité signifie que l'information résultante contient au moins toutes les informations fournies par les sources. Des relations d'ordre partiel, de réduction des cubes et d'anti-unification sont définies sur l'espace des cubes pour construire un treillis de cubes qui représente les informations et leurs résultats de fusion. Considérons l'exemple de différents capteurs qui fournissent les informations suivantes [45] :

- e : *L'objet est un camion qui roule à* 25 *km/h*
- e' : *L'objet est une voiture qui roule.*

Le cube qui correspond l'information e : {**Type(camion),vitesse(25)**} et celui qui représente e' : {**Type(voiture),vitesse(x)**}. L'espace de fusion défini par Chaudron et Maille contient les informations fournies par les sources et leurs résultats de fusion. Il contient 8 éléments montrés sur la Figure 2.3. Plus l'opérateur de fusion est haut dans le treillis, plus son information est riche, puisque la relation d'ordre traduit l'enrichissement de l'information.

A priori, l'utilisateur choisit le cube le plus général tout en haut du treillis (le top du treillis). Mais des contraintes empêchent parfois de choisir ce cube.

Par exemple, sur la Figure 2.3, le cube représentant l'information générale est {**Type(camion), Type(voiture),vitesse(25)**}. Cette information signifie que l'objet est une voiture et un camion à la fois ce qui est impossible puisque l'utilisateur sait que cela ne correspond à rien. Par conséquent l'utilisateur est amené à choisir un cube plus bas dans le treillis. Néanmoins, après la réduction du nombre de cubes, il ne reste que les résultats les plus riches. Dans l'exemple ici, les deux cubes ayant des objets de type voiture et camion à la fois sont retirés et les plus riches sont

les cubes {**Type(camion),vitesse(25)**} et {**Type(voiture),vitesse(25)**}. L'utilisateur, pour choisir entre les deux informations résultants, devra utiliser des contraintes comme par exemple la fiabilité des capteurs, une préférence sur les capteurs.

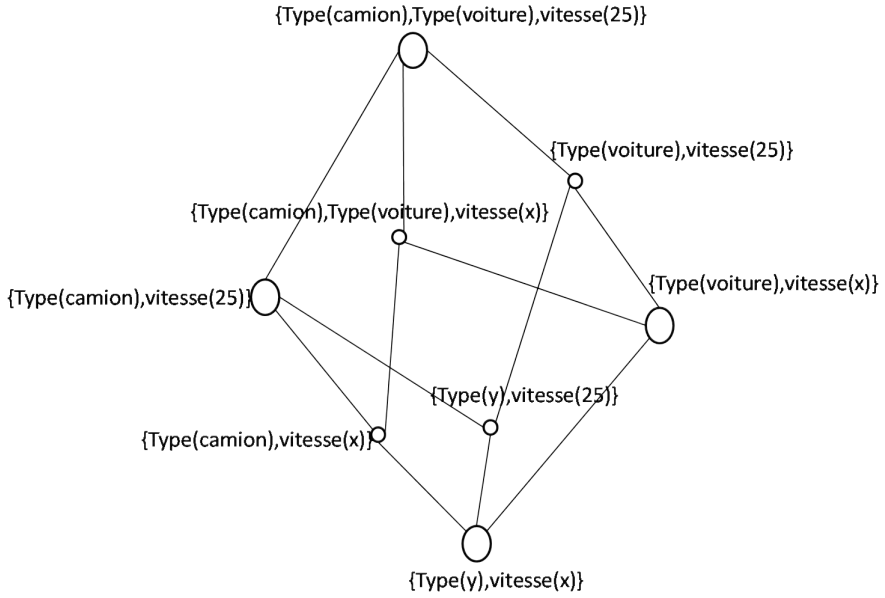


FIGURE 2.3 – Treillis des cubes obtenu à partir de l'exemple de Maille [45]

3

Fusion d'intervalles par l'analyse formelle de concepts

Sommaire

Dans ce chapitre, nous traitons le problème de la fusion d'informations modélisées par des intervalles. Nous illustrons nos exemples à partir de la Table 3.1 présentant quatre sources d'informations délivrant des valeurs imprécises pour deux variables. L'ensemble des sources joue le rôle de l'ensemble des objets dans un contexte formel (ou une structure de patrons). La relation de subsomption définie dans le chapitre précédent sera définie dans ce chapitre en respectant l'opérateur infimum choisi.

Définition 8 (Espace de fusion) *Un espace de fusion D_m, est composé des informations fournies par les sources pour la variable m, et de tous leurs résultats de fusion possibles suivant un opérateur de fusion ϕ_m.*

Par exemple, pour la variable m_1 de la Table 3.1 et ϕ_m représente l'intersection des intervalles : $D_m = \{[1,5],[4,7],[6,10],[2,3],\ [4,5],[6,7],\emptyset\}$.

	m_1	m_2
g_1	$[1,5]$	$[1,7]$
g_2	$[2,3]$	$[1,3]$
g_3	$[4,7]$	$[6,7]$
g_4	$[6,10]$	$[8,9]$

TABLE 3.1 – Intervalles fournis par les sources

3.1 Fusion conjonctive des intervalles dans le treillis de concepts pour une variable

L'opérateur de fusion conjonctif est utilisé quand toutes les sources sont fiables. Dans cette section, nous supposons l'hypothèse de fiabilité des sources.

3.1.1 L'opérateur de fusion conjonctif comme infimum

L'opérateur de fusion conjonctif appliqué aux intervalles ($\phi_m = \cap$) est une opération commutative, associative et idempotente. Elle vérifie donc les conditions d'un infimum [11] et $(D_m, \cap)$ est un inf-demi-treillis où D_m est l'espace de fusion de la variable m. Les patrons de D sont ordonnés, pour tout c et d dans D_m :

$$c \cap d = c \Leftrightarrow c \subseteq d \tag{3.1}$$

Cet ordre est une instance de la relation de subsomption définie pour les structures de patrons (Section 2.4.1). Dans l'exemple de la Table 3.1, tout intervalle est subsumé par un intervalle plus large, par exemple $[2,3] \sqsubseteq [1,5]$ puisque $[2,3] \subseteq [1,5]$. Nous avons donc $[2,3] \sqcap [1,5] = [2,3] \Leftrightarrow [2,3] \sqsubseteq [1,5]$ dans le demi-treillis qui correspond à $[2,3] \cap [1,5] = [2,3] \Leftrightarrow [2,3] \subseteq [1,5]$ en terme d'inclusion d'intervalles.

Le triplet $(G, (D_m, \cap), \delta)$ est une structure de patrons où G désigne l'ensemble des sources, $(D_m, \cap)$ est l'inf-demi-treillis et δ associe à chaque source dans G sa description dans D_m. L'infimum de deux descriptions est leur résultat de fusion, c-à-d leur intersection.

Les opérateurs de dérivation correspondant à l'opérateur conjonctif et qui forment la connexion de Galois sont :

$$\begin{array}{lcl} A^{\square} & = & \bigcap_{g \in A} \delta(g) \\ d^{\square} & = & \{g \in G | d \subseteq \delta(g)\} \end{array} \tag{3.2}$$

Le premier opérateur (Eq. 3.2) retourne pour un sous-ensemble d'objets $A \subseteq G$ leur description commune, qui est le résultat de la fusion conjonctive des éléments d'information fournis par les

11. Un infimum est une opération commutative, associative et idempotent

sources du sous-ensemble A. Le deuxième opérateur (Eq. 3.2) retourne pour toute description d l'ensemble des objets ayant une description contenant d, ce qui se traduit par l'ensemble des sources de G qui ont fourni l'information dans d.

3.1.2 Construction du treillis de concepts

Pour construire le treillis de concepts, il suffit de calculer les concepts et de les ordonner. Des nombreux algorithmes discutés dans [127] existent pour extraire et ordonner les concepts d'un contexte formel. Ainsi, ces algorithmes sont adaptés pour les structure de patrons [94, 120, 123]. Considérons l'exemple de la Table 3.1, nous avons donc $(G, (D_{m_1}, \cap), \delta)$ une structure de patrons. $(D_{m_1}, \cap)$ est l'espace de fusion de m_1 muni de l'intersection. A l'aide de la connexion de Galois, nous pouvons calculer les concepts et les ordonner. Revenons à l'exemple de la Table 3.1, les descriptions des sources g_1 et g_2 sont respectivement $\delta(g_1) = [1, 5]$ et $\delta(g_1) = [2, 3]$, et on a :

$$\begin{array}{lcl} \{g_1, g_2\}^{\square} & = & [2, 3] \cap [1, 5] = \phi_{m_1}([1, 5], [2, 3]) = [2, 3] \\ [2, 3]^{\square} & = & \{g \in G \mid [2, 3] \sqsubseteq \delta(g)\} = \{g \in G \mid [2, 3] \subseteq \delta(g)\} = \{g_1, g_2\} \end{array}$$

Puisque $\{g_1, g_2\}^{\square} = [2, 3]$ et $[2, 3]^{\square} = \{g_1, g_2\}$, la paire $(\{g_1, g_2\}, [2, 3])$ est un concept. L'ensemble des concepts ordonnés sur la Figure 3.1 représente le treillis des résultats partiels et global de la fusion des informations des sources de l'ensemble G à partir de l'exemple de la Table 3.1.

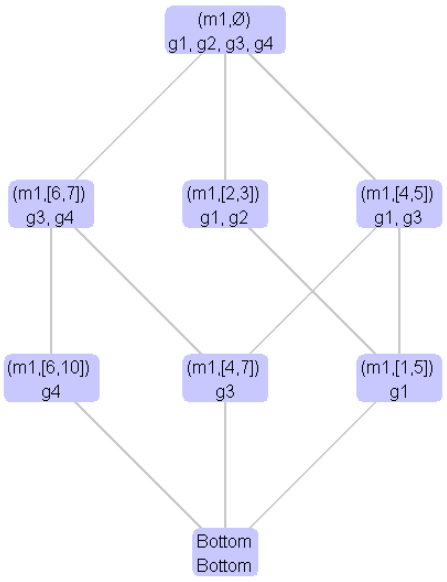


FIGURE 3.1 – Treillis des résultats de la fusion conjonctive des intervalles de la Table 3.1

3.1.3 Interprétation du treillis de concepts

Le treillis de concepts résultant de la fusion conjonctive des sources de la Table 3.1 est donné sur la Figure 3.1. Chaque nœud représentant un concept est donné par son extension et son intension. Par exemple, le concept situé en haut à droite $(\{g_1, g_3\}, (m_1, [4, 5]))$ fournit le résultat de la fusion des informations des sources de son extension (ainsi que des extensions de ses sous-concepts $(\{g_1\}, (m_1, [1, 5]))$ et $(\{g_3\}, (m_1, [4, 7])))$, $[4, 5]$ est l'information résultant de la fusion conjonctive des sources g_1 et g_3.

Le treillis des résultats conjonctifs comprend tous les sous-ensembles maximaux possibles des sources avec leurs résultats de fusion. Le concept le plus général du treillis comporte l'ensemble G avec le résultat de fusion de tous ses éléments, qui est le résultat global de la fusion conjonctive, calculé au sens des travaux [22, 79, 58, 75, 70, 81]. Plus un concept est haut dans le treillis, plus son extension est large et plus petit est l'intervalle de son intension, voire vide comme le montre l'exemple de la Table 3.1. Ainsi, l'information contenue dans un concept est plus précise et fiable que les informations de ses sous-concepts.

Considérons par exemple le concept $(A, d) = (\{g_1, g_2\}, (m_1, [2, 3]))$ de la Figure 3.1, on a :

- Son intension d fournit le résultat de fusion des informations des objets de son extension A, par exemple l'intervalle $[2, 3]$ est la fusion conjonctive (l'intersection) résultante des sources g_1 et g_2.
- Le sous-ensemble A est maximal : tout ajout d'un objet dans A conduit à un changement de d, par exemple $\{g_1, g_2\}$ est le sous-ensemble maximal des sources qui ont une intersection $[2, 3]$.
- L'extension A garde l'origine de l'information, e.g. l'information $[2, 3]$ provient des sources g_1 et g_2.
- Si A représente toutes les sources de G, c-à-d $A = G$, d désigne le résultat global de la fusion.

La navigation dans le treillis offre donc plus de flexibilité aux utilisateurs. La classification des informations, grâce à la relation d'ordre dans le treillis, montre les liens qui existent entre les informations. Par exemple, dans les travaux cités, la fusion de l'ensemble de toutes les sources correspond au concept le plus général ($\top$) dans le treillis. Ce résultat ne permet pas souvent de décider (par exemple, l'intersection est vide dans notre exemple, ce qui réduit l'utilité du résultat). Par conséquent, la navigation dans le treillis permet d'identifier des sous-ensembles d'objets et leur fusion, ainsi que de répondre à des questions du type : *"Quel est le résultat de la fusion conjonctive des sources g_1, g_2 et g_3"?*. On connaît de plus les liens entre les différents résultats possibles.

3.2 Fusion disjonctive d'intervalles pour une variable

L'opérateur de fusion disjonctif est moins contraignant que l'opérateur conjonctif. Il est utilisé quand au moins une source est fiable sans savoir laquelle. Dans cette section, nous supposons l'hypothèse de fiabilité d'une des sources sans savoir laquelle.

3.2.1 L'opérateur de fusion disjonctif comme infimum

L'opérateur de fusion disjonctif $\phi_m = \cup$ est un opérateur commutatif, associatif et idempotent. Par conséquent, l'opérateur disjonctif peut être considéré comme un infimum dans un treillis. Par conséquent, l'espace de fusion D_m obtenu à partir de l'opérateur disjonctif muni de l'opérateur disjonctif, qui est l'union des intervalles, est un inf-demi-treillis. Dans l'exemple de la Table 3.1, pour la variable m_1, l'espace de fusion est donné par $D_{m_1} = \{[1, 5], [4, 7], [6, 10], [4, 10], [1, 7], [1, 5] \cup [6, 10], [1, 10]\}$.

Donc le triplet $(G, (D_m, \cup), \delta)$ est une structure de patrons. G représente l'ensemble des objets, $(D_m, \cup)$ est l'inf.-demi-treillis des informations et δ une fonction qui associe à chaque élément de G sa description dans D_m.

Les éléments de D_m sont ordonnés, pour tout c et d, par la relation

$$c \cup d = c \Leftrightarrow c \supseteq d \tag{3.3}$$

3.2.2 Construction du treillis de concepts

Les opérateurs de dérivation sont définis pour tout sous-ensemble d'objets A et toute description d comme suit :

$$\begin{array}{lcl} A^{\square} & = & \bigcup_{g \in A} \delta(g) \\ d^{\square} & = & \{g \in G | d \supseteq \delta_(g)\} \end{array} \tag{3.4}$$

Le premier opérateur (Equ 3.4) retourne pour un sous-ensemble d'objets $A \subseteq G$ l'union de leurs descriptions qui représente le résultat de la fusion des informations des sources dans le sous-ensemble A en utilisant un opérateur disjonctif. Le second opérateur (Eq 3.4) retourne pour toute description d les objets ayant une description contenue dans d, autrement dit, les objets qui ont fourni une information contenue dans d. Pour construire le treillis de concepts, il suffit de calculer les concepts et de les ordonner.

Pour l'illustrer, considérons l'exemple de la table 3.1 et la variable m_1, nous avons alors $(G, (D_{m_1}, \cup), \delta)$ une structure de patrons. $(D_{m_1}, \cup)$ est l'espace de fusion de m_1 muni de l'union. A l'aide de la connexion de Galois, nous pouvons calculer les concepts.

Revenons à l'exemple de la Table 3.1, les descriptions des sources g_1 et g_2 sont respectivement $\delta(g_1) = [1, 5]$ et $\delta(g_1) = [2, 3]$, et on a :

$$\begin{array}{lcl} \{g_1, g_2\}^{\square} & = & [2, 3] \cup [1, 5] = \phi_{m_1}([1, 5], [2, 3]) = [1, 5] \\ [1, 5]^{\square} & = & \{g \in G \mid [1, 5] \sqsubseteq \delta(g)\} = \{g \in G \mid [1, 5] \supseteq \delta(g)\} = \{g_1, g_2\} \end{array}$$

Puisque $\{g_1, g_2\}^{\square} = [1, 5]$ et $[1, 5]^{\square} = \{g_1, g_2\}$, la paire $(\{g_1, g_2\}, [1, 5])$ est un concept formel.

L'ensemble de tous les concepts est ordonné et le résultat est donné sur la Figure 3.2.

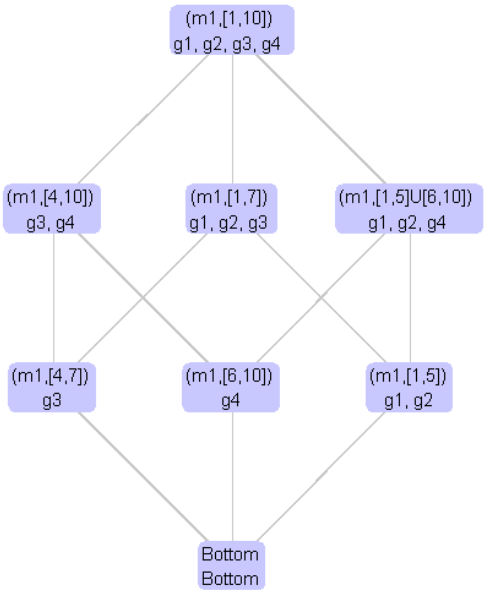


FIGURE 3.2 – Treillis de concepts des résultats de la fusion disjonctive pour la variable m_1 de la Table 3.1

3.2.3 Interprétation du treillis de concepts

Le treillis (Figure 3.2) contient les sous-ensembles maximaux de sources de G avec leurs résultats disjonctifs. Le treillis offre une hiérarchisation des résultats de la fusion disjonctive.

chaque nœud correspond à un sous-ensemble de sources (extension) et à son résultat disjonctif (intension). Par exemple, le concept $(\{g_1, g_2, g_4\}, (m_1, [1, 5] \cup [6, 10]))$ de la Figure 3.2 contient les sources g_1, g_2 et g_4 dans son extension ayant le résultat $[1, 5] \cup [6, 10]$. Le concept le plus général du treillis $(\{g_1, g_2, g_3, g_4\}, (m_1, [1, 10]))$ sur la Figure 3.2, représente le résultat calculé classiquement pour toutes les sources.

Plus les concepts sont hauts dans le treillis, plus leurs extensions des concepts contiennent plus de sources et plus leurs intensions sont larges. Ainsi, pour le treillis des résultats disjonctifs, plus on monte dans le treillis moins l'information est fiable et précise. La valeur d'une intension d'un concept est moins précise que celles des intensions de ses sous-concepts. Par exemple, le concept $\top$ est moins précis que son sous-concept $(\{g_1, g_2, g_3\}, [1, 7])$. Par conséquent, la navigation dans le treillis offre plus de flexibilité pour aider à la décision. Elle permet d'identifier un (ou plusieurs) sous-ensembles maximaux permettant de mieux agir et décider suivant le contexte de l'étude. Nous pourrons ainsi répondre aux questions comme "quelle est la fusion de g_i et g_j en utilisant un opérateur disjonctif". De plus, nous reconnaîtrons les liens entre les différents résultats possibles. Nous pouvons ainsi introduire des informations concernant les sources ou les variables pendant la construction du treillis pour éliminer des concepts qui ne sont pas utiles pour la décision.

Pour résumer, pour tout concept (A, d) de $(G, (D, \cup), \delta)$ (par exemple $(\{g_1, g_2\}, (m_1[1, 5]))$), on a :

- L'intension d fournit le résultat de la fusion des objets dans l'extension, par exemple $[1, 5]$ est le résultat de la disjonction des deux sources g_1 et g_2.
- Le sous-ensemble A est maximal, tout ajout d'un objet dans A conduit à un changement de d.
- L'extension A garde l'origine de l'information d, dans notre exemple l'information $[1, 5]$ provient de g_1 et g_2.
- Si $A = G$, le concept $\top$ désigne le résultat global de la fusion de toutes les sources.

3.3 Approche fondée sur les SMC

Comme indiqué dans la section 1.7, les opérateurs de compromis sont utilisés dans le but de gagner le maximum d'informations et d'équilibrer la fiabilité de l'information. Lorsque plusieurs sources d'informations fournissent des informations concernant une variable ou un paramètre dont la valeur exacte est inconnue, les opérateurs de combinaison convexes peuvent être critiquées et peuvent donner des valeurs que les sources initiales ne les considèrent pas comme plausibles. Les opérateurs non-adaptatifs autres que les combinaisons convexes comme les opérateurs fondés sur des moyennes pondérées (voir Yager [180, 181]) ont les mêmes inconvénients que les combinaisons convexes. Les opérateurs adaptatifs dépendent du contexte de l'étude et fournissent un résultat reflétant le conflit des informations initiales. Ils consistent à appliquer des opérateurs disjonctifs et conjonctifs. Les opérateurs adaptatifs sont étudiés dans le cas où aucune méta-information n'existe sur les sources par exemple la fiabilité des sources (fusion des opinions d'experts sur une grandeur physique).

Des opérateurs de fusion dans le cadre de la théorie des possibilités ont été introduits [155, 81, 181, 55] pour traiter le conflit parmi les informations. Dans cette section, nous nous intéressons à la fusion utilisant la notion logique de sous-ensembles maximaux cohérents (SMC). Nous détaillons comment extraire les SMC des intervalles et des distributions de N sources et la méthode de fusion pour N sources. Ensuite, nous détaillons la méthode fondée sur l'AFC pour la fusion des SMC et l'organisation des informations et de leurs résultats de fusion au sein de

treillis de concepts.

3.3.1 Définition de la méthode fondée sur les SMC

La notion des sous-ensembles maximaux cohérents est une notion originaire de la logique pour traiter l'inconsistance, et introduite par [164]. L'idée d'utilisation des SMC n'est pas nouvelle dans les théories de l'incertain. Dans la théorie des probabilités imprécises, cette notion a été étudiée par Walley [173]. Elle est aussi utilisée en pré-traitement des données [147] et le résultat de la méthode proposé dans [172] peut être vu comme une moyenne pondérée des SMC des sources. Dans le cadre de la théorie d'évidence, la notion est utilisée pour détecter des sous-ensembles maximaux cohérents des capteurs [14] et dans le contexte de la théorie des possibilités, la notion a été introduite par Dubois et al. [81] dans le cadre des informations modélisées par des intervalles réels classiques et généralisée pour les distributions de possibilités dans les travaux de Destercke [58, 55] dont le résultat de la fusion est, en général, une structure de croyance floue.

L'utilisation de la méthode de fusion fondée sur la notion de sous-ensembles maximaux cohérents permet d'avoir un maximum d'information sur l'ensemble d'informations fournies en faisant au mieux avec la fiabilité des sources [22, 79, 58, 75, 70, 81, 69].

Cette méthode consiste à fusionner au moyen d'un opérateur conjonctif des sous-ensembles de sources cohérentes entre elles, et d'utiliser un opérateur disjonctif sur ces derniers sous-ensembles. Le résultat obtenu est proche d'une conjonction si les informations sont trés cohérentes entre elles, et il est proche de la disjonction si les sources fournissent des informations incohérentes et très contradictoires.

L'approche utilisant les SMC répond à deux buts : (1) elle garde le maximum d'informativité dans le résultat et (2) elle fournit un résultat qui est cohérent avec toutes les informations fournies par les sources au départ. La méthode fondée sur les SMC est simple, attractive et elle peut s'appliquer d'une manière générale à plusieurs cadres comme montré dans les travaux de Destercke [55]. Nous rappelons d'abord comment calculer les SMC d'un ensemble d'intervalles ensuite comment calculer les SMC pour des distributions de possibilités.

3.3.2 Extraction des SMC dans les intervalles

Etant donnés N intervalles $\mathbb{I} = \{I_1, I_2, \ldots, I_N\}$, un sous-ensemble $\mathbb{K} \subseteq \mathbb{I}$ est dit maximal si $\bigcap_{i=1}^{|\mathbb{K}|} K_i \neq \emptyset$ avec $K_i \in \mathbb{K}$ et il n'existe aucun sous-ensemble non vide $\mathbb{K}' \subseteq \mathbb{I}$ tel que $\mathbb{K} \subseteq \mathbb{K}'$ avec $\bigcap_{i=1}^{|\mathbb{K}'|} K'_i \neq \emptyset$ et $K'_i \in \mathbb{K}'$.

Par exemple, dans la Table 3.1, le sous-ensemble $\mathbb{K} = \{I_1, I_2\} = \{[1,5],[2,3]\}$ est maximal pour la variable m_1 car $I_1 \cap I_2 \neq \emptyset$ et il est maximal pour la propriété d'intersection, c'est-à-dire tout ajout d'un intervalle à $\mathbb{K}$ conduit à une intersection vide.

En général, rechercher les SMC est un problème de complexité exponentielle e.g. [135]. Dubois et al. [69] ont introduit un algorithme linéaire pour trouver les SMC de N intervalles. L'algorithme 1 donné dans la suite est fondé sur l'ordre croissant des points extrêmes a_i et b_i des intervalles dans une nouvelle séquence (c_q) $q = 1, \ldots, 2N$. L'intersection maximale cohérente d'intervalles est atteinte seulement quand un élément c_q de type a (i.e. une borne inférieure d'un intervalle) est suivi par un élément c_{q+1} de type b (i.e. une borne supérieure d'un intervalle).

La méthode de fusion consiste donc à appliquer l'union aux sous-ensembles maximaux cohérents obtenus. Pour N intervalles ayant p sous-ensembles maximaux cohérents, la méthode s'écrit :

$$\hat{I} = \bigcup_{j=1,\ldots,p} \bigcap_{i=1,\ldots,|\mathbb{K}_j|} K_i \tag{3.5}$$

Algorithme 1: Sous-ensembles maximaux cohérents sur des intervalles

Entrées : n intervalles $[a_i, b_i]$

Sorties : Liste de p sous-ensembles maximaux cohérents $\mathbb{K}_j$

Liste=$\emptyset$;$j = 1$;$\mathbb{K} = \emptyset$
Mettre en ordre croissant $\{a_i, i = 1, \ldots, N\} \cup \{b_i, i = 1, \ldots, N\}$
Les renommer $\{c_q, q = 1, \ldots, 2N\}$
avec $type(c_q) = a$ si $c_q = a_k$ et $type(c_q) = b$ si $c_q = b_k$ et $k \in \{1, \ldots, N\}$
pour chaque $q = 1$ *to* $2n - 1$ **faire**
  si $type(c_q) = a$ **alors**
    Ajouter la source g_k à $\mathbb{K}$ t.q. $c_q = a_k$
    si $type(c_{q+1}) = b$ **alors**
      Ajouter $\mathbb{K}$ à Liste ($\mathbb{K}_j = \mathbb{K}$)
      $j = j + 1$
    fin
  fin
  sinon
    Enlever la source g_k de $\mathbb{K}$ t.q. $c_q = b_k$
  fin
fin

Par exemple, dans la Table 3.1, Les intervalles provenant de l'ensemble des sources de G pour la variable m_1 sont : $I_1 = [1, 5]$, $I_2 = [2, 3]$, $I_3 = [4, 7]$ et $I_4 = [6, 10]$. Suivant l'algorithme, nous avons l'ordre suivant :

$$a_1 = 1, a_2 = 2, b_2 = 3, a_3 = 4, b_1 = 5, a_4 = 6, b_3 = 7, b_4 = 10$$

L'algorithme 1 trouve le sous-ensemble $\mathbb{K} = \{I_1, I_2\}$. En croisant b_2 le sous-ensemble $\mathbb{K} = \{I_1, I_2\}$ est un sous-ensemble maximal cohérent. I_2 est retiré de la liste et nous ajoutons I_3, d'où $\mathbb{K} = \{I_1, I_3\}$ et on a ensuite b_1 donc $\mathbb{K} = \{I_1, I_3\}$ est un SMC. Enfin on enlève I_1 de la liste et on ajoute I_4, puisqu'on trouve b_3 alors le sous-ensemble est gardé et $\mathbb{K} = \{I_3, I_4\}$ est un SMC. Par conséquent, les sous-ensembles maximaux cohérents des intervalles $\{I_1, I_2, I_3, I_4\}$ de la variable m_1 sont $\{I_1, I_2\}$, $\{I_1, I_3\}$ et $\{I_3, I_4\}$, et qui correspondent respectivement aux valeurs $[2, 3]$, $[4, 5]$ et $[6, 7]$ (Figure 3.3). Donc le résultat de la fusion des SMC est donné par la disjonction des SMC recherchés dont la valeur est $\{I_1, I_2\} \cup \{I_1, I_3\} \cup \{I_3, I_4\} = [2, 3] \cup [4, 5] \cup [6, 7]$.

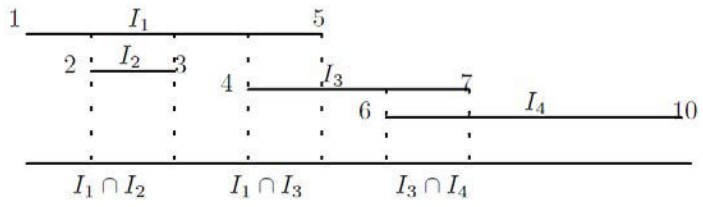


FIGURE 3.3 – Les SMC de la variable m_1 de la Table 3.1

3.3.3 L'opérateur de fusion fondée sur les SMC comme infimum

L'opérateur de fusion fondée sur les SMC est un opérateur commutatif, idempotent mais il n'est pas associatif.

Par exemple dans la Table 3.1, $f_{m_1}(f_{m_1}([1,5],[2,3]),[4,7]) = [2,3] \cup [4,7]$ et $f_{m_1}(f_{m_1}([1,5],[4,7]),[2,3]) = [2,3] \cup [4,5]$. Par conséquent, l'opérateur de fusion ne peut pas être considéré directement comme un infimum pour construire un treillis.

Cependant, d'après la définition de l'opérateur fondé sur les SMC, nous remarquons qu'on peut considérer l'union des SMC et construire un treillis à partir du contexte contenant les SMC des sources du contexte initial.

Nous allons donc effectuer un pré-traitement des données afin d'obtenir la structure de patrons contenant les SMC des informations correspondant aux sous-ensembles maximaux des sources qui les ont donnés. Ensuite, nous allons construire le treillis en considérant l'opérateur disjonctif comme un infimum.

3.3.4 Construction du treillis de concepts

Formellement, nous considérons pour une variable m le triplet $(\mathcal{O}, (F_m, \cup), \delta)$ comme une structure de patrons où l'ensemble des objets de $\mathcal{O}$ sont des sous-ensembles maximaux des objets de $\subseteq G$ qui ont fourni les SMC des intervalles de l'ensemble $\mathbb{I}$ pour une variable m. F_m est l'espace de fusion considéré à partir des informations des objets et de leur union et δ la fonction qui associe à chaque objet de $\mathcal{O}$ son information dans F_m. Considérons l'exemple de la Table 3.1, les SMC des intervalles de m_1 sont $[2,3]$, $[4,5]$ et $[6,7]$ provenant respectivement des sous-ensembles de sources $\{g_1, g_2\}$, $\{g_1, g_3\}$ et $\{g_3, g_4\}$.

Par la suite, l'ensemble $\mathcal{O}$ dans l'exemple de la variable m_1 représente l'ensemble $\{\mathbb{K}_1, \mathbb{K}_2, \mathbb{K}_3\}$ où $\mathbb{K}_1 = (g_1, g_2)$, $\mathbb{K}_2 = (g_1, g_3)$ et $\mathbb{K}_3 = (g_3, g_4)$. Les descriptions de ces objets sont $\delta((g_1, g_2)) = [2,3]$ (qui signifie que l'intervalle $[2,3]$ est attaché aux sources g_1 et g_2), $\delta((g_1, g_3)) = [4,5]$ et $\delta((g_3, g_4)) = [6,7]$. Par conséquent, l'espace de fusion peut se définir à partir de ces informations. L'espace de fusion est $F_m = \{[2,3], [4,5], [6,7], [2,3] \cup [4,5], [2,3] \cup [6,7], [4,5] \cup [6,7]\}$, où les éléments sont ordonnés par l'inclusion des intervalles comme détaillé dans la section 3.2. La structure de patrons résultante, pour la variable m_1, est donnée dans la Table 3.2.

	m_1
$\mathbb{K}_1 = (g_1, g_2)$	$[2,3]$
$\mathbb{K}_2 = (g_1, g_3)$	$[4,5]$
$\mathbb{K}_3 = (g_3, g_4)$	$[6,7]$

TABLE 3.2 – La structure de patrons résultant du pré-traitement pour la variable m_1 de la Table 3.1

Le treillis résultant de l'application de l'union à la structure de patrons donnée dans la Table 3.2 est donné sur la Figure 3.4.

3.3.5 Interprétation du treillis de concepts

Le treillis de concepts obtenu fournit une classification des sources avec leurs résultats de fusion en utilisant une union sur les SMC des sources, ce qui se traduit par la fusion fondée sur les SMC pour un sous-ensemble de sources donné.

Sur la Figure 3.4, chaque concept est donné par son extension et son intension. La première ligne des concepts juste au dessus du concept "bottom" ($\perp$) désigne les SMC recherchés dans le contexte initial, ici on a les intervalles $[2,3]$, $[4,5]$ et $[6,7]$. Plus haut dans le treillis on trouve les résultats partiels obtenus en utilisant une union sur ces derniers et le concept le plus général

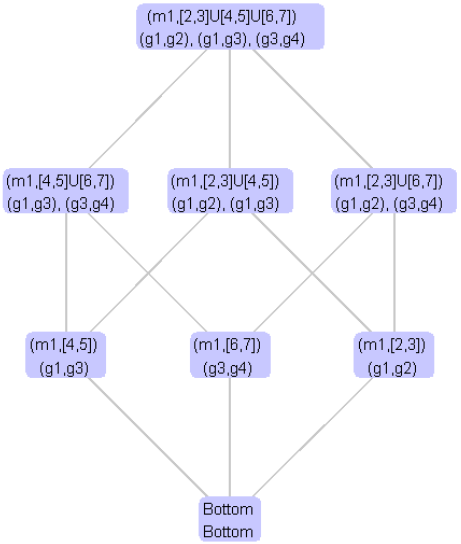


FIGURE 3.4 – Treillis de concepts des SMC pour la variable m_1 de la Table 3.1

du treillis désigne le résultat de la fusion globale de toutes les sources dans G, et constitue le résultat de la fusion en utilisant la méthode fondée sur les SMC des sources sur l'ensemble G.

Par exemple sur la Figure 3.4, le concept $\{((g_1, g_2), (g_1, g_3)), [2, 3] \cup [4, 5]\}$ présente les valeurs de m_1 suivant les objets $\mathbb{K}_1 = (g_1, g_2)$ et $\mathbb{K}_2 = (g_1, g_3)$. Ce qui signifie que l'information $[2, 3] \cup [4, 5]$ est fournie par les sources "g_1 et g_2" ou "g_1 et g_3". Notons qu'un objet du contexte initial (ici la Table 3.1) peut appartenir à plusieurs concepts qui se situent après le "bottom", à la différence des treillis obtenus avec la conjonction et la disjonction, puisqu'il peut être consistant avec un ou plusieurs objets de ces concepts. Par exemple l'objet g_1 appartient aux extensions des deux concepts $\big((g_1, g_2), [2, 3]\big)$ et $\big((g_1, g_3), [4, 5]\big)$ sur la Figure 3.4.

De plus, le treillis permet d'obtenir les sous-ensembles maximaux des sources de G qui ont fourni l'information contenue dans les intensions des concepts à partir des objets des extensions. Par exemple, les valeurs $[2, 3] \cup [4, 5]$ représentent le résultat de la fusion du sous-ensemble d'objets $\{g_1, g_2, g_3\}$ en utilisant l'opérateur fondé sur les SMC. Par conséquent, le concept $\top$, qui correspond à l'union de tous les SMC, est le résultat global de la fusion fondée sur les SMC appliquée à toutes les sources.

Ainsi, pour les concepts situés juste au-dessus du concept ($\perp$) (la première ligne des concepts dans le treillis), le sous-ensemble maximal des sources contient les sources données par leurs extensions. Pour les autres concepts, nous considérons l'union des sous-ensembles d'objets des extensions de ses sous-concepts si les extensions ont au moins une source en commun. Revenons à la Figure 3.4 et considérons le concept $\big((g_1, g_2), (g_1, g_3), [2, 3] \cup [4, 5]\big)$: les valeurs $[2, 3] \cup [4, 5]$ proviennent de $\{g_1, g_2\} \cup \{g_1, g_3\} = \{g_1, g_2, g_3\}$. Mais, dans le cas où les objets des extensions n'ont aucune intersection, le sous-ensemble maximal de sources est l'union des objets en excluant l'objet cohérent avec au moins un objet de chaque concept. Par exemple sur la Figure 3.4, considérons le concept $\big(((g_1, g_2), (g_3, g_4)), [2, 3] \cup [6, 7]\big)$, le sous-ensemble maximal des sources fournissant les valeurs $[2, 3] \cup [6, 7]$ est $\{g_2, g_3, g_4\}$ puisque l'objet g_1 est cohérent avec g_2 d'une part et avec l'objet g_3 d'autre part. La Figure 3.5 illustre le même treillis de la Figure 3.4 mais en calculant les sous-ensembles de sources de G fournissant les valeurs.

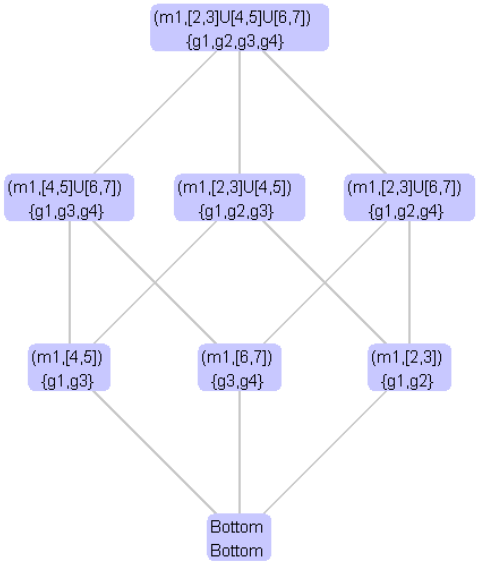


FIGURE 3.5 – Les sous-ensembles des sources fournissant les SMC de m_1 de la Table 3.1

Nous remarquons sur la Figure 3.5 que la méthode utilisée pour la construction des treillis des résultats de la fusion des SMC ne contient pas tous les sous-ensembles de sources de G, à la différence des treillis des résultats conjonctifs et disjonctifs qui considèrent tous les sous-ensembles maximaux des sources et leurs résultats de fusion. La méthode de pré-traitement utilisée pour le treillis fondé sur les SMC considère les SMC maximaux des sources et leurs résultats de fusion. Cependant, il ne contient que les SMC maximaux des sources de G. Par exemple, nous ne trouvons pas le sous-ensemble $\{g_1, g_4\}$ sur le treillis de concepts même si l'intersection de g_1 et g_4 est vide puisque $\{g_1, g_4\} \subseteq \{g_1, g_3, g_4\}$. Ceci est dû à la non-associativité de l'opérateur de fusion des SMC et à l'utilisation de l'union au sein des SMC des intervalles initiaux et non directement sur ces derniers eux-mêmes. Néanmoins, le treillis nous permet de garder l'origine de l'information et offre plus de flexibilité aux les analystes et aux utilisateurs pour le choix d'un sous-ensemble de sources dans plusieurs champs d'application quand il s'agit d'un problème de fusion de données hétérogènes et (trop) contradictoires ayant un résultat de fusion qui n'est pas utile pour la décision.

3.4 Introduction de similarités

Si nous considérons la taille des intervalles dans les treillis conjonctifs et disjonctifs, nous remarquons que la taille du premier treillis de conjonction (intervalles et distributions) est en général limitée, et que si le nombre de sources augmente et si les sources donnent des valeurs incohérentes (distinctes) alors la conjonction est vide et le nombre de concepts du treillis ne pose pas de problème du point de vue de la classification, dont le but est de regrouper des objets ayant des valeurs proches. Mais en considérant le treillis obtenu dans le cas de la fusion disjonctive (ainsi que dans le cas de la fusion fondée sur les SMC puisque c'est une union), nous remarquons que les intervalles les plus hauts sont les plus larges (en prenant pour la taille d'un intervalle l'écart entre ses deux bornes inférieure et supérieure, c-à-d la différence entre sa borne supérieure et sa borne inférieure). Ainsi, ces concepts plus haut placés possèdent plus d'objets

dans leur extensions. Notre but dans ce travail est de réduire au maximum l'incertitude des variables d'entrées dans un domaine d'application (les données agronomiques) que nous allons détailler plus tard. Les concepts les plus bas dans le treillis contiennent peu de sources mais avec des informations spécifiques en comparaison avec les informations plus haut. Si le nombre des sources (fournissant des valeurs incohérentes (au moins disjonctifs)) augmente, alors le nombre de concepts explose puisque tous les résultats possibles partiels et global de la fusion disjonctive sont calculés.

Une des manières de réduire le nombre de concepts et de faciliter le choix des utilisateurs parmi les résultats de fusion est de ne considérer que les concepts dont l'intension n'est pas plus large qu'un seuil donné, ce qui revient à filtrer les concepts à l'aide de contraintes.

Pour un concept ayant pour intension la description $d = [a, b]$, on conserve le concept si $|b - a| \leq \theta_s$, où θ_s est un seuil fixé par rapport au contexte de l'étude, par expertise, et suivant la nature de la variable. Cette contrainte nous permet d'éliminer tous les super-concepts d'un concept qui ne la vérifie pas puisqu'elle est monotone. Plusieurs travaux ont été développés dans le cadre des classifications de données numériques en utilisant l'analyse formelle de concepts sans aucune transformation de données (voir [140, 121, 7, 116, 118, 119]) et ont proposé des méthodes fondées sur la définition de notion de similarité entre les objets et les valeurs des attributs dans un contexte numériques.

Dans ce travail, nous allons utiliser une notion de similarité fondée sur une distance entre les bornes inférieure et supérieure des intervalles pour pouvoir les regrouper. En effet, les résultats de la fusion (disjonctive) sont généralement non convexes, l'intervalle pouvant alors être large et inutile pour la décision dans certaines situations. Par conséquent le fait d'utiliser cette notion évite de regrouper (fusionner) des objets qui n'ont pas des valeurs similaires par rapport à un seuil donné. On définit ainsi la notion de similarité entre sources et le seuil de similarité pour chaque attribut.

3.4.1 Similarité entre valeurs

Deux valeurs d'un même attribut sont similaires s'il n'y a pas une grande différence entre elles. Par exemple, considérons les valeurs suivantes pour une variable m : 10, 15, 22 et 25. La différence maximale (la différence entre la valeur maximale et la valeur minimale) obtenue vaut 15. En considérant cette valeur maximale, toutes les valeurs se regroupent ensemble et elles sont similaires.

Définition 9 (Seuil de similarité) *Un seuil de similarité (seuil de variation) est la variation maximale autorisée entre deux valeurs similaires. Pour toutes deux valeurs a_i et a_j pour une variable m, on dit que a_i est similaire à a_j ssi :*

$$|a_j - a_i| \leq \theta_m \tag{3.6}$$

Par exemple, si nous considérons la valeur $\theta_m = 6$, les valeurs 10 et 15 de la variable m sont similaires puisque $15 - 10 = 5 < 6$, mais les valeurs 22 et 25 ne le sont pas puisque leur différence excède la valeur du seuil θ_m.

Lorsque les objets décrivent les variables par des intervalles, la définition de la similarité entre les deux objets sera définie comme étant la variation maximale autorisée entre les valeurs du plus petit intervalle fermé qui contient ces intervalles.

3.4.2 Intervalles similaires

Pour un seuil de similarité θ_m, et deux intervalles $[a_i, b_i]$ et $[a_j, b_j]$, ces deux intervalles sont dits similaires si et seulement si :

$$|\max(b_i, b_j) - \min(a_i, a_j)| \leq \theta_m \tag{3.7}$$

Revenons à l'exemple de la Table 3.1, nous avons $\delta(g_1) = [1, 5]$, $\delta(g_3) = [4, 7]$ et $\delta(g_4) = [6, 10]$. Considérons le seuil $\theta_{m_1} = 6$, alors $\delta(g_1)$ est similaire à $\delta(g_3)$ mais elle n'est pas similaire à $\delta(g_4)$. Par conséquent, les deux objets g_1 et g_3 sont similaires mais les objets g_1 et g_4 ne le sont pas.

La relation de similarité n'est pas transitive, c-à-d si a est similaire à b et b est similaire à c, on n'a pas forcément a similaire à c. Considérons les trois objets g_1, g_3 et g_4, g_3 est similaire aux deux objets g_1 et g_4 mais g_1 et g_4 ne sont pas similaires puisqu'ils ne vérifient pas la définition de la similarité.

Le choix d'un seuil de similarité θ_m pour une variable m permet de juger si deux objets sont similaires (c-à-d ils ont des valeurs similaires) pour la variable m.

Considérer deux objets similaires implique la définition de deux objets différents (non similaires) d'où l'introduction d'un élément $*$ qui désigne la non similarité de deux objets, ce qui se traduit par le rejet du résultat de la fusion dans le cas des informations qui ne sont pas similaires. L'élément $*$ est subsumé par tout intervalle d dans le demi-treillis d'intervalles :

$$* \sqcap d = * \Leftrightarrow * \sqsubseteq d$$

Deux intervalles c et d sont dissimilaires si et seulement si $c \sqcap d = *$.

Donc, pour un seuil de similarité $\theta \in \mathbb{R}$, et deux intervalles $[a_i, b_i]$ et $[a_j, b_j]$, le résultat de leur infimum contraint par θ est donné par :

$$[a_i, b_i] \sqcap_\theta [a_j, b_j]) = \begin{cases} [\min(a_i, a_j), \max(b_i, b_j)] & \text{si } |\max(b_i, b_j) - \min(a_i, a_j)| \leq \theta_m \\ * & \text{sinon} \end{cases} \tag{3.8}$$

Néanmoins, l'ensemble des objets n'est pas maximal et il arrive des cas où on regroupe des objets qui ne sont pas similaires entre eux. Pour calculer un ensemble d'objets valide à regrouper, Ganter et Kuznetsov [94] ont introduit la notion de projection. Une projection ψ est une application de D dans D (D est l'inf-demi-treillis) qui associe à tout patron $d \in D$ un nouveau patron plus général $\psi(d)$ avec $\psi(d) \sqsupseteq d$, ce qui signifie dans un contexte numérique que l'on remplace tout intervalle d par un intervalle $\psi(d)$ plus large ($\psi(d) \supseteq d$).

Pour calculer la projection de chaque patron, il suffit de trouver les patrons qui lui sont similaires, puis de retirer les patrons qui ne sont pas similaires entre eux enfin de calculer leur infimum.

3.4.3 Fusion d'informations et similarité

Comme mentionné dans les sections précédentes, les résultats de la fusion (notamment) disjonctive sont nombreux, ce qui rend le travail d'un analyste plus difficile. L'approche décrite dans la section 3.4.2 est donc particulièrement intéressante dans le cas des contextes volumineux. Cette approche permet de réduire la taille du treillis de concepts des résultats disjonctifs en éliminant un nombre des concepts qui ne semble pas utiles pour l'utilisateur. Dans cette partie, nous nous intéressons à la fusion disjonctive des intervalles (voir section 3.2). En introduisant le seuil θ_m,

l'opérateur infimum qui représente l'opérateur de fusion disjonctif défini dans la section 3.2 est défini pour une variable m comme suit :

$$\phi_m([a_i, b_i], [a_j, b_j]) = \begin{cases} [\min(a_i, a_i), \max(b_i, b_j)] & \text{si } |\max(b_i, b_j) - \min(a_i, a_j)| \leq \theta_m \\ * & \text{sinon} \end{cases} \quad (3.9)$$

Ainsi le résultat de la fusion de deux intervalles similaires est la plus petite enveloppe convexe qui les contient. Par conséquent, les résultats de la fusion organisés dans le treillis obtenu à partir d'un contexte sont convexes. Le seuil θ permet donc de contrôler la convexification des résultats. Ainsi, cette approche ne permet que la fusion des informations qui sont similaires par rapport au seuil θ choisi et elle ne garde que les concepts ayant des sources similaires deux à deux. Les tailles de tous les intervalles résultants ne dépassent pas le seuil θ. Considérons l'exemple de la variable m donné dans la Table 3.3 où il y a douze objets et une variable. Le treillis des résultats disjonctifs correspondant à l'exemple de la Table 3.3 contient 90 concepts. Pour un seuil $\theta = 10$, le treillis obtenu est donné sur la Figure 3.6. Le nombre de concepts est réduit ainsi que les résultats de la fusion sont convexes.

	m
g_1	[1,5]
g_2	[2,3]
g_3	[1,3]
g_4	[4,7]
g_5	[6,7]
g_6	[16,20]
g_7	[8,12]
g_8	[4,14]
g_9	[11,13]
g_{10}	[15,20]
g_{11}	[1,7]
g_{12}	[10,18]

TABLE 3.3 – Information donnée pour une variable m

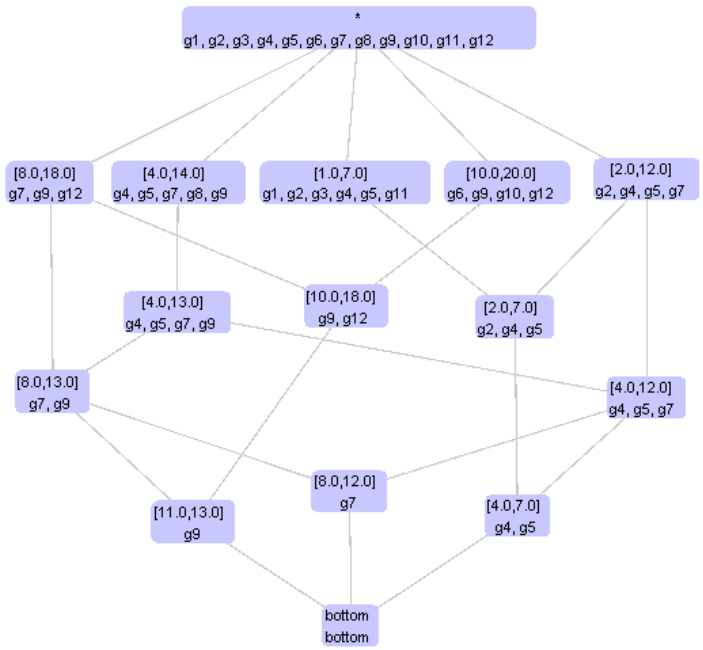


FIGURE 3.6 – Le treillis de concepts correspondant au contexte de la Table 3.3

3.4.4 Variation de la précision dans le treillis

La variation du seuil θ peut entraîner l'apparition et/ou la disparition de concepts dans le treillis de concepts obtenu à partir d'un contexte numérique. Une variation se traduit par un changement de la contrainte à vérifier par les valeurs des variables du contexte. L'augmentation de la valeur du seuil θ entraîne un relâchement et le nombre de concepts augmente puisque de nouveaux groupements apparaissent. À l'inverse, si la valeur du seuil θ diminue, moins d'objets sont similaires puisque la contrainte est renforcée à cause de la diminution de la valeur du seuil.

Par exemple, dans la Table 3.3, $\delta(g_6) = [16, 20]$ et $\delta(g_{12}) = [10, 18]$. Pour $\theta_m = 9$ les deux objets g_6 et g_{12} sont similaires et ils apparaissent ensemble dans le treillis de concepts obtenu à partir du contexte pour le seuil $\theta_m = 9$. Cependant, si on diminue le seuil $\theta_m = 8$ les objets g_6 et g_{12} ne sont plus similaires. Ainsi des liens entre les concepts disparaissent et d'autres apparaissent comme le montrent les treillis de la Table 3.4.

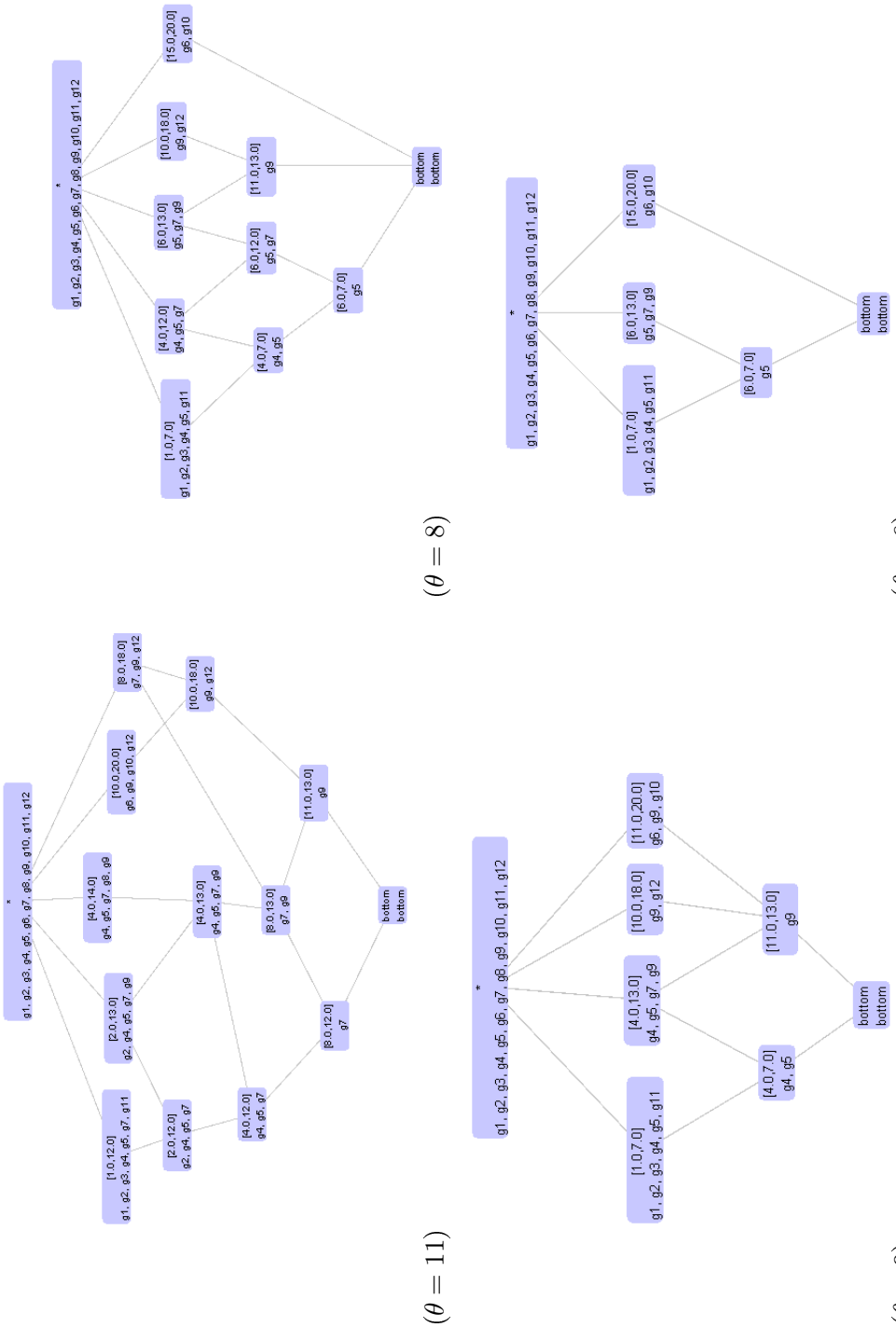


TABLE 3.4 – Les treillis de concepts des résultats disjonctifs convexes calculés à partir de la Table 3.3 pour les seuils $\theta = 11$, $\theta = 8$, $\theta = 9$ et $\theta = 6$.

3.5 Fusion pour plusieurs variables dans le treillis

3.5.1 Conjonction et disjonction

Dans la Section 2.4.1, nous avons montré comment traiter plusieurs variables dans une structure de patrons et comment construire le treillis. Dans cette section, nous considérons que les sources fournissent des valeurs pour différentes variables. Ceci peut être intéressant pour calculer le résultat de la fusion pour toutes les variables simultanément.

Par exemple, la Table 3.1 présente les objets décrits par des *vecteurs d'intervalles*. Nous considérons un *vecteur d'intervalles* à p dimensions, composé de p intervalles où chaque intervalle correspond à une variable unique, par exemple la description de l'objet g_1 est $\delta(g_1) = \langle [1,5],[1,9] \rangle$.

Pour formaliser une structure de patrons dans le cas de la fusion d'informations, il suffit de définir un infimum, i.e. un opérateur de fusion, pour chaque dimension (variable). L'opérateur de fusion s'applique sur deux ou plusieurs vecteurs. Les vecteurs doivent être de la même taille et les variables sont supposées indépendantes. On suppose un ordre canonique sur les dimensions des vecteurs.

La fusion des vecteurs d'intervalles est définie comme étant la fusion entre les intervalles de chaque dimension et en utilisant la relation d'ordre partiel des descriptions des objets définie dans [120] (voir Section 2.4.1).

Le triplet $(G, (D, \sqcap), \delta)$ détermine une structure de patrons, où G est un ensemble des sources, $(D, \sqcap)$ est un inf-demi-treillis de patrons où D est l'union des espaces de fusion de toutes les variables et δ est la fonction qui associe à tout $g \in G$ sa description dans D. Notons que $\sqcap$ représente l'opérateur de fusion qui peut être choisi pour chaque dimension.

Revenons à la Table 3.1 et considérons les deux objets g_1 et g_2, $\sqcap$ représente l'intersection d'intervalles. Le résultat de la fusion conjonctive pour les variables m_1 et m_2 est respectivement $[2,3]$ et $[1,3]$ puisque $\langle [1,5],[1,7] \rangle \cap \langle [2,3],[1,3] \rangle = \langle [1,5] \cap [2,3],[1,7] \cap [1,3] \rangle = \langle [2,3],[1,3] \rangle$.

Le treillis résultant du traitement de plusieurs variables permet d'avoir simultanément le résultat de la fusion de ces variables. La Figure 3.7 illustre un exemple de traitement de plusieurs variables pour la Table 3.1. Considérons par exemple le concept $(\{g_1, g_2\}, ([2,3],[1,3]))$, le résultat de la fusion conjonctive des informations provenant des sources g_1 et g_2 pour la variable m_1 (respectivement pour m_2) est $[2,3]$ (respectivement $[1,3]$ pour m_2). Le concept général du treillis ($\top$) représente le résultat conjonctif pour les deux variables (qui est vide dans notre exemple).

Nous pouvons également opérer une fusion simultanée pour plusieurs variables dans le cas de l'information modélisée par des distributions de possibilités. Il suffit de définir les vecteurs d'intervalles et de choisir un opérateur de fusion. Notons que nous pouvons utiliser des opérateurs de fusion différents pour chaque dimension : par exemple, dans la Table 3.1, en considérant l'opérateur conjonctif pour m_1 et l'opérateur disjonctif pour m_2.

3.5.2 Treillis des SMC d'intervalles pour plusieurs variables

Dans le cas où les sources d'informations fournissent des informations pour plusieurs variables comme dans la Table 3.1, nous utilisons la méthode de pré-traitement des données détaillée dans la section 3.3.3 pour construire le treillis de concepts. Le treillis obtenu permet dans le cadre de la fusion d'informations d'obtenir les résultats de la fusion pour plusieurs variables simultanément en considérant que les variables sont indépendantes.

Nous considérons donc d'abord les SMC des variables et nous construisons ensuite la structure de patrons telle que les descriptions des objets sont les valeurs des SMC pour la variable et vide sinon.

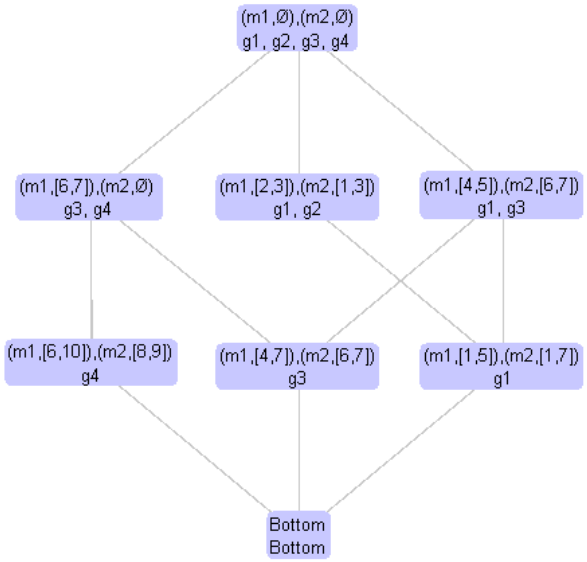


FIGURE 3.7 – Le treillis d'intersection pour les deux variables m_1 et m_2 de la Table 3.1

Considérons l'exemple de la Table 3.1 avec deux variables m_1 et m_2. Les SMC des intervalles de la variable m_1 sont $[2,3]$, $[4,5]$ et $[6,7]$ donnés par les sous-ensembles de sources $\{g_1,g_2\}$, $\{g_1,g_3\}$ et $\{g_3,g_4\}$. Les SMC des intervalles de la variable m_2 sont $[1,3]$ et $[6,7]$ provenant des sous-ensembles $\{g_1,g_2\}$ et $\{g_1,g_3\}$. Nous remarquons que les sous-ensembles $\{g_1,g_2\}$ et $\{g_1,g_3\}$ sont des SMC pour m_1 et m_2, tandis que le sous-ensemble $\{g_3,g_4\}$ est un SMC pour m_1 mais pas pour m_2. Par conséquent l'ensemble d'objets de la structure de patrons des SMC est tel que $\mathcal{O} = \{(g_1,g_2),(g_1,g_3),(g_3,g_4)\}$. Pour les descriptions des objets dans $\mathcal{O}$, seul le sous-ensemble $\{g_3,g_4\}$ n'est pas un SMC pour m_2 puisqu'il ne fournit pas d'information pour m_2 donc $\delta(g_3,g_4) = \emptyset$ et la structure de patrons résultante est donnée dans la Table 3.5.

	m_1	m_2
(g_1,g_2)	$[2,3]$	$[1,3]$
(g_1,g_3)	$[4,5]$	$[6,7]$
(g_3,g_4)	$[6,7]$	$\emptyset$
(g_4)	$\emptyset$	$[8,9]$

TABLE 3.5 – Table résultant du pré-traitement des deux variables

Le treillis de concepts résultant pour deux variables est donné sur la Figure 3.8. Les extensions des concepts sont données pour chaque concept. Les intensions sont données dans la Table 3.6. Nous trouvons les résultats de la fusion des SMC pour les deux variables simultanément. Le concept le plus général du treillis représente le résultat de la fusion des SMC des deux variables m_1 et m_2.

Les sous-ensembles maximaux cohérents des sources de G (les origines des SMC) sont donnés sur le diagramme de la Figure 3.9. Comme dans le cas d'une seule variable, nous remarquons sur le treillis de concepts de la Figure 3.9 qu'il ne contient pas tous les sous-ensembles possibles de sources de G pour chacune des variables m_1 et m_2. En effet, pour calculer le sous-ensemble maximal de sources de G qui a fourni l'information dans l'intension des concepts du treillis, il

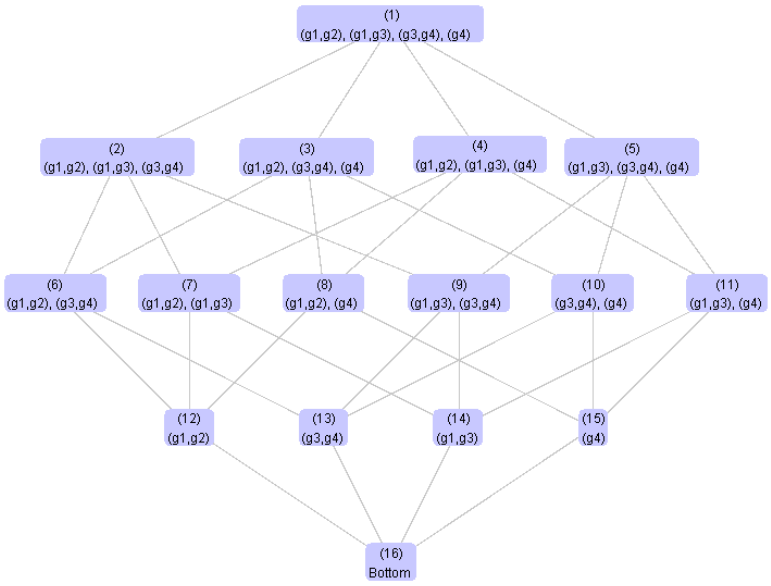


FIGURE 3.8 – Treillis de concepts des SMC pour deux variables

suffit de faire l'union des objets de l'extension pour une même variable si leur intersection n'est pas vide. Sinon, l'union des objets de l'extension des concepts est à considérer en excluant les sources de G qui sont cohérentes avec au moins une source appartenant aux objets de $\mathcal{O}$ dans le treillis.

Considérons par exemple, le concept marqué (7) sur le treillis de la Figure 3.8 ayant comme extension le sous-ensemble d'objets $\{(g_1, g_2), (g_1, g_3)\}$ et comme intension $([2,3] \cup [4,5], [1,3] \cup [6,7])$ (voir la Table 3.6). Par conséquent le sous-ensemble maximal de sources de G, pour les deux variables m_1 et m_2, qui ont fourni les SMC de l'intension est $\{g_1, g_2\} \cup \{g_1, g_3\} = \{g_1, g_2, g_3\}$ puisque $\{g_1, g_2\}$ et $\{g_1, g_3\}$ ont une source en commun (qui est g_1).

Si nous considérons le concept (3) sur la Figure 3.8 $\big(\{(g_1, g_2), (g_3, g_4), g_4\}, [2,3] \cup [6,7], [1,3] \cup [8,9]\big)$:

- Pour la variable m_1, les sous-ensembles de sources $\{g_1, g_2\}$ et $\{(g_3, g_4\}$ n'ont pas d'intersection. De plus, la source g_1 du sous-ensemble $\{g_1, g_2\}$ est cohérente avec la source g_3 du sous-ensemble $\{(g_3, g_4\}$. Par conséquent, le sous-ensemble maximal de sources de G qui fournit le SMC de l'intension pour m_1 est $\{\{g_1, g_2\} \cup \{g_3, g_4\}\} \backslash \{g_1\} = \{g_2, g_3, g_4\}$.
- Pour la variable m_2, les sous-ensembles de sources $\{g_1, g_2\}$ et $\{(g_4)\}$ n'ont pas d'intersection, aucun des sources du sous-ensemble $\{g_1, g_2\}$ n'est cohérente avec au moins une des sources du sous-ensemble $\{(g_4)\}$, donc l'union des deux sous-ensembles est à considérer et le sous-ensemble maximal de sources est $\{g_1, g_2, g_4\}$.

Numéro de concept	intension
(1)	$(m_1, [2,3] \cup [4,5] \cup [6,7]), (m_2, [1,3] \cup [6,7] \cup [8,9])$
(2)	$(m_1, [2,3] \cup [4,5] \cup [6,7]), (m_2, [1,3] \cup [6,7])$
(3)	$(m_1, [2,3] \cup [6,7]), (m_2, [1,3] \cup [8,9])$
(4)	$(m_1, [2,3] \cup [4,5]), (m_2, [1,3] \cup [6,7] \cup [8,9])$
(5)	$(m_1, [4,5] \cup [6,7]), (m_2, [6,7] \cup [8,9])$
(6)	$(m_1, [2,3] \cup [6,7]), (m_2, [1,3])$
(7)	$(m_1, [2,3] \cup [4,5]), (m_2, [1,3] \cup [6,7])$
(8)	$(m_1, [2,3]), (m_2, [1,3] \cup [8,9])$
(9)	$(m_1, [4,5] \cup [6,7]), (m_2, [6,7])$
(10)	$(m_1, [6,7]), (m_2, [8,9])$
(11)	$(m_1, [4,5]), (m_2, [6,7] \cup [8,9])$
(12)	$(m_1, [2,3]), (m_2, [1,3])$
(13)	$(m_1, [6,7]), (m_2, \emptyset)$
(14)	$(m_1, [4,5]), (m_2, [6,7])$
(15)	$(m_1, \emptyset), (m_2, [8,9])$
(16)	$(m_1, \emptyset), (m_2, \emptyset)$

TABLE 3.6 – Intensions des concepts du treillis des SMC des deux variables m_1 et m_2 de la Table 3.1

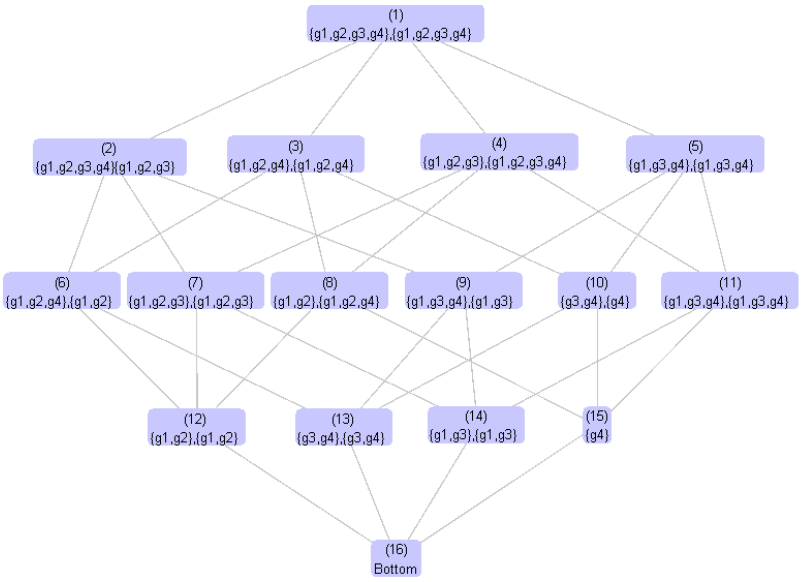


FIGURE 3.9 – Les sous-ensembles de G fournissant les SMC pour les deux variables m_1 et m_2

4

Fusion de distributions de possibilité

Sommaire

Dans ce chapitre, nous traitons le problème de la fusion d'informations modélisées par des intervalles flous. Nous illustrons nos exemples à partir de la Table 4.1 où il y a quatre sources d'informations délivrant des valeurs imprécises pour deux variables modélisées avec une distribution de possibilités dans $[0, 1]$. Une distribution de possibilités sera notée, pour une variable m et une source $g \in G$, par $\pi_{m(g)}$. Nous considérons la notation suivante pour les intervalles emboîtés : pour tous les intervalles I et J de $\mathbb{R}$, on écrit $(\lceil I; J \rceil, [\theta, \gamma]) = \{[x, y] \subseteq \mathbb{R} | I \supseteq [x, y] \supseteq J\}$ entre deux niveaux θ et γ, ainsi une distribution de possibilités est un ensemble d'intervalles emboîtés qui sont ses α-coupes, on écrit $\langle [1, 5], [2, 3] \rangle = (\lceil [1, 5]; [2, 3] \rceil, [0, 1]) = \{[x, y] \subseteq \mathbb{R} | [1, 5] \supseteq [x, y] \supseteq [2, 3]\}$ entre l'α-coupe $\theta = 0$ et l'α-coupe $\gamma = 1$. Avant d'introduire la notion d'espace de fusion de distributions, nous rappelons comment calculer les valeurs d'intersection des sources. Destercke et al. [58] ont introduit un algorithme qui calcule les valeurs β_q. Ces dernières correspondent aux hauteurs des distributions $\min(\pi_i, \pi_j)$ pour toutes les paires de distributions π_i et π_j, $\forall i, j \in 1, \ldots, N$. Une valeur β_q correspond à celle de l'α-coupe au-delà de laquelle π_i et π_j ne sont plus cohérentes (elles n'ont plus d'intersections). L'algorithme introduit dans les travaux de Destercke [58, 55] est donné ci-dessous (voir Algorithme 2).

	m_1		m_2	
	Support	Noyau	Support	Noyau
g_1	$[1,5]$	$[2,3]$	$[1,7]$	$[2,5]$
g_2	$[2,3]$	$[3,3]$	$[1,3]$	$[2,2]$
g_3	$[4,7]$	$[5,6]$	$[6,7]$	$[6,6]$
g_4	$[6,10]$	$[8,9]$	$[8,9]$	$[8,8]$

TABLE 4.1 – Informations modélisées par des intervalles flous

Algorithme 2: Valeurs β_k du résultat de la fusion des SMC pour une variable m

Entrées : n distributions de possibilité π_i

Sorties : Liste des valeurs β_k

Liste=$\emptyset$;$i = 1$
pour chaque $k = 1, \ldots, N$ **faire**
  pour chaque $l = 1, \ldots, N$ **faire**
    $\beta_i = \max(\min(\pi_k, \pi_l))$
    $i = i + 1$
    Ajouter β_i à la liste
  fin
fin
Ordonner les éléments de l'ensemble Liste par ordre croissant

4.1 Fusion conjonctive de distributions de possibilités

Définition 10 (Espace de fusion de distributions) *Un espace de fusion de distributions Q_m est composé des informations fournies par les sources pour la variable m munies de leurs intervalles de confiance, et de tous leurs résultats de fusion possibles suivant un opérateur de fusion ϕ_m.*

Par exemple, dans la Table 4.1, en considérant l'opérateur conjonctif et la variable m_1, l'espace de fusion de distributions est donné par :

$$\begin{aligned} Q_m = \quad & \{(\lceil [1,5],[2,3] \rceil,]0,1]), \\ & (\lceil [2,3],[3,3] \rceil,]0,1]), \\ & (\lceil [4,7],[5,6] \rceil,]0,1]), \\ & (\lceil [6,10],[8,9] \rceil,]0,1]), \\ & (\lceil [2,3],[3,3] \rceil,]0,1]), \\ & (\lceil [4,5],[\tfrac{13}{3},\tfrac{13}{3}] \rceil,]0,\tfrac{1}{3}]), \\ & (\lceil [6,7],[\tfrac{20}{3},\tfrac{20}{3}] \rceil,]0,\tfrac{1}{3}]), \emptyset\} \end{aligned}$$

Une remarque relative à la fusion conjonctive est que le résultat n'est pas forcément normalisé. Ceci est dû au conflit entre les sources puisqu'aucune valeur n'est considérée simultanément et complètement possible par les sources. Moins le résultat de la conjonction est normalisé, plus le conflit entre les sources est important, et moins il est légitime de supposer que toutes les sources du sous-ensemble A sont fiables. En revanche, on peut continuer de supposer que les sources sont fiables et renormaliser le résultat de la fusion conjonctive, en divisant par la hauteur de la distribution résultante de la fusion qui représente le degré de conflit entre les sources.

4.1.1 L'opérateur de fusion conjonctif comme infimum

Le résultat de la fusion conjonctive de N distributions de possibilités est donné par :

$$\hat{\pi}_m = \bigcap_{i=1,\ldots,N} \pi_{m(g_i)} \tag{4.1}$$

Comme indiqué dans la section 1.8, l'opérateur conjonctif est une opération commutative, associative et idempotente. L'opérateur de fusion conjonctif des distributions peut donc être considéré comme un infimum. Par conséquent, $(Q_m, \cap)$ est un inf-demi-treillis.

La relation d'ordre des intervalles classiques est étendue pour les sous-ensembles flous. Par la suite, les éléments de Q_m sont ordonnés tels que $R_1 \cap R_2 = R_1 \Leftrightarrow R_1 \subseteq R_2$. La relation d'ordre appliquée sur les intervalles classiques est étendue aux sous-ensembles flous grâce à l'équation $E \subseteq F \Leftrightarrow \mu_E(x) \leq \mu_F(x), \forall x$ pour toute paire de sous-ensembles flous E et F ayant des fonctions d'appartenance respectives μ_E et μ_F.

4.1.2 Construction du treillis de concepts

Le triplet $(G, (Q_m, \cap), \delta)$ est une structure de patrons où G est l'ensemble de sources, $(Q_m, \cap)$ désigne l'inf-demi-treillis et δ est la fonction qui associe à chaque objet dans G son information dans Q_m.

Les opérateurs de dérivation pour le mode de fusion conjonctif sont définis par :

$$\begin{array}{rcl} A^{\square} & = & \bigcap_{g \in A} \pi_{m(g)} \\ d^{\square} & = & \{g \in G | d \subseteq \pi_{m(g)}\} \end{array} \tag{4.2}$$

Le premier opérateur (Eq 4.2) retourne le sous-ensemble flou commun à tous les sous-ensembles flous des objets de A et le second (Eq 4.2) fournit pour chaque distribution d le sous-ensemble de sources de G qui la contiennent.

Revenons à l'exemple de la Table 4.1, considérons les objets g_1, g_2 et g_3 ayant les descriptions respectives $\delta(g_1) = (\lceil [1,5],[2,3] \rceil,]0,1])$, $\delta(g_2) = (\lceil [2,3],[3,3] \rceil,]0,1])$ et $\delta(g_3) = \lceil [4,7],[5,6] \rceil,]0,1])$:

$$\begin{array}{rcl} \{g_1, g_2\}^{\square} & = & \pi_{m_1(g_1)} \cap \pi_{m_1(g_2)} \\ & = & \langle [1,5]; [2,3] \rangle \cap \langle [2,3]; [3,3] \rangle \\ & = & \langle [2,3]; [3,3] \rangle \\ & = & (\lceil [2,3],[3,3] \rceil,]0,1]) \\ \langle [2,3]; [3,3] \rangle^{\square} & = & \{g \in G \mid \langle [2,3]; [3,3] \rangle \sqsubseteq \pi_{m(g)}\} \\ & = & \{g_1, g_2\} \end{array}$$

$$\begin{array}{rcl} \{g_1, g_3\}^{\square} & = & \pi_{m_1(g_1)} \cap \pi_{m_1(g_3)} \\ & = & \langle [1,5]; [2,3] \rangle \cap \langle [4,7]; [5,6] \rangle \\ & = & (\lceil [4,5]; [\frac{13}{3}, \frac{13}{3}] \rceil,]0, \frac{1}{3}]) \\ (\lceil [4,5]; [\frac{13}{3}, \frac{13}{3}] \rceil,]\frac{1}{3}, \frac{1}{3}])^{\square} & = & \{g \in G \mid (\lceil [4,5]; [\frac{13}{3}] \rceil,]\frac{1}{3}, \frac{1}{3}]) \sqsubseteq \pi_{m(g)}\} \\ & = & \{g_1, g_3\} \end{array}$$

Le résultat de la fusion des éléments d'informations du sous-ensemble $\{g_1, g_2\}$ est normalisé et celui de $\{g_1, g_3\}$ n'est pas normalisé.

Puisque $\{g_1, g_2\}^{\square} = (\lceil [2,3],[3,3] \rceil,]0,1])$ et $(\lceil [2,3],[3,3] \rceil,]0,1])^{\square} = \{g_1, g_2\}$, la paire $(\{g_1, g_2\}, (\lceil [2,3],[3,3] \rceil,]0,1]))$ est un concept formel.

L'ensemble de tous les concepts ordonnés forme le treillis correspondant au triplet $(G, (Q_m, \cap), \delta)$ donné sur le diagramme de la Figure 4.1.

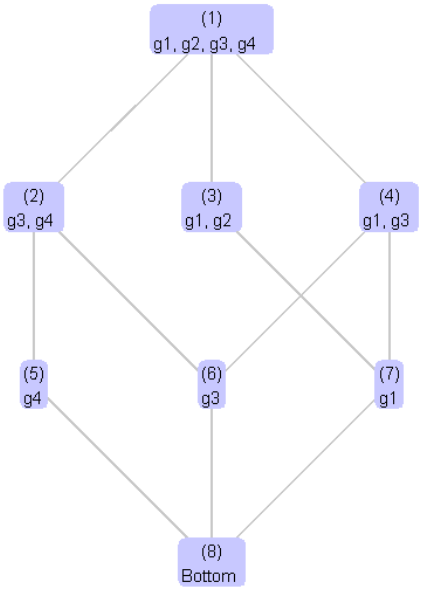


FIGURE 4.1 – Treillis des résultats conjonctifs pour la variable m_1 de la Table 4.1

Numéro de concept	Intension
(1)	$\emptyset$
(2)	$(\lceil[6,7],[\frac{20}{3},\frac{20}{3}]\rceil,]0,\frac{1}{3}])$
(3)	$(\lceil[2,3],[3,3]\rceil,]0,1])$
(4)	$(\lceil[4,5],[\frac{13}{3},\frac{13}{3}]\rceil,]0,\frac{1}{3}])$
(5)	$(\lceil[6,10],[8,9]\rceil,]0,1])$
(6)	$(\lceil[4,7],[5,6]\rceil,]0,1])$
(7)	$(\lceil[1,5],[2,3]\rceil,]0,1])$

TABLE 4.2 – Résultats conjonctifs de la variables m_1 de la Table 4.1

4.1.3 Interprétation du treillis de concepts

Les extensions sont données pour chaque concept sur la Figure 4.1. Les intensions sont données dans la Table 4.2. Le treillis nous permet de plus de conserver l'origine de l'information ainsi que le niveau de confiance du résultat de fusion partiel obtenu. Les intensions représentent le résultat de fusion des sources de l'extension. Le concept général $\top$ est le résultat de fusion global. L'intervalle de confiance donné dans chaque intension montre si les sources sont consistantes (distribution résultante normalisée) ou elles sont partiellement consistantes (distribution résultante non normalisée).

Ce treillis est une généralisation du treillis obtenu pour les intervalles classiques puisque les distributions de possibilités sont des intervalles emboîtés.

Tout concept (A, d) est intéressant de plusieurs points de vue. Par exemple, pour le concept $(\{g_1, g_2\}, (\lceil[2,3],[3,3]\rceil,]0,1]))$ (marqué (3) sur la Figure 4.1), on a :

- L'intension d qui représente l'ensemble des intervalles emboîtés entre deux α-coupes spécifiques est le résultat de la fusion des informations des sources de l'extension des sources dans A, par exemple $(\lceil[2,3],[3,3]\rceil,]0,1])$ est le résultat de la fusion conjonctive du sous-ensemble $\{g_1, g_2\}$.

- L'intervalle de confiance montre la consistance entre les sources de A, g_1 et g_2 sont totalement consistante puisque l'intervalle de la quantité de confiance a pour maximum 1.
- Le sous-ensemble A est maximal.
- L'extension préserve l'origine de l'information, $(\lceil [2,3],[3,3] \rceil,]0,1])$ provient de g_1 et g_2.
- Le concept général $\top$ représente le résultat global de la fusion de toutes les sources.

4.2 Fusion disjonctive de distributions de possibilités

Dans cette section, nous généralisons le cas des intervalles pour la fusion disjonctive et nous reprenons le cas de la fusion de distributions de possibilités. Nous illustrons nos exemples à partir de la Table 4.1 pour la variable m_1.

L'opérateur de fusion disjonctif appliqué à N sources fournit le résultat suivant :

$$\hat{\pi}_m = \bigcup_{i=1,\ldots,N} \pi_{m(g_i)}$$

4.2.1 L'opérateur de fusion disjonctif comme infimum

L'opérateur disjonctif est commutatif, associatif et idempotent (voir section 1.8). L'opérateur disjonctif est donc un infimum et l'espace de fusion de distributions muni de la disjonction est un inf-demi-treillis.

Soit Q_m l'espace de fusion de distributions en utilisant l'opérateur de fusion disjonctif. Les éléments de Q_m sont des intervalles emboîtés et sont ordonnés tels que $R_1 \cup R_2 = R_2 \Leftrightarrow R_1 \supseteq R_2$. La relation d'ordre des intervalles flous pour l'union est étendue à partir de la relation d'ordre appliquée sur les intervalles classiques.

Le triplet $(G, (Q_m, \cup), \delta)$ est une structure de patrons où G est l'ensemble de sources, $(Q_m, \cup)$ est l'inf-demi-treillis ordonné et δ est une fonction qui associe à chaque sources dans G sa description dans Q_m.

4.2.2 Construction du treillis de concepts

Les opérateurs de dérivation pour le mode de fusion disjonctif sont définis par :

$$\begin{array}{lcl} A^{\square} & = & \bigcup_{g \in A} \pi_{m(g)} \\ d^{\square} & = & \{g \in G | d \supseteq \pi_{m(g)}\} \end{array} \tag{4.3}$$

L'opérateur donné dans l'équation (4.3) retourne le résultat de la fusion disjonctive des distributions des objets dans A. L'opérateur dans l'équation (4.3) retourne pour chaque description d le sous-ensemble de sources de G qui l'ont fournie.

Revenons à l'exemple de la Table 4.1 et considérons les objets g_1, g_2 et g_3. On a $\delta(g_1) = (\lceil [1,5],[2,3] \rceil,]0,1])$, $\delta(g_2) = (\lceil [2,3],[3,3] \rceil,]0,1])$ et $\delta(g_3) = \lceil [4,7],[5,6] \rceil,]0,1])$:

$$\begin{array}{rcl} \{g_1, g_2\}^{\square} & = & \pi_{m_1(g_1)} \cup \pi_{m_1(g_2)} \\ & = & \langle [1,5]; [2,3] \rangle \cup \langle [2,3]; [3,3] \rangle \\ & = & \langle [1,5]; [2,3] \rangle \\ & = & (\lceil [1,5],[2,3] \rceil,]0,1]) \\ \langle [1,5]; [2,3] \rangle^{\square} & = & \{g \in G \mid \langle [1,5]; [2,3] \rangle \sqsubseteq \pi_{m(g)}\} \\ & = & \{g_1, g_2\} \end{array}$$

Puisque $\{g_1, g_2\}^{\square} = (\lceil [1,5],[2,3] \rceil,]0,1])$ et $(\lceil [1,5],[2,3] \rceil,]0,1])^{\square} = \{g_1, g_2\}$, la paire $(\{g_1, g_2\}, (\lceil [1,5],[2,3] \rceil,]0,1]))$ est un concept formel. L'ensemble de tous les concepts ordonnés forme le treillis correspondant au triplet $(G, (Q_m, \cup), \delta)$ donné sur le diagramme de la Figure 4.2.

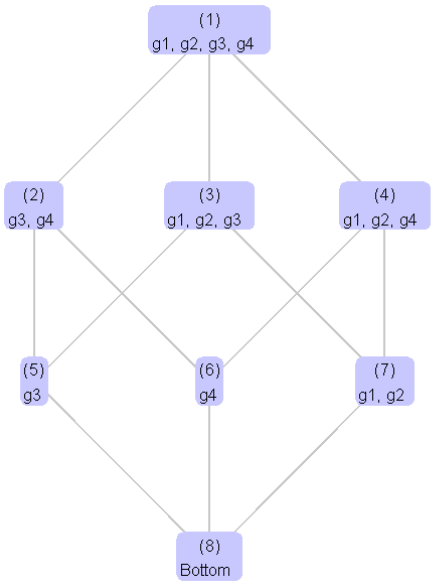


FIGURE 4.2 – Treillis des résultats disjonctifs de la variable m_1 de la Table 4.1

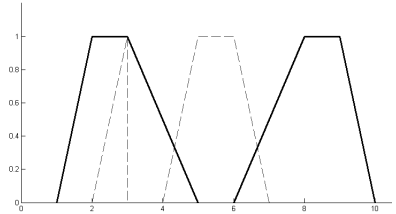
4.2.3 Interprétation du treillis de concepts

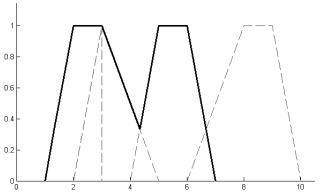
Le treillis obtenu est donné sur Figure 4.2. Les extensions sont données pour chaque concept. Les intensions sont données dans la Table 4.3. La Table 4.4 montre un exemple des distributions de plusieurs concepts du treillis. Comme dans le cas des intervalles, le treillis nous permet d'identifier des sous-ensembles maximaux de sources avec leurs résultats de fusion disjonctifs où l'information est modélisée par une distribution de possibilités. L'intervalle de confiance montre le conflit entre les sources. Le treillis construit à partir de l'opérateur disjonctif pour la fusion des distributions de possibilités est le treillis généralisé du treillis obtenu dans le cas des intervalles classiques puisque les distributions de possibilités sont des intervalles emboîtés. Le treillis possède les mêmes propriétés qu'un treillis obtenu dans le cas des intervalles classiques c-à-d tout concept contient dans son intension le résultat de la fusion pour les sources dans son extension. Un concept est moins informatif que son sous-concept puisqu'on enrichit l'information avec une nouvelle information qui n'est pas nécessairement fiable.

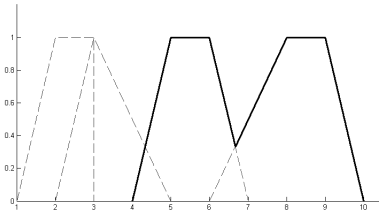
Notons que les résultats de fusion ne sont pas nécessairement convexes. Il est toujours possible de considérer l'enveloppe convexe du résultat. On y perd de l'information mais on gagne de l'efficacité dans le calcul. Dans le processus de décision, l'expert pourra trouver tous les sous-ensembles maximaux possibles des sources de G puisque le résultat global n'est pas souvent utile et est imprécis, surtout lorsque les sources fournissent des informations incohérentes et contradictoires.

Numéro	Intension
(1)	$(\lceil[1,10],[\frac{4}{3},\frac{29}{3}]\rceil,]0,\frac{1}{3}]) \cup (\lceil[\frac{4}{3},\frac{13}{3}],[2,3.0]\rceil,]\frac{1}{3},1]) \cup (\lceil[\frac{13}{3},\frac{20}{3}],[5,6]\rceil,]\frac{1}{3},1]) \cup (\lceil[\frac{20}{3},\frac{29}{3}],[8,9]\rceil,]\frac{1}{3},1])$
(2)	$(\lceil[4,10.0],[\frac{13}{3},\frac{29}{3}]\rceil,]0,\frac{1}{3}]) \cup (\lceil[\frac{13}{3},\frac{20}{3}],[5,6]\rceil,]\frac{1}{3},1]) \cup (\lceil[\frac{20}{3},\frac{29}{3}],[8,9]\rceil,]\frac{1}{3},1]$
(3)	$(\lceil[1,7],[\frac{4}{3},\frac{20}{3}]\rceil,]0,\frac{1}{3}]) \cup (\lceil[\frac{4}{3},\frac{13}{3}],[2,3.0]\rceil,]\frac{1}{3},1]) \cup (\lceil[\frac{13}{3},\frac{20}{3}],[5,6]\rceil,]\frac{1}{3},1])$
(4)	$(\lceil[1,5],[2,3]\rceil,]0,1]) \cup (\rceil[6,10.0],[8,9]\rceil,]0,1])$
(5)	$(\lceil[4,7],[5,6]\rceil,]0,1])$
(6)	$(\lceil[6,10.0],[8,9]\rceil,]0,1])$
(7)	$(\lceil[1,5],[3.0,3.0]\rceil,]0,1])$

TABLE 4.3 – Intensions du treillis des résultats disjonctifs des distributions

(4) 

(3) 

(2) 

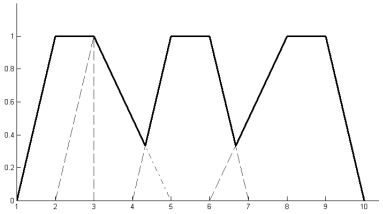
(1) 

TABLE 4.4 – Exemples de distributions obtenues dans le treillis de concepts

4.3 Fusion des SMC des distributions

4.3.1 Extraction des SMC dans les distributions de possibilités

Pour l'opérateur de fusion fondé sur les SMC, nous calculons d'abord les SMC des distributions et nous appliquons l'opérateur disjonctif sur ces derniers. Pour extraire les SMC des distributions, nous rappelons la méthode introduite par Destercke pour les distributions de possibilités qui consiste à calculer des valeurs particulières des niveaux α pour trouver le résultat de la fusion des SMC des distributions.

Une distribution de possibilités, par définition, est un ensemble d'intervalles emboîtés qui sont ses α-coupes.

En effet, à chaque niveau α, les α-coupes des distributions forment un ensemble d'intervalles dont nous pourrons calculer les résultats de leur fusion en utilisant l'algorithme 1. Par conséquent, pour tout $\mathbb{I}^\alpha = \{I_1^\alpha, \ldots, I_N^\alpha\}$ qui correspond à l'ensemble de n α-coupes I_i^α $(i = 1, \ldots, N(\alpha))$, on obtient un résultat de fusion des SMC des intervalles comme détaillé dans la section 3.3.2.

Formellement, soit $\mathbb{K}^\alpha$ les sous-ensembles d'intervalles de $\mathbb{I}^\alpha$ tels que $\bigcap_{i=1,\ldots,|\mathbb{K}^\alpha|} K_i^\alpha \neq \emptyset$, avec $K_i^\alpha \in \mathbb{I}^\alpha$, la méthode de fusion s'écrit alors : $\mathcal{K}^\alpha = \bigcup_{j=1,\ldots,p} \mathbb{K}_j^\alpha$ où p est le nombre des sous-ensembles maximaux cohérents pour un niveau α donné. En général, la méthode s'écrit pour N sources ayant $N(\alpha)$ SMC pour la coupe de niveau α :

$$\hat{\pi}^\alpha = \bigcup_{j=1,\ldots,N(\alpha)} \bigcap_{i=1,\ldots,|\mathcal{E}_j^\alpha|} K_i^\alpha \tag{4.4}$$

En général, $\mathcal{K}^\alpha$ est une union d'intervalles disjoints, et nous n'avons pas $\mathcal{K}^\alpha \supset \mathcal{K}^\beta$, $\forall\, \beta > \alpha$, par conséquent le résultat de la fusion globale de N sources n'est pas une distribution de possibilités, puisque les α-coupes ne sont pas emboîtées. Cependant il existe un ensemble fini de valeurs $\beta_q \in]0, 1]$ telles que $0 = \beta_1 \leq \ldots \leq \beta_q \leq \beta_{q+1} = 1$ et les ensembles $\mathcal{K}^\alpha$ sont emboîtés pour $\alpha \in]\beta_q, \beta_{q+1}]$. En utilisant l'Algorithme 2, nous pouvons calculer ces valeurs. Considérons l'exemple de la Table 4.1, les distributions de possibilités des sources de G pour la variable m_1 sont : $\pi_{m_1(g_1)} = (\lceil [1, 5], [2, 3] \rceil,]0, 1])$, $\pi_{m_1(g_2)} = (\lceil [2, 3], [3, 3] \rceil,]0, 1])$, $\pi_{m_1(g_3)} = (\lceil [4, 7], [5, 6] \rceil,]0, 1])$, et $\pi_{m_1(g_4)} = (\lceil [6, 10], [8, 9] \rceil,]0, 1])$. En appliquant l'algorithme 2, nous trouvons les valeurs $\beta = 1$ pour le sous-ensemble de distributions $\{\pi_{m_1(g_1)}, \pi_{m_1(g_2)}\}$, $\beta = \frac{1}{3}$ pour le sous-ensemble de distributions $\{\pi_{m_1(g_2)}, \pi_{m_1(g_3)}\}$, et $\beta = \frac{1}{3}$ pour le sous-ensemble $\{\pi_{m_1(g_3)}, \pi_{m_1(g_4)}\}$. Alors, $Liste = \{1, \frac{1}{3}, \frac{1}{3}\}$. Les valeurs de β_q ordonnées correspondantes à notre exemple sont donc $0 < \beta_1 = \beta_2 = \frac{1}{3} < \beta_3 = 1$. Comme mentionné plus haut, à partir du niveau $\alpha = \frac{1}{3}$, les distributions qui représentent $\pi_{m_1(g_2)}$ et $\pi_{m_1(g_3)}$ (respectivement $\pi_{m_1(g_3)}$ et $\pi_{m_1(g_4)}$) ne sont plus cohérentes. Par conséquent, le résultat de la fusion des SMC est un ensemble d'intervalles emboîtés entre $]0, \frac{1}{3}]$ et $]\frac{1}{3}, 1]$

Le résultat de la fusion des SMC est, en général, une structure de croyance floue [58]. Si nous appliquons l'équation 4.3.1 pour $\alpha \in]\beta_q, \beta_{q+1}]$, nous obtenons un ensemble flou non normalisé dont le degré d'appartenance varie dans l'intervalle $]\beta_q, \beta_{q+1}]$ puisque les ensembles sont emboîtés entre ces valeurs. Cet ensemble peut être normalisé en changeant la fonction d'appartenance à ce sous sous-ensemble μ par

$$\frac{\max(\mu - \beta_q, 0)}{\beta_{q+1} - \beta q}$$

et en assignant le poids $\lambda_i = \beta_{q+1} - \beta q$ à cet ensemble flou. Le poids λ_i peut s'interpréter comme étant la quantité de confiance donnée à la représentation de l'ensemble flou.

La Figure 4.3 présente le résultat de la fusion des SMC de la variable m_1. Si nous prenons l'α-coupe stricte (c-à-d les supports), nous retrouvons le résultat montré sur la Figure 3.3 pour

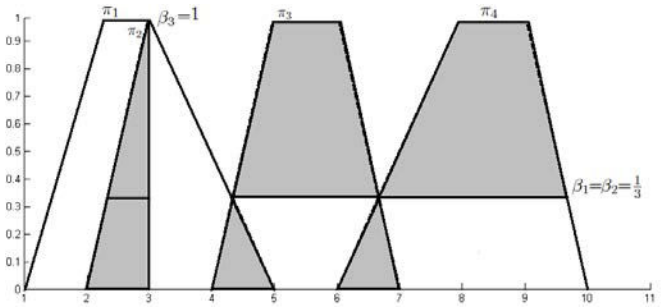


FIGURE 4.3 – Résultat de la fusion des SMC des distributions de l'exemple de la Table 4.1

les intervalles. Notons également que si toutes les sources s'accordent sur au moins une valeur commune (c-à-d il existe une seule valeur pour α), le résultat est un ensemble flou unique qui représente la conjonction usuelle. Au contraire, si les sources sont disjonctives deux à deux (en conflit) alors le résultat est un ensemble flou unique qui représente la disjonction usuelle. La structure obtenue est interprétée comme une bonne représentation de toutes les informations fournies par les sources. Cependant, il semble aussi difficile de tirer des conclusions ou des informations directement de la structure. Dans les sections suivantes nous présentons la méthode fondée sur l'AFC pour extraire de l'information à partir de la méthode de fusion des SMC pouvant être utile pour l'analyste.

4.3.2 Pré-traitement des données

Nous utilisons une méthode de traitement des données similaire à celle introduite dans la section 3.3.3, i.e. nous calculons les SMC avec leurs valeurs pour la variable m et construisons le treillis en utilisant l'opérateur disjonctif. Comme mentionné ci-dessus, les résultats de la fusion des SMC ne sont pas des distributions, mais des sous-ensembles de valeurs munis des intervalles de confiance. Par conséquent, pour construire le treillis des SMC, il faut considérer plusieurs cas suivant les valeurs de α.

Revenons à l'exemple de la Table 4.1, où on a $0 < \beta_1 = \beta_2 = \frac{1}{3} < \beta_3 = 1$. A partir du niveau $\alpha = \frac{1}{3}$ les sources g_1 et g_3 (respectivement g_3 et g_4) ne sont plus cohérentes. Formellement, on considère $(\mathcal{O}, (F, \cup), \delta)$ comme une structure de patrons où les objets dans $\mathcal{O} \subseteq G$ sont les origines des SMC des distributions pour la variable m en respectant les valeurs de α. $(F, \cup)$ est l'inf-demi-treillis des descriptions muni de l'union et δ est une fonction qui associe à chaque objet dans $\mathcal{O}$ sa description dans F. Chaque objet est décrit par sa valeur et un intervalle de degré d'appartenance, c'est-à-dire un ensemble d'intervalles emboîtés entre deux niveaux. Par exemple, dans la Table 4.1, le sous-ensemble maximal cohérent des sources g_3 et g_4, pour la variable m_1, est

$$\begin{cases} \left\lceil [6,7]; [\frac{20}{3}, \frac{20}{3}] \right\rceil & \text{si } \alpha \in [0, \frac{1}{3}] \\ \left\lceil [\frac{13}{3}, \frac{20}{3}]; [5,6] \right\rceil \cup \left\lceil [\frac{20}{3}, \frac{29}{3}]; [8,9] \right\rceil & \text{si } \alpha \in]\frac{1}{3}, 1] \end{cases}$$

Les éléments de l'espace de fusion F sont des couples $(R, [\gamma_1, \gamma_2])$ où R est l'ensemble des intervalles emboîtés entre les deux niveaux γ_1 et γ_2. Ces éléments sont ordonnés, pour tout $d_1 = (R_1; [\gamma_1, \zeta_1])$ et $d_2 = (R_2; [\gamma_2, \zeta_2])$, par :

$$d_1 \cup d_2 = d_1 \Leftrightarrow d_1 \supseteq d_2$$

$\mathbb{K}_1 = (g_1, g_2)$	$\left\lceil [2,3]; [\frac{7}{3}, 3] \right\rceil, [0, \frac{1}{3}]$
$\mathbb{K}_2 = (g_1, g_3)$	$\left\lceil [4,5]; [\frac{13}{3}, \frac{13}{3}] \right\rceil, [0, \frac{1}{3}]$
$\mathbb{K}_3 = (g_3, g_4)$	$\left\lceil [6,7]; [\frac{20}{3}, \frac{20}{3}] \right\rceil, [0, \frac{1}{3}]$
$\mathbb{K}_4 = (g_1, g_2)$	$\left\lceil [\frac{7}{3}, 3]; [3,3] \right\rceil, [\frac{1}{3}, 1]$
$\mathbb{K}_5 = (g_3)$	$\left\lceil [\frac{13}{3}, \frac{20}{3}]; [5,6] \right\rceil, [\frac{1}{3}, 1]$
$\mathbb{K}_6 = (g_4)$	$\left\lceil [\frac{20}{3}, \frac{29}{3}]; [8,9] \right\rceil, [\frac{1}{3}, 1]$

TABLE 4.5 – La structure de patrons des SMC de la variable m_1 de la Table 4.1

ceci est équivalent à la relation suivante :

$$\begin{aligned}(R_1; [\gamma_1, \zeta_1]) \cup (R_2; [\gamma_2, \zeta_2]) = (R_1; [\gamma_1, \zeta_1]) \quad &\Leftrightarrow \quad R_1 \cup R_2 = R_1 \text{ et } [\gamma_1, \zeta_1] \cup [\gamma_2, \zeta_2] = [\gamma_1, \zeta_1] \\ &\Leftrightarrow \quad R_1 \supseteq R_2 \text{ et } [\gamma_1, \zeta_1] \supseteq [\gamma_2, \theta_2]\end{aligned}$$

L'ensemble des concepts ordonnés forme le treillis de concepts dont le $\top$ est le résultat de la fusion globale des SMC pour les valeurs de $\alpha \in [0, 1]$. Ainsi, le treillis permet d'identifier des sous-ensembles maximaux avec leur résultat de fusion et un intervalle de degré d'appartenance au sous-ensemble obtenu tout en gardant l'origine de l'information.

4.3.3 Construction du treillis de concepts

L'opérateur de fusion fondée sur la recherche des SMC des distributions est un opérateur commutatif, idempotent mais il n'est pas associatif. Par conséquent, l'opérateur fondé sur les SMC ne permet pas de construire un treillis de concepts des informations et de leurs résultats de fusion puisqu'il ne peut pas être un infimum.

Partant du pré-traitement que nous avions utilisé dans le cas des intervalles et de la définition du résultat de fusion des SMC sur les distributions de possibilités, nous proposons ici une méthode permettant de calculer le treillis des SMC avec un intervalle de degré d'appartenance d'une part et d'autre part leur union qui représente la fusion des SMC des sources du contexte initial.

Considérons l'exemple de la Table 4.1 et la variable m_1. La structure de patrons est donnée dans la Table 4.5. En considérant le triplet $(\mathcal{O}, (F, \cup), \delta)$ et les opérateurs de dérivation de l'opérateur disjonctif, nous pouvons calculer les concepts et les ordonner. Le treillis de concept obtenu permet d'avoir une classification des informations suivant leurs résultats de fusion munis d'une quantité de confiance représentée par l'intervalle de degré d'appartenance entre deux α-coupes spécifiques. Le treillis contient tous les groupement possibles et dans notre exemple pour la variable m_1 de la Table 4.1, le treillis a 63 concepts. Par exemple : $\big(\{\mathbb{K}_1, \mathbb{K}_3\}, \left\lceil [6,7]; [\frac{20}{3}, \frac{20}{3}],]0, \frac{1}{3}] \right\rceil \cup \left\lceil [2,3]; [\frac{7}{3}, 3],]0, \frac{1}{3}] \right\rceil\big)$.

Les concepts obtenus ne sont pas tous utiles. En effet, il y a des concepts pour lesquels les analystes ne peuvent pas tirer de conclusions à partir du résultat de fusion obtenu, comme par exemple le concept $(\{\mathbb{K}_1, \mathbb{K}_5\}, \left\lceil [2,3]; [\frac{7}{3}, 3],]0, \frac{1}{3}] \right\rceil \cup \left\lceil [\frac{13}{3}, \frac{20}{3}]; [5,6],]0, 1] \right\rceil)$, où le résultat de la fusion est donné pour deux sous-ensembles avec deux intervalles de degré d'appartenance différents. La renormalisation et la convexification de ce sous-ensemble nous permet d'avoir un intervalle de degré d'appartenance de $]0, 1]$. Par conséquent, le concept devient un concept calculé par le treillis de concept. Pour cela nous extrayons parmi les concepts résultants les concepts qui ont des degrés de confiance égaux et qui permettent de mieux décider. La Figure 4.4 illustre les concepts extraits à partir du treillis obtenu. Le nombre de concepts restant est de 18 concepts. Les extensions et les intensions de ces concepts sont données dans la Table 4.6.

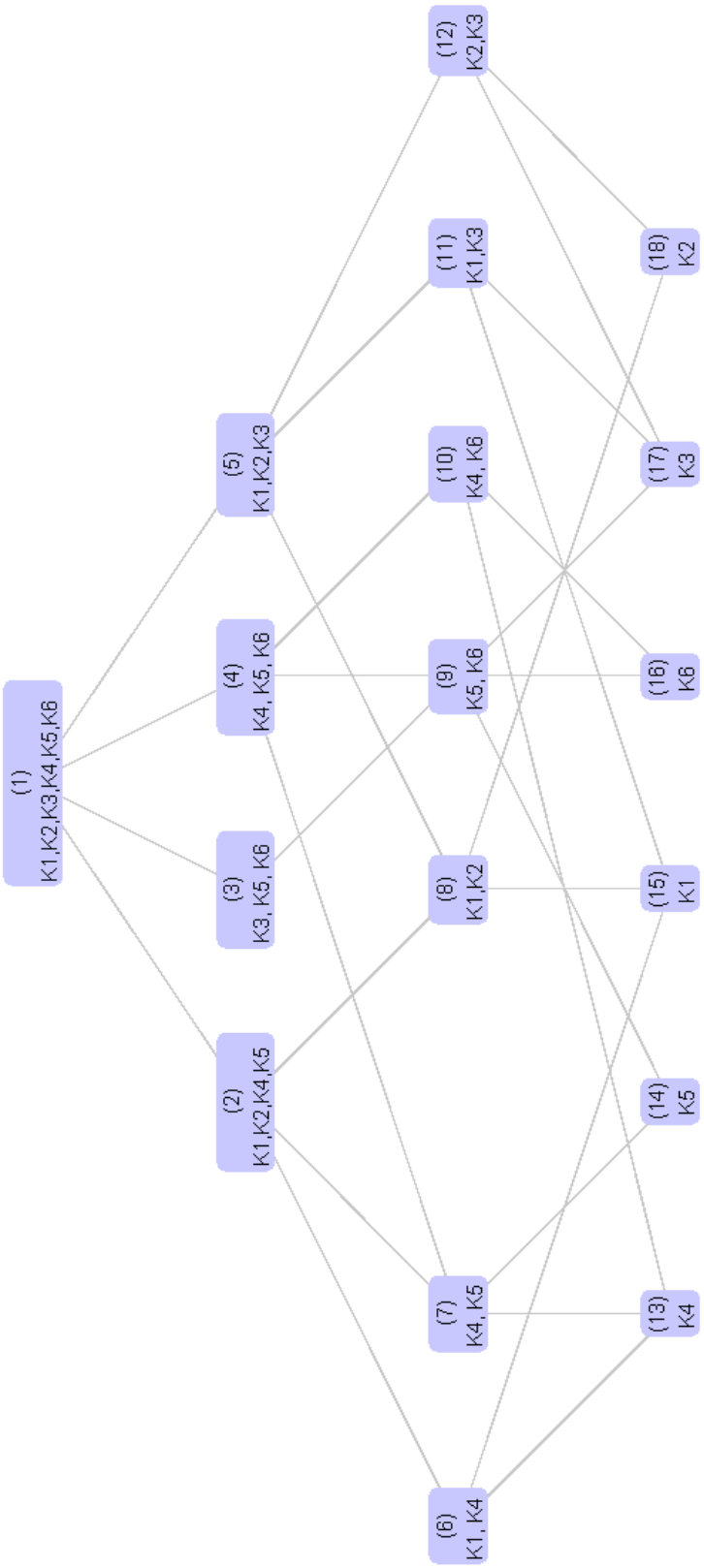


FIGURE 4.4 – Concepts extraits à partir du treillis des SMC de la variable m_1 de la Table 4.1

Extension	Intension
$\mathbb{K}_1$	$[2,3];[\frac{7}{3},3],]0,\frac{1}{3}]]$
$\mathbb{K}_2$	$[4,5];[\frac{13}{3},\frac{13}{3}],]0,\frac{1}{3}]]$
$\mathbb{K}_3$	$[6,7];[\frac{20}{3},\frac{20}{3}],]0,\frac{1}{3}]]$
$\mathbb{K}_4$	$[\frac{7}{3},3];[3,3],]\frac{1}{3},1]]$
$\mathbb{K}_5$	$[\frac{13}{3},\frac{20}{3}];[5,6],]\frac{1}{3},1]]$
$\mathbb{K}_6$	$[\frac{20}{3},\frac{29}{3}];[8,9],]\frac{1}{3},1]]$
$\mathbb{K}_1,\mathbb{K}_2$	$[4,5];[\frac{13}{3},\frac{13}{3}],]0,\frac{1}{3}]] \cup [2,3];[\frac{7}{3},3],]0,\frac{1}{3}]]$
$\mathbb{K}_1,\mathbb{K}_3$	$[6,7];[\frac{20}{3},\frac{20}{3}],]0,\frac{1}{3}]] \cup [2,3];[\frac{7}{3},3],]0,\frac{1}{3}]]$
$\mathbb{K}_1,\mathbb{K}_4$	$[2,3];[\frac{7}{3},3],]0,\frac{1}{3}]] \cup [[\frac{7}{3},3];[3,3],]\frac{1}{3},1]]$
$\mathbb{K}_2,\mathbb{K}_3$	$[4,5];[\frac{13}{3},\frac{13}{3}],]0,\frac{1}{3}]] \cup [6,7];[\frac{20}{3},\frac{20}{3}],]0,\frac{1}{3}]]$
$\mathbb{K}_4,\mathbb{K}_5$	$[\frac{7}{3},3];[3,3],]\frac{1}{3},1]] \cup [[\frac{13}{3},\frac{20}{3}];[5,6],]\frac{1}{3},1]]$
$\mathbb{K}_4,\mathbb{K}_6$	$[\frac{7}{3},3];[3,3],]\frac{1}{3},1]] \cup [[\frac{20}{3},\frac{29}{3}];[8,9],]\frac{1}{3},1]]$
$\mathbb{K}_5,\mathbb{K}_6$	$[\frac{13}{3},\frac{20}{3}];[5,6],]\frac{1}{3},1]] \cup [[\frac{20}{3},\frac{29}{3}];[8,9],]\frac{1}{3},1]]$
$\mathbb{K}_1,\mathbb{K}_2,\mathbb{K}_3$	$[45];[\frac{13}{3},\frac{13}{3}],]0,\frac{1}{3}]] \cup [2,3];[\frac{7}{3},3],]0,\frac{1}{3}]] \cup [67];[\frac{20}{3},\frac{20}{3}],]0,\frac{1}{3}]]$
$\mathbb{K}_4,\mathbb{K}_5,\mathbb{K}_6$	$[\frac{7}{3},3];[3,3],]\frac{1}{3},1]] \cup [\frac{20}{3},\frac{29}{3}];[8,9],]\frac{1}{3},1]] \cup [[\frac{13}{3},\frac{20}{3}];[5,6],]\frac{1}{3},1]]$
$\mathbb{K}_3,\mathbb{K}_5,\mathbb{K}_6$	$[6,7];[\frac{20}{3},\frac{20}{3}],]0,\frac{1}{3}]] \cup [[\frac{20}{3},\frac{29}{3}];[8,9],]\frac{1}{3},1]] \cup [[\frac{13}{3},\frac{20}{3}];[5,6],]\frac{1}{3},1]]$
$\mathbb{K}_1,\mathbb{K}_2,\mathbb{K}_4,\mathbb{K}_5$	$[4,5];[\frac{13}{3},\frac{13}{3}],]0,\frac{1}{3}]] \cup [[\frac{13}{3},\frac{20}{3}];[5,6],]\frac{1}{3},1]] \cup [[\frac{7}{3},3];[3,3],]\frac{1}{3},1]] \cup [2,3];[\frac{7}{3},3],]0,\frac{1}{3}]]$
$\mathbb{K}_1,\mathbb{K}_2,\mathbb{K}_3,$ $\mathbb{K}_4,\mathbb{K}_5,\mathbb{K}_6$	$[4,5];[\frac{13}{3},\frac{13}{3}],]0,\frac{1}{3}]] \cup [[\frac{20}{3},\frac{29}{3}];[8,9],]\frac{1}{3},1]] \cup [[\frac{13}{3},\frac{20}{3}];[5,6],]\frac{1}{3},1]] \cup$ $[[\frac{7}{3},3];[3,3],]\frac{1}{3},1]] \cup [[\frac{20}{3},\frac{29}{3}];[8,9],]\frac{1}{3},1]] \cup [[\frac{13}{3},\frac{20}{3}];[5,6],]\frac{1}{3},1]]$

TABLE 4.6 – Extensions et intensions des concepts de la Figure 4.4

4.3.4 Interprétation du treillis de concepts

Le treillis calculé à partir de la structure de patrons $(\mathcal{O}, (F, \cup), \delta)$ fournit tous les sous-ensembles possibles maximaux de $\mathcal{O}$ avec leurs résultats de fusion disjonctifs. Ce qui se traduit par le résultat de la fusion d'un sous-ensemble de sources de G avec leurs résultats en utilisant l'opérateur de fusion fondé sur la recherche des SMC. Le première ligne tout en bas de la Figure 4.4 désigne les SMC des sources pour les deux intervalles de degré d'appartenance $]0, \frac{1}{3}]$ et $]\frac{1}{3}, 1]$. Le concept général du treillis marqué 1 sur la Figure 4.4 représente le résultat de la fusion globale de toutes les sources de G.

Le calcul de ces sous-ensembles se fait de la même manière expliquée dans la section 3.3.3 en ajoutant que l'union se fait par niveaux de degré d'appartenance. Par exemple, si on considère le concept $(\{\mathbb{K}_2, \mathbb{K}_3\}, \lceil [4,5]; [\frac{13}{3}, \frac{13}{3}],]0, \frac{1}{3}] \rceil \cup \lceil [6,7]; [\frac{20}{3}, \frac{20}{3}],]0, \frac{1}{3}] \rceil)$, le sous-ensemble $\{\mathbb{K}_2, \mathbb{K}_3\} \subseteq \mathcal{O}$ fournit un résultat disjonctif représenté par un ensemble d'intervalles emboîtés donné dans l'intension du concept avec un degré d'appartenance $]0, 1]$. Par conséquent le sous-ensemble des sources de G fournissant le SMC est donné par $\mathbb{K}_2 \cup \mathbb{K}_3 = \{g_1, g_3\} \cup \{g_3, g_4\} = \{g_1, g_3, g_4\}$ puisque $\mathbb{K}_2$ et $\mathbb{K}_3$ ont l'objet g_3 en commun. Si on considère le concept $(\{\mathbb{K}_1, \mathbb{K}_2\}, \lceil [4,5]; [\frac{13}{3}, \frac{13}{3}],]0, \frac{1}{3}] \rceil \cup \lceil [2,3]; [\frac{7}{3}, 3],]0, \frac{1}{3}] \rceil)$, nous avons $\mathbb{K}_1 = \{g_1, g_2\}$ et $\mathbb{K}_2 = \{g_3, g_4\}$, donc le sous-ensemble de sources de G qui a fourni le SMC est $\{g_1, g_2\} \cup \{g_3, g_4\} = \{g_2, g_3, g_4\}$ puisque g_1 est cohérente avec g_3.

4.3.5 Treillis des SMC de distributions pour plusieurs variables

Pour le traitement de deux variables, nous utilisons la méthode de pré-traitement détaillée pour les intervalles classiques dans la section 3.5. Nous supposons que les variables sont indépendantes. Nous cherchons les SMC pour les variables. Par conséquent, le triplet $(\mathcal{O}, (F, \cup), \delta)$ où $\mathcal{O}$ est l'ensemble des sous-ensembles de G des SMC pour toutes les variables. L'espace de fusion est l'union des espaces de fusion de toutes les variables. Les descriptions des objets de $\mathcal{O}$ sont les valeurs de SMC pour la variable m_i pour l'intervalle de degré d'appartenance $[\beta_i, \beta_j]$ et $\emptyset$ sinon. L'infimum est l'opérateur de fusion disjonctif. Le but est d'obtenir une information à partir de la fusion des SMC des variables qui soit utile et facile à manipuler par l'analyste. Ce traitement permet d'avoir les résultats de fusion pour plusieurs variables simultanément.

5

Application à l'évaluation environnementale : indicateur "pesticides"

Sommaire

Face à l'impossibilité d'effectuer des mesures de terrain systématiques pour des raisons de coûts et de temps, et le manque d'outils de prévisions précis qui soient opérationnels, il est souvent fait appel à des indicateurs qui reposent sur un compromis entre les contraintes de précision scientifique et de faisabilité. C'est dans ce contexte que s'inscrivent les travaux de l'équipe Agriculture Durable de l'Institut national de recherche agronomique (INRA) de Colmar et Nancy qui ont conduit au développement des indicateurs agri-environnementaux constituants la méthode INDIGO. Dans ce chapitre, nous définissons la méthode INDIGO, les indicateurs agri-environnementaux : la démarche de leur construction, leurs objectifs, leurs modes de calcul et les imperfections entachant leurs variables. Nous considérons l'exemple de l'indicateur "pesticide" pour une matière active et pour un programme de traitement. Enfin nous appliquerons les méthodes détaillées dans le chapitre 3 et 4 pour calculer le domaine de variation de l'indicateur en tenant compte de variation de ses variables.

5.1 La méthode INDIGO

5.1.1 Contexte de création

Indigo (indicateurs de diagnostic global à la parcelle) est une méthode scientifique d'évaluation de l'impact environnemental des pratiques agricoles sur l'air, le sol, l'eau de surface et l'eau souterraine. Elle a été mise au point par l'INRA, aux centres de Colmar et de Nancy, en collaboration avec l'Association pour la relance agronomique en Alsace (ARAA).

Couplé à une application informatique, INDIGO est un outil de diagnostic et d'aide à la décision, destiné aux techniciens, conseillers, ingénieurs agronomes et agriculteurs qui souhaitent améliorer leurs pratiques pour les rendre plus durables.

5.1.2 Méthode

Grâce à une série d'indicateurs, spécifiques de chaque culture (10 indicateurs pour les grandes cultures), la méthode d'évaluation INDIGO permet de faire un diagnostic des pratiques agricoles à partir de leurs risques potentiels de pollution de l'air, du sol, de l'eau de surface et de l'eau souterraine.

Les indicateurs d'INDIGO permettent d'évaluer l'impact de l'utilisation de l'azote, du phosphore, des produits phytosanitaires, de l'eau, de la matière organique, des ressources énergétiques non renouvelables, ainsi que l'impact de la gestion de la rotation des cultures et de l'assolement et enfin l'impact des structures non productives (haies, bords de champs, fossés,...).

5.1.3 Mise en oeuvre

Calculés par un logiciel adapté, les indicateurs présentent l'avantage d'être simples à utiliser et faciles à interpréter. L'utilisateur introduit des données techniques, qu'il possède en général,

puis lance le calcul. Le résultat est une valeur comprise entre 0 et 10.

Indigo donne aussi le détail des résultats intermédiaires pour une analyse plus fine de la note finale. L'outil Indigo relève les points forts et les points faibles des pratiques analysées et identifie des causes possibles. En plus du diagnostic global à la parcelle, Indigo donne les éléments pour bâtir un conseil agronomique personnalisé. L'utilisateur peut alors apporter des améliorations pour rendre ses pratiques plus respectueuses de l'environnement.

5.2 Les indicateurs agri-environnementaux

5.2.1 Définitions des indicateurs

Plusieurs définitions des indicateurs sont disponibles [131] et en voici deux qui caractérisent les indicateurs agri-environnementaux :

- Les indicateurs sont des mesures qui fournissent des renseignements sur des phénomènes difficiles à mesurer. Les indicateurs servent aussi de repère pour prendre une décision [104].
- Ils fournissent des informations au sujet d'un système complexe en vue de faciliter sa compréhension aux utilisateurs de sorte qu'ils puissent prendre des décisions appropriées qui mènent à la réalisation d'objectifs [142].

Ces deux définitions posent les bases de la construction de l'indicateur. Ce dernier peut résulter d'une mesure, d'une observation, d'une donnée statistique, d'un calcul. Il peut être une sortie de modèle dans le cas d'indicateurs simples ou une agrégation d'indicateurs (appelés sous-indicateurs) pour des indicateurs composites [31, 99, 101]. Deux aspects ressortent "fournir un renseignement, une information" au sujet d'une autre grandeur ou d'un système difficile à mesurer ou à décrire directement, ceci pour aider à "prendre une décision". L'indicateur a donc une vocation d'être utilisé par des acteurs et doit donc répondre à des critères de faisabilité, de simplicité et de lisibilité.

5.2.2 Types et validation des indicateurs

Les indicateurs agri-environnementaux sont souvent employés comme outil de diagnostic pour des objectifs précis [101]. Les indicateurs peuvent être de deux types : simples ou composites. L'indicateur simple est issu de mesures directes ou d'une estimation obtenue à partir d'un modèle [30] tandis que l'indicateur composite provient de l'agrégation de plusieurs variables ou de plusieurs indicateurs simples mesurés ou estimés. Pour les agrosystèmes, les indicateurs composites doivent être flexibles et pragmatiques tout en restant fondés sur les connaissances scientifiques [101]. Ces indicateurs sont construits à l'aide de plusieurs variables dont l'importance varie selon leur impact supposé sur l'environnement [101]. C'est de cette façon qu'a été construit par exemple un indicateur de gestion des produits phytosanitaires [176]. Par souci de lisibilité et de clarté pour les utilisateurs, les indicateurs sont exprimés sur une même échelle après transformation mathématique : en classes [100] ou en rangs [13]. CependantGirardin et al. [100] constatent que classer les variables peut aboutir à un biais sachant que les classes sont des entités limitées et que leur amplitude est déterminée plus ou moins empiriquement. De ce fait, il traduit son résultat d'indicateur sous forme d'un indice calculé par rapport à une référence. De cette façon grâce à l'indicateur, on peut suivre la déviation de la situation étudiée par rapport à la référence. Girardin et al. [100] présentent les indicateurs de la méthode Indigo sur une échelle de 0 (risque maximal pour l'environnement) à 10 (risque minimal pour l'environnement) avec une référence à 7 (risque pour l'environnement acceptable par la société). La fixation de la valeur de référence 7

résulte de processus de consensus entre les références bibliographiques et des avis d'experts. La validation des indicateurs fait partie intégrante de leur construction.

5.3 Démarche de construction d'un indicateur

Nous reprendrons les différentes étapes proposées par [100] et modifiées dans [102]. Les auteurs dans [100, 102] insistent sur la nécessité de bien clarifier les choix préalables à la construction de ces outils avant de commencer la construction et le développement de l'outil. En effet, la construction de l'outil doit dépendre directement des choix de départ et des hypothèses qui en découlent. La phase de construction doit être suivie d'une phase de tests. La démarche de construction de l'indicateur se fait en cinq étapes :

- Choix des utilisateurs et des objectifs.
- Construction de l'indicateur.
- Choix de la référence.
- Test de sensibilité.
- Validations de l'indicateur.

5.3.1 Choix des utilisateurs et des objectifs

La première étape dans la construction d'un indicateur est le choix des utilisateurs potentiels et des objectifs. L'objectif des indicateurs est de fournir des outils de diagnostic et d'aide à la décision aux agriculteurs et aux conseillers agricoles qui sont les utilisateurs potentiels.

L'objectif général se décompose en objectifs opérationnels qui reposeront sur le choix des impacts potentiels à évaluer ou objectifs environnementaux. Il faut donc définir les pratiques agricoles et les composantes environnementales qui sont susceptibles d'être modifiés par ces pratiques agricoles. Girardin et al. [102] proposent d'utiliser une matrice agri-environnementale qui permet de rendre transparents les choix effectués, qui peuvent se faire à partir de discussions avec différents acteurs et experts impliqués dans la démarche d'évaluation. Une première matrice agri-environnementale avait été mise au point en grandes cultures. Pour chaque système de culture étudié, il est nécessaire d'adapter cette matrice aux spécificités du système étudié et de ses impacts sur les composantes de l'environnement. La matrice agri-environnementale se présente sous forme d'un tableau à double entrée croisant les différents composantes de l'environnement avec les pratiques agricoles susceptibles d'être mises en oeuvre par l'agriculteur sur une parcelle. La matrice est présentée sur la Table 5.1.

Les pratiques agricoles illustrées en colonnes regroupent principalement la gestion des intrants (énergie, azote et produits phytosanitaires) et la gestion de l'espace et des stocks (matière organique et couverture du sol). L'environnement est décomposé en plusieurs composantes : l'eau, l'air et le sol, les ressources non renouvelables (les matières fossiles, les matières premières) et les ressources biotiques (la faune auxiliaire) et le paysage.

Chaque intersection entre une pratique agricole et une composante de l'environnement est appelée "module d'évaluation" et correspond à l'impact environnemental potentiel d'une pratique agricole sur une composante de l'environnement. Lorsque la pratique agricole ne présente aucun impact significatif sur l'environnement, la case d'intersection reste vide. A l'inverse, s'il a été montré que la pratique agricole peut causer un problème environnemental, alors l'impact environnemental est notifié dans la matrice.

Deux sortes d'indicateurs peuvent être élaborées à partir de la matrice agri-environnementale :

- Un indicateur agri-environnemental qui évalue l'impact d'une pratique agricole sur l'ensemble des composantes de l'environnement, il représente une colonne entière de la matrice.

			Pratiques culturales									
			Facteurs de production							Espace		
			Produits phytosanitaires	Azote	Phosphore	Irrigation	Énergie	Matière Organique	Travaux du sol	Assolement	Couverture de sol	Éléments non productifs
Milieu	Eau	Surface	×		×	×			×		×	×
		Profondeur	×	×	×	×						
	Air	Qualité	×	×			×					
	Sol	Quantité				×		×	×		×	×
		Structure						×	×			
		Qualité			×							
	Ressources non renouvelables				×							
	Faune		×			×			×	×	×	×
	Paysage									×		×

TABLE 5.1 – Matrice agri-environnementale servant à la construction des indicateurs de la méthode Indigo

Chaque colonne est considérée comme un indicateur composite (par exemple l'indicateur des produits phytosanitaires I_{Phy}, 1ère colonne de la Table 5.1, qui correspond à l'impact des pesticides sur l'ensemble des composantes du milieu eau de surface, eau de profondeur et air [176]).

– Un indicateur d'impact environnemental qui évalue l'impact de toutes les pratiques agricoles sur une seule composante du milieu, il représente une ligne entière de la matrice (par exemple l'indicateur "qualité des eaux de surface" [102] représentée par la première ligne de la Table 5.1).

La matrice agri-environnementale nous permet d'acquérir une vision globale des pratiques agricoles susceptibles d'avoir un impact sur l'environnement [102].

5.3.2 Construction des indicateurs

Les utilisateurs potentiels de ces indicateurs sont les gestionnaires de l'environnement, les techniciens, les agriculteurs et les conseillers agricoles. L'indicateur doit donc être lisible et simple à utiliser. Un des principes qui guide les indicateurs de la méthode Indigo est que les données nécessaires à leurs calculs sont fournies par l'agriculteur. Ces données servent à renseigner les variables qui composent l'indicateur. Pour ce qui est échelle spatiale, les indicateurs de la méthode Indigo se calculent à la parcelle et ils peuvent être pondérés *au prorata* des surfaces des parcelles pour obtenir une moyenne au niveau de l'exploitation agricole. La suite de l'étape de construction proprement dite de l'indicateur est le processus d'agrégation de ces variables (voir section 5.4).

5.3.3 Choix de la référence

Tout indicateur de la méthode Indigo correspond à une valeur positionnée par rapport à une référence ou un seuil. Cette référence peut être une norme, une valeur fixée scientifiquement ou issue d'un compromis entre experts [33].

5.3.4 Test de sensibilité

Les tests de sensibilité permettent de vérifier la capacité d'estimer le poids respectif des variables composant l'indicateur. Le test est réalisé en faisant varier une variable et en observant l'impact de ces variations sur la valeur finale de l'indicateur. Il est possible de faire des tests de sensibilité pour tous les niveaux d'agrégation d'un indicateur. Cette phase est indispensable pour connaître le comportement de l'indicateur et fournit des informations sur le niveau de précision requis pour chacune des variables qui composent l'indicateur. Dans certains cas, il peut conduire à supprimer une variable qui aurait un très faible poids à moins de la conserver pour des raisons pédagogiques [158, 170, 178].

5.3.5 Validation de l'indicateur

La validation de l'outil de diagnostic sert principalement à obtenir un outil fondé scientifiquement tout en fournissant une information concise et simple qui reflète la réalité de terrain [32]. La validation des indicateurs fait partie intégrante de leur construction. Bockstaller et al. [32] présentent 3 niveaux de validation des indicateurs : la validation de la conception, la validation de sortie et la validation d'usage [29, 32, 100].

5.4 Données, mode de calcul et exemples des indicateurs

Les données utilisées pour calculer les indicateurs de la méthode Indigo sont disponibles sur l'exploitation ou accessibles régionalement, elles sont fournies par les agriculteurs et les techniciens. Des données proviennent de différentes sources d'informations. Une source d'information peut être un expert, ressource bibliographique, base de données... Les données de l'indicateur peuvent être classées en deux types :

- les variables enregistrées par les agriculteurs ou les techniciens. Par exemple, la quantité de pesticide utilisée dans le champ ou la surface de la parcelle.
- les variables dont les valeurs proviennent de plusieurs sources d'informations ou elles peuvent être la sortie d'un modèle ou la conclusion d'une règle issue d'une arbre de décision, par exemple, le potentiel de lessivage est la sortie d'un système de règle de décision dont les prémisses représentent les variables taux de matière organique, texture de sol et profondeur : ***Si*** *le taux de matière organique est inférieure à* 3% *et le sol est filtrant et profond* ***Alors*** *le potentiel de lessivage vaut* 1.

Il existe plusieurs modes de calcul.

1. les indicateurs simples qui se calculent à l'aide d'équations mathématiques (un rapport ou une multiplication entre leurs variables d'entrées) comme par exemple l'indicateur matière organique I_{MO}. Ce dernier évalue l'impact des pratiques culturales (succession, travail de sol, gestion de résidus, amendements organiques) sur la quantité de matière organique dans le sol. Il est calculé au niveau de la parcelle comme étant le rapport de deux facteurs A_X et A_R, il s'écrit sous la forme :

$$I_{MO} = 7.\frac{A_X}{A_R}$$

où A_X représente les apports moyens en humus pour les quatre dernières cultures et A_R représente les apports nécessaires pour maintenir le sol à long terme à une teneur d'équilibre qui soit satisfaisante.

2. les indicateurs composites qui se calculent à partir de sous-indicateurs, comme par exemple, l'indicateur azote I_N. Cet indicateur a pour objectif d'aider l'agriculteur à adapter sa fertilisation azotée aux principes de la fertilisation raisonnée et sa gestion de l'azote durant l'interculture, de manière à minimiser ces pertes. I_N est exprimé comme étant le *minimum* de trois sous-indicateurs I_{NH3}, I_{N_2O} et I_{NO_3} et on écrit :

$$I_N = \min(I_{NH_3}, I_{N_2O}, I_{NO_3})$$

Avec l'indicateur I_{NH_3} évalue le risque potentiel de la gestion des fertilisants azotés sur le sol et la biodiversité de manière indirecte par la volatilisation de l'ammoniac. L'indicateur I_{N_2O} évalue le risque potentiel des pratiques culturales sur la qualité de l'air par des émissions du protoxyde d'azote. L'indicateur I_{NO_3} qui évalue l'impact des pratiques culturales sur la qualité des eaux souterraines au travers du lessivage des nitrates.

3. Les indicateurs qui se calculent en utilisant des systèmes d'inférences floues comme, par exemple, les indicateurs évaluant l'impact des produits phytosanitaires [176]. Un système d'inférence flou est formé de trois blocs (voir figure 5.1) : la fuzzification, l'inférence et la défuzzification.

La fuzzification est la phase de transformation des valeurs numériques en terme linguistique et le calcul des fonctions d'appartenance aux classes correspondantes (partition choisie pour chacune des variables). Les fonctions triangulaires et trapézoïdales sont, en général, les plus utilisées. Les fonctions d'appartenance utilisées dans les indicateurs phytosanitaires sont sinusoïdales. Le deuxième bloc est l'inférence qui met en évidence la relation entre les variables d'entrées et la variable de sortie et il constitue la base de règles floues (ou la base de connaissance) [12]. Les bases de règles utilisées pour calculer les indicateurs phytosanitaires sont des règles données par les experts. Le troisième bloc est la défuzzification qui consiste, si nécessaire, d'inférer une valeur à partir de l'agrégation des règles. Plusieurs méthodes de défuzzification sont proposées [141].

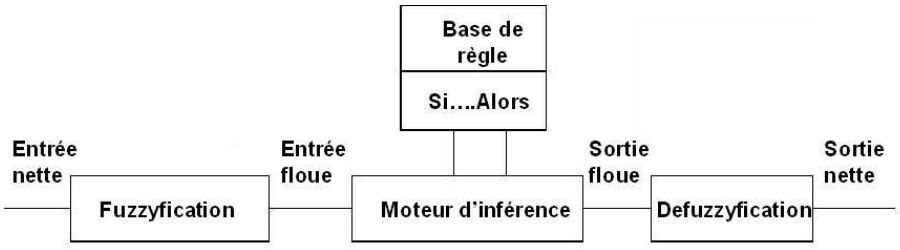


FIGURE 5.1 – Un système d'inférence floue

L'indicateur des produits phytosanitaires noté I_{phy} qui évalue l'impact de l'utilisation des pesticides et les pratiques culturales de l'agriculteur sur les composantes du milieu (eaux souterraines, eaux superficielles et air). L'indicateur I_{phy} est une agrégation de trois

12. Une règle floue est une règle de la forme *Si j'ai une telle situation alors j'ai une telle conclusion.* La situation est appelée prémisse, entrée, antécédant de la règle et elle est définie par des relations de la forme x est A pour chacune des composantes de l'entrée. La partie conclusion de la règle est appelée sortie de la règle. Les sorties dépendent de la base de règles, de l'opérateur d'agrégation et de défuzzification

sous-indicateurs I_{eso}, I_{esu}, I_{air} avec la dose de pesticide utilisée dans le champ. Les sous-indicateurs I_{eso}, I_{esu}, I_{air} se calculent aussi en utilisant des systèmes d'inférences floues. On présentera le mode de calcul détaillé de l'indicateur des produits phytosanitaires et ses sous-indicateurs (voir Section 5.6.1).

5.5 Indicateurs et imperfection

Plusieurs travaux existent sur le traitement, la manipulation et la représentation des données imparfaites dans le domaine de l'environnement. Bardossy et al [36] ont simulé l'écoulement dans un milieu non saturé à l'aide des règles de types floues. Dou et al. [66] ont utilisé des nombres flous pour représenter l'imprécision liée aux paramètre d'un modèle d'écoulement en régime permanent. C. Freissinet [92] a utilisé les sous-ensembles flous pour la représentation des imprécisions dans la modélisation du devenir des produits phytosanitaires dans le sol. Baudrit et al. [17] ont traité l'imperfection pour l'évaluation des risques liés aux sols pollués.

Le cadre général qui regroupe les outils de représentation de l'incertain est le cadre des probabilités imprécises [174]. Ce cadre permet de combiner différents types des connaissances et donc de propager les différentes types d'imperfection associées. Dans ce cadre, Guyonnet et al. [107] ont développé une méthode hybride combinant le calcul possibiliste avec la méthode de Monte Carlo utilisé dans [17].

Les indicateurs environnementaux sont calculés à partir des informations accessibles sur le terrain, des mesures de laboratoire et d'autres données difficiles à calculer qui sont choisies par les experts [176]. La précision de ces informations peut être très variable. Ces informations sont entachées d'imperfection. Elles sont imprécises, incertaines et/ou incomplètes. Dans ce chapitre, nous nous intéressons à l'indicateur phytosanitaire et ses composants. Ces indicateurs se calculent à partir de plusieurs informations numériques et symboliques. Les informations numériques sont des mesures de terrain (la distance de la position d'application d'un pesticide à un cours d'eau), des valeurs calculées dans les laboratoires (analyse de sol) et des valeurs estimées par des experts à partir de sources d'informations (comme les caractéristiques des pesticides). Les informations symboliques sont les informations données par l'expert (ou le technicien du terrain) sous forme de terme linguistique (la profondeur de sol : profond, moyen...). Une imperfection supplémentaire est due aux choix des conclusions des règles dans les systèmes d'inférences floues ainsi qu'aux choix des partitions floues des données d'entrées de ces systèmes (classes favorable et défavorable).

La prise en compte de l'imperfection des indicateurs agri-environnementaux n'était jamais traitée par les chercheurs qui proposent ces indicateurs. Mais le besoin de donner une précision à leur outil et ainsi donner plus de flexibilité à l'agriculteur pour choisir ces pratiques a mené les chercheurs à traiter cette problématique. Le traitement de l'imperfection permet de garder la qualité scientifique de leur outil et il permet aussi d'avoir plus de flexibilité aux utilisateurs de cet indicateur.

Dans ce travail, nous nous intéressons à l'indicateur "pesticides" ou l'indicateur des produits phytosanitaires [13]. Ce dernier permet d'évaluer le risque des produits phytosanitaires sur l'environnement.

13. Un produit phytosanitaire est un ensemble des produits chimiques (pesticides) destiné à protéger les plantes cultivées des maladies et des insectes. Un produit phytosanitaire est composé d'une ou plusieurs matière active. Un pesticide est un produit chimique employé contre les parasites animaux et végétaux des plantes. Il existe trois types de pesticides : les fongicides, les herbicides et les insecticides. Un pesticide est appelé aussi matière active ou substance active.

5.6 Indicateur des produits phytosanitaires de grandes cultures

5.6.1 Définition de l'indicateur I_{phy} et son objectif

L'indicateur phytosanitaire I_{phy} est un indicateur agro-écologique d'impact calculé sur la parcelle.

L'indicateur phytosanitaire a pour objectif de fournir un outil de raisonnement du choix des matières actives et des programmes de traitements, outil qui permet d'évaluer les risques environnementaux de l'application de chaque substance active et de comparer les programmes de traitements. L'indicateur I_{phy} est destiné avant tout aux techniciens et aux agriculteurs, pour leur permettre d'établir un diagnostic sur les pratiques de l'année et d'effectuer des simulations en vue d'apporter des améliorations aux programmes de traitement prévus.

Malgré tous les progrès de la technologie, une grosse partie du produit appliqué ne touche pas sa cible et va dans l'environnement (sol, eau, air, etc) où il peut passer d'un compartiment à l'autre (du sol vers la rivière par ruissellement ou par érosion, de l'air vers l'eau via les pluies, etc). En plus les matières se révèlent plus ou moins toxiques pour divers organismes de ces compartiments de l'environnement.

L'indicateur I_{phy} se limite à quatre types de risque correspondant à quatre "sous-indicateurs" :

- Un risque d'entraînement vers les eaux de profondeur (eaux souterraines) par lessivage, qui est aggravé si la substance active est toxique pour l'homme (de manière chronique). Cette toxicité est évaluée par la dose journalière admissible (DJA).
- Un risque d'entraînement vers les eaux de surface (eaux superficielles) par ruissellement/érosion et/ou par dérive. Ce risque est aggravé si la substance est toxique pour les organismes aquatiques (poisson, daphnies, algues).
- Un risque de propagation vers l'air par volatilisation qui est aggravé si la substance active est toxique pour l'homme (la toxicité est mesurée par la DJA).
- Le dernier risque appelé "dose" est simplement lié à la quantité de la substance active. Plus la dose est élevée, plus le risque pour l'environnement est élevé.

Les trois risques liés aux eaux de surface, de profondeur et l'air sont indépendantes de la dose appliquée. L'indicateur phytosanitaire se calcule pour une matière active en première étape, ensuite, il est calculé pour un programme de traitements composé de plusieurs matières actives. Dans la suite, nous détaillerons, pour une matière active, le mode de calcul de l'indicateur ainsi que de ses sous-indicateurs. Ceux-ci seront agrégés au dernier risque lié à la dose pour donner l'indicateur I_{phy} pour une matière active. La figure 5.2 résume le calcul de l'indicateur pour une matière active et aussi pour un programme de traitements de plusieurs matières actives.

5.7 Données et échelles de calcul

L'indicateur phytosanitaire se calcule au niveau de la parcelle pour une matière active et pour un programme de traitements.

Certaines informations sont de mesures de terrain (donnée par l'agriculteur comme la positon d'application de la matière active sur le sol). D'autres informations sont issues d'un consensus entre experts, comme par exemple, le potentiel de lessivage. Les informations multi-sources provenant de plusieurs sources d'informations (bases de données, experts, bibliographie, etc.) par exemple les caractéristiques physico-chimiques des pesticides. Pour l'indicateur I_{Phy}, ces valeurs sont fixées dans une base de données construite à partir de différentes bases de données françaises

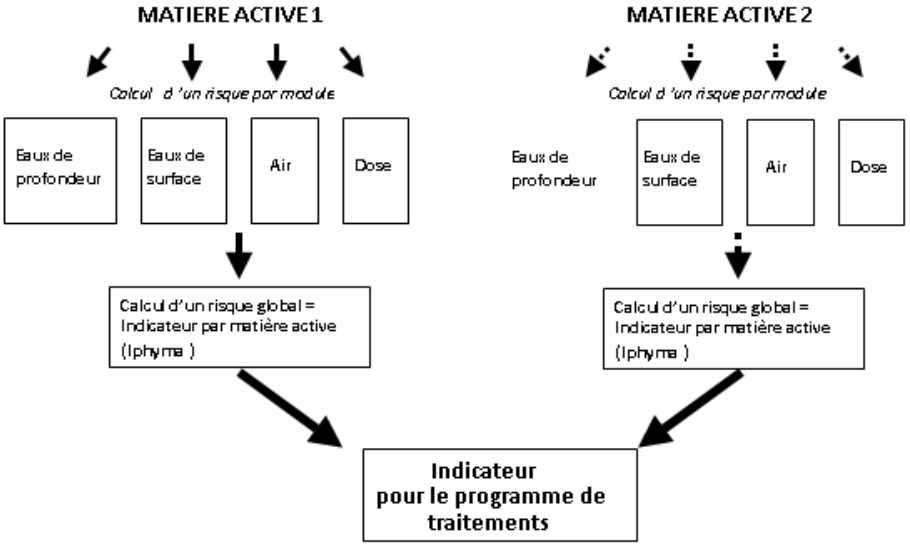


FIGURE 5.2 – I_{phy} pour une matière active et un programme de traitement

et internationales ("Agritox" de l'INRA [14], base de données du comité (version de 1998), base de données américaine (ARS sur internet 1995 [15]), base de données néerlandaise (RIVM, 1994) [16], et le *Pesticide Manual* d'Angleterre [33]).

Les variables d'entrées de l'indicateur phytosanitaire sont données dans la Table 5.2 et elles sont décrites dans la suite.

5.7.1 Caractéristiques de la matière active

Le temps de demie-vie au champ

Le temps de demie-vie au champ $DT50$ est le temps mis par une substance active pour perdre la moitié de son activité pharmacologique, physiologique ou radioactive. Il permet également d'évaluer la persistance de la substance active et d'estimer le risque d'accumulation dans l'environnement dans les conditions réelles. Il s'exprime en *jour*.

Le domaine de variation est divisé en trois classes : favorable, défavorable et la classe floue (valeurs qui ne sont pas complètement favorables ni complètement défavorables). Les valeurs en dessous d'1 jour de $DT50$ sont favorables, au delà de 30 jours elles sont défavorables. Entre les deux valeurs la fonction d'appartenance est donnée par :

$$f(DT50) \quad = \quad 0.5 + 0.5\cos\left(\frac{3.14(DT50-1)}{29}\right) \tag{5.1}$$

Le coefficient de partage carbone organique-eau

Le coefficient de partage carbone organique-eau *koc* est le rapport entre la quantité absorbée de la matière active par unité de poids de carbone organique du sol et la concentration de cette même matière organique en solution aqueuse à l'équilibre. Cette caractéristique caractérise

14. http ://www.dive.afssa.fr/agritox/index.php
15. http ://www.ars.usda.gov/Services/docs.htm ?docid=14199
16. http ://www.rivm.nl/en/

Variables	Unité	Dose	I_{eso}	I_{esu}	I_{air}
Variables liées au pesticide					
DT50	jours			×	
GUS	-		×		
KH		-			×
DJA	$mg.kg^{-1}$		×	×	
Aquatox	$mg.l^{-1}$			×	
Variables liées au milieu					
Potentiel de lessivage	-		×		
Potentiel de ruissellement	-			×	
Potentiel de dérive	-			×	
Variables liées au conditions d'application					
Dose de pesticide	$g.ha^{-1}$	×			
Position d'application	-		×	×	×

TABLE 5.2 – Variable entrant dans le calcul de chacun des indicateurs phytosanitaires

indirectement la mobilité de la substance active dans le sol. Elle est mesurée en $L.kg^{-1}$ et elle est considérée comme un paramètre dans le calcul de l'indicateur phytosanitaire des eaux de surface. Plus le coefficient *koc* est grand, plus la substance est "liée" au sol, au contraire si *koc* est petit, le pesticide a tendance à se trouver dissout dans l'eau. Aucune partition des valeurs de *koc* est proposée dans la littérature.

L'indice d'ubiquité de Gustafson GUS

La variable GUS permet de distinguer entre les pesticides lessivés et ceux qui ne le sont pas. Elle est disponible pour toutes les substances actives. Elle se calcule à partir des variables $DT50$ et *koc* suivant l'équation :

$$GUS = \log_{10}(DT50)(4 - \log_{10}(Koc)) \tag{5.2}$$

Le domaine des valeurs de GUS est divisé en trois classes. Les valeurs plus petites que 1.8 sont favorables, et celles plus grandes que 2.8 sont défavorables et entre les deux les valeurs, la fonction d'appartenance est donnée par :

$$f(GUS) = 0.5 + 0.5\cos(3.14(GUS - 1.8)) \tag{5.3}$$

La dose journalière admissible

La dose journalière admissible DJA (appelée aussi la dose journalière acceptable ou tolérable) est la quantité d'une matière active, par kilo de poids corporel, que pourrait absorber une personne quotidiennement durant sa vie, sans que cela lui pose des problèmes de santé. Elle s'exprime en $mg/kg/jour$. Les valeurs de DJA sont décomposées en trois classes. Si $\log_{10}(DJA) < \log_{10}(0.0001)$ alors DJA est favorable et si $\log_{10}(DJA) > \log_{10}(1)$ alors DJA est défavorable, sinon les degrés d'appartenance des valeurs appartenant à la classe floue se calculent en utilisant la fonction :

$$f(DJA) = 0.5 + 0.5\sin(3.14\frac{\log_{10}(DJA) - \log_{10}(0.0001)}{\log_{10}(1) - \log_{10}(0.0001)} - 0.5) \tag{5.4}$$

La constante de Henry

La constante de Henry KH décrit la solubilité physique d'un gaz dans l'eau. Elle correspond au rapport de la pression de vapeur sur la solubilité. Elle n'a pas d'unité. Si $\log_{10}(kh) < \log_{10}(0.00000265)$ alors KH est favorable, si $\log_{10}(KH) > \log_{10}(0.000265)$ alors KH est défavorable sinon la fonction d'appartenance à la classe floue est donnée par :

$$f(KH) = 0.5 + 0.5\cos(3.14\frac{\log_{10}(KH)-\log_{10}(0.00000265)}{\log_{10}(0.000265)-\log_{10}(0.00000265)}) \quad (5.5)$$

La toxicité aquatique *Aquatox*

La toxicité aquatique *aquatox* est le risque de toxicité pour les organismes aquatiques. C'est la toxicité maximale entre la toxicité vis-à-vis des algues, celle vis-à-vis des poissons et celle vis-à-vis des daphnies. Elle s'exprime en $mg.l^{-1}$. Si $\log_{10}(aquatox) < \log_{10}(0.01)$ alors *aquatox* est favorable, si $\log_{10}(aquatox) > \log_{10}(100)$ alors *aquatox* est défavorable, sinon *aquatox* appartient à la classe floue et sa fonction d'appartenance est donnée par :

$$f(aquatox) = 0.5 + 0.5\sin(3.14\frac{\log_{10}(aquatox)-\log_{10}(0.01)}{\log_{10}(100)-\log_{10}(0.01)} - 0.5)) \quad (5.6)$$

5.7.2 Variables liées au milieu

Le potentiel de lessivage

Le potentiel de lessivage estime la quantité de pesticide lessivée dans le sol. Le potentiel de lessivage dépend de la texture de sol (filtrant ou non filtrant lié à la nature de sol), le taux de matière organique (pauvre, moyen, riche), le profondeur de sol et le pH. Les valeurs du potentiel de lessivage obtenues à partir des conclusions de règles de décision, qui varient dans $[0, 1]$. Elles sont choisies par un expert puisque la valeur réelle du potentiel de lessivage n'est pas facile à trouver. L'arbre de décision est donné sur la Figure 5.3. Le potentiel de lessivage est une variable d'entrée de l'indicateur phytosanitaire des eaux souterraines I_{eso}. La valeur 0 est favorable et la valeur 1 est défavorable. La fonction d'appartenance du sous-ensemble flou est donnée par :

$$f(lessivage) = 0.5 + 0.5\cos(3.14 lessivage) \quad (5.7)$$

Le potentiel de ruissellement

Le potentiel de ruissellement est la quantité de pesticide qui peut ruisseler. Il dépend de la texture du sol, de la topographie de la parcelle et de la pente. Une faible pente suffit pour permettre un ruissellement, à la différence de l'érosion proprement dite qui est fortement liée à la pente. L'état de surface et notamment la présence d'une croûte de battance, ou des problèmes d'infiltration dus à la présence d'hydromorphie à faible profondeur, jouent un rôle important dans la détermination du ruissellement. Les valeurs du potentiel de ruissellement sont données dans la Table 5.7.2. Elles sont choisies par des experts dans $[0, 1]$. La valeur 0 est une valeur favorable et 1 est la limite défavorable. La fonction d'appartenance est définie par :

$$f(ruissellement) = 0.5 + 0.5\cos(3.14 ruissellement) \quad (5.8)$$

	Sableux	Limoneux Battance		Argileux Hydromorphie	
		Non	Oui	Non	Oui
Nulle	0	0	0	0	0.25
Faible (0-2 %)	0.1	0.25	0.4	0.25	0.5
Moyenne (2- 5 %)	0.3	0.5	0.7	0.5	0.75
Forte ($>$ 5 %)	0.6	0.75	1	0.75	1

TABLE 5.3 – Les valeurs du potentiel de ruissellement

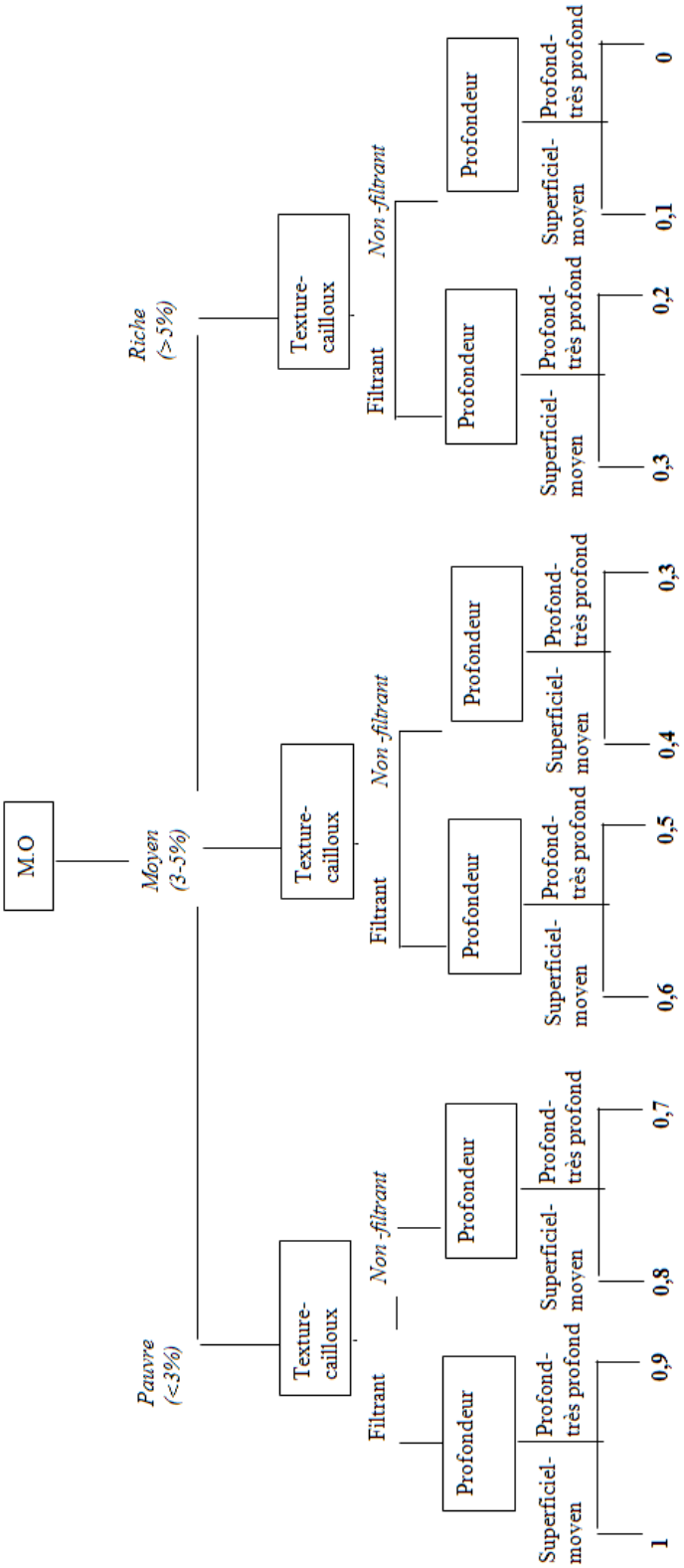


FIGURE 5.3 – Arbre de décision du potentiel de lessivage

Le potentiel de dérive

Lorsque l'agriculteur pulvérise un produit phytosanitaire sur une parcelle, ce dernier se retrouve en partie sur la plante et le reste n'atteint pas sa cible et se retrouve dans le sol ou dans l'air. Par ailleurs, toute la quantité du produit phytosanitaire déposée sur la plante ne va pas avoir le temps d'agir. En effet, une partie peut ruisseler le long des feuilles de la plante et se retrouver dans le sol. Un troisième phénomène s'ajoute aux deux cités précédemment : la dérive aérienne et le dépôt de produits phytosanitaires plus loin sur la parcelle ou sur une autre parcelle, parfois à grande distance [165]. Le potentiel de dérive estime la quantité de produit susceptible de se retrouver dans le cours d'eau. Le potentiel de dérive dépend des paramètres de la parcelle qui favorisent la dérive des produits phytosanitaires : la distance au cours d'eau et le mode de traitement du produit phytosanitaire. Des valeurs estimées par des experts sont données dans la Table 5.7.2 . Elles se présentent comme des valeurs entre 0 et 1. La valeur 0 (respectivement 1) est favorable (défavorable). La fonction dans l'intervalle de valeurs $]0,1[$ est donnée par l'équation :

$$f(derive) \quad = \quad 0.5 + 0.5 * cos(3.14 derive) \tag{5.9}$$

	Distance à la rivière en mètre			
Méthodes d'application	< 3	$[3,6[$	$[6,12[$	≥ 12
Traitement en plein	1	0.7	0.3	0
Traitement sur le rang	0.5	0.3	0.1	0

TABLE 5.4 – Les valeurs du potentiel de dérive

5.7.3 Variables liées aux conditions d'application

La position d'application

Cette variable traite la façon d'appliquer le produit phytosanitaire au champ. La valeur de la position d'application est calculée pour les trois sous-indicateurs I_{esu}, I_{eso} et I_{air} à partir de la couverture de sol :

$$position \quad = \quad \frac{couverture}{100} \tag{5.10}$$

Le lieu d'application du produit phytosanitaire (sur le plante, sur le sol, ou dans le sol) joue un rôle significatif sur le lessivage, le ruissellement et la dérive. En effet l'application de produit sur la surface de sol est défavorable à la qualité des eaux de surface car il est alors possible de trouver une forte corrélation entre les concentrations en substances actives dans l'eau de ruissellement et dans les 10 premiers millimètre du sol tandis que l'application de produit phytosanitaire dans le sol par injection ou par une incorporation suite au traitement est favorable [130]. Une forte couverture du sol par la culture permet de réduire la quantité de produit phytosanitaire qui atteint le sol et donc de réduire le risque de lessivage et la position de l'application a les mêmes effets sur l'air que sur le ruissellement. Tandis que si le produit est incorporé ou injecté dans le sol, alors la position d'application est défavorable pour les eaux souterraines et comme pour l'air et l'eau de surface.La fonction d'appartenance, pour la classe floue (ensemble de valeurs dans $]0,1]$), est définie par :

$$0.5 + 0.5 \sin(3.14(position - 0.5)) \tag{5.11}$$

La couverture de sol

La couverture de sol par les plantes est un des facteurs principaux de la limitation de l'érosion et du lessivage. Elle dépend de la culture présente et de la date de traitement. Comme par exemple, pour un traitement du produit au début de mois de juillet sur une culture de maïs, le sol est couvert à 75%.

La dose

La dose est la quantité de produit utilisée au champ. Elle est donnée par les techniciens et les agriculteurs. Le domaine des valeurs de la variable dose est divisé en trois classes. Si $\log_{10}(dose) < \log_{10}(10)$ alors la dose est favorable, si $\log_{10}(dose) > \log_{10}(10000)$ alors la dose est défavorable pour l'environnement sinon la valeur appartient à la classe floue dont la fonction d'appartenance est définie par :

$$f(dose) \quad = \quad 0.5 + 0.5\cos(3.14\frac{\log_{10}(dose)-\log_{10}(10)}{\log_{10}(10000)-\log_{10}(10)}) \tag{5.12}$$

5.8 Indicateur des produits phytosanitaires vis-à-vis des eaux souterraines

L'indicateur phytosanitaire vis-à-vis des eaux souterraines I_{eso} est un indicateur qui évalue le risque des pesticides sur les eaux souterraines. Il est calculé au niveau de la parcelle pour une matière active en utilisant un système inférence floue ayant GUS, *position*, DJA et *potentiel de lessivage*. La base de règle est donnée sur l'arbre de décision de la figure 5.4. Le nombre de règles

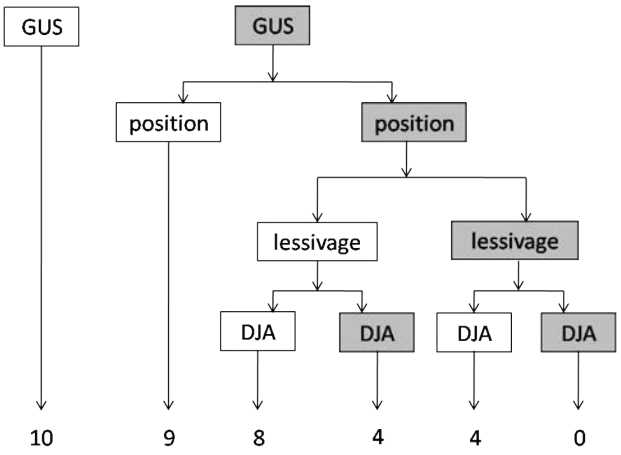


FIGURE 5.4 – Arbre de décision de l'indicateur I_{eso}

de la base est 16. Les valeurs favorables et défavorables servant au calcul de la valeur de vérité sont données dans la section 5.7 et on a :

$$I_{eso} \quad = \quad \frac{\sum_{i=1}^{16} w_i x_i}{\sum_{i=1}^{16} w_i} \tag{5.13}$$

où w_i est la valeur de vérité qui est la conjonction des fonctions d'appartenance aux classes favorable et défavorable pour la règle i et x_i est sa conclusion.

5.9 Indicateur des produits phytosanitaires vis-à-vis des eaux de surface

L'indicateur phytosanitaire vis-à-vis des eaux de surface évalue le risque d'utilisation des pesticides sur les eaux superficielles puisque les pesticides se trouvent dans l'eau de surface via ruissellement ou dérive. I_{esu} est calculé en utilisant un système d'inférence floue où les variables d'entrées sont : $DT50$, *position*, *aquatox*, *potentiel de ruissellement* et *potentiel de dérive*. Les fonctions d'appartenance sont données dans la section 5.7 et la base de connaissance est formée de 32 règle données sur l'arbre de décision de la Figure 5.5. La valeur finale de I_{esu} est donnée par :

$$I_{esu} = \frac{\sum_{i=1}^{32} w_i x_i}{\sum_{i=1}^{32} w_i} \tag{5.14}$$

où w_i est la valeur de vérité qui est la conjonction des fonctions d'appartenance aux classes favorable et défavorable pour la règle i et x_i est sa conclusion.

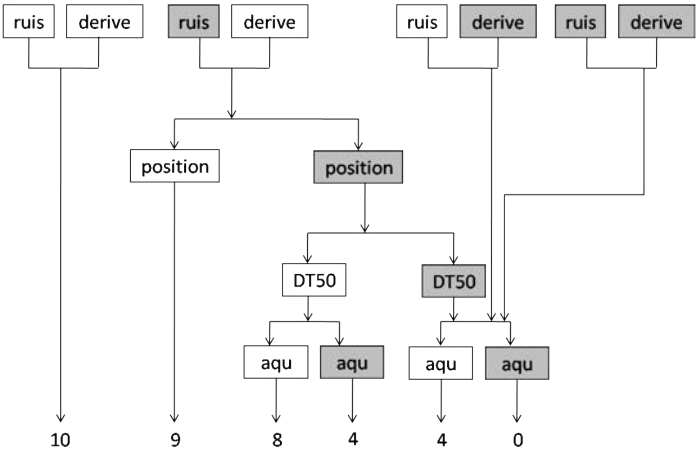


FIGURE 5.5 – Arbre de décision de l'indicateur I_{esu}

5.10 Indicateurs des produits phytosanitaires vis-à-vis de l'air

L'indicateur phytosanitaires vis-à-vis de l'air, I_{air}, évalue le risque de l'utilisation des produits phytosanitaires sur l'air puisque les pesticides se déplacent dans l'air via la volatilisation (représenté par la variable constante de Henri KH). Comme les autres indicateurs phytosanitaires I_{eso} et I_{esu}, l'indicateur I_{air} se calcule en utilisant un système d'inférence floue où les variables d'entrées sont $DT50$, KH, DJA et *Position*. Avec les fonctions d'appartenance données dans la section 5.7 et la base de règles est formée de 16 règles (voir figure 5.6).

La valeur finale de I_{air} est une moyenne pondérée comme suit :

$$I_{air} = \frac{\sum_{i=1}^{16} w_i x_i}{\sum_{i=1}^{16} w_i} \tag{5.15}$$

où w_i est la valeur de vérité qui est la conjonction des fonctions d'appartenance aux classes favorable et défavorable pour la règle i, et, x_i est sa conclusion.

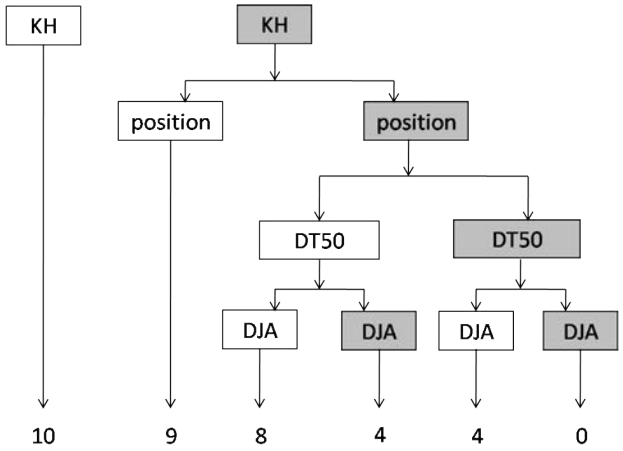


FIGURE 5.6 – Arbre de décision de l'indicateur I_{air}

5.11 Indicateur des produits phytosanitaires pour une matière active

L'indicateur phytosanitaire évalue le risque des produits phytosanitaires sur tous les compartiments de milieu (eau et air). Les variables d'entrée du système d'inférence de I_{phy} sont : (I_{eso},I_{esu}, et I_{air}) et la variable *dose*. La base des règles de I_{phy} est formée de 16 règles données sur l'arbre de décision de la figure 5.7. Les fonction d'appartenance de la variable dose est donnée

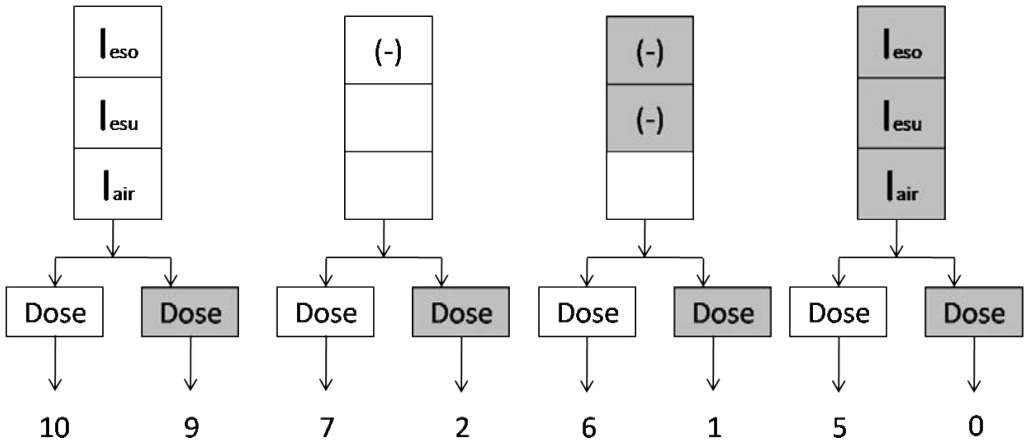


FIGURE 5.7 – Arbre de décision de l'indicateur I_{phy} où (*) désigne un des indicateurs I_{eso}, I_{esu} ou I_{air}

dans la section 5.7. Les limites favorables (respectivement défavorable) pour les trois indicateurs (I_{eso},I_{esu}, et I_{air}) sont 10 (resp. 0). Les fonctions d'appartenance sont données respectivement pour les trois indicateurs I_{eso},I_{esu}, et I_{air} par les équations (5.16), (5.17) et (5.18).

$$f(I_{eso}) = 0.5 + 0.5\sin(3.14\frac{I_{eso}}{10} - 0.5) \tag{5.16}$$

$$f(I_{esu}) = 0.5 + 0.5\sin(3.14\frac{I_{esu}}{10} - 0.5) \tag{5.17}$$

$$f(I_{air}) = 0.5 + 0.5\sin(3.14\frac{I_{air}}{10} - 0.5) \tag{5.18}$$

La valeur de I_{phy} est une moyenne pondérée donnée par :

$$I_{phy} = \frac{\sum_{i=1}^{16} w_i x_i}{\sum_{i=1}^{16} w_i} \tag{5.19}$$

où w_i est la valeur de vérité qui est la conjonction des fonctions d'appartenance aux classes favorable et défavorable pour la règle i, et, x_i est sa conclusion.

5.12 Exemple de la matière active *sulcotrione*

Les quatre indicateurs I_{phy}, I_{eso}, I_{esu} et I_{air} se calculent de la même manière pour la matière active *sulcotrione*. Dans cet exemple, nous détaillerons que le calcul de l'indicateur I_{eso}.

5.12.1 Données

Données liées à la matière active

La Table 5.5 présente les valeurs des variables de la matière active *sulcotrione* fournies par les experts. Ces valeurs sont choisies dans un ensemble de valeurs proposées par plusieurs sources d'informations. Les experts favorisent les valeurs proposées par la base de données française Agritox puisqu'elles sont calculés à partir des contextes similaires au contexte de notre étude (nature de sol, matériel disponible, etc.). La variable GUS se calcule en utilisant l'équation (5.2),

$DT50$	koc	DJA
6	17	0.00005

TABLE 5.5 – Valeurs des caractéristiques $DT50$, koc et DJA de la matière active *sulcotrione*

et $GUS = 2.15$.

Données liées aux conditions d'application

L'agriculteur a appliquée le sulcotrione au mois de mars, le sol est couvert à 10%, c-à-d $couverture = 10$ et $position = 0.1$ (voir l'équation 5.10).

Données liées au milieu

La seule variable liée au milieu utilisée dans le calcul de l'indicateur phytosanitaires vis-à-vis des eaux souterraines : le potentiel de lessivage ($lessivage$). En considérant les caractéristiques du milieu précisée ici, on obtient $lessivage = 0.9$ (voir section 5.7).

5.12.2 Calcul des degrés d'appartenance

Les degrés d'appartenance aux classes pour chacune des variables d'entrée sont calculés à l'aide des équations correspondantes données dans la section 5.7. La Table 5.6 présente les degrés d'appartenance aux classes d'appartenance des variables d'entrée de l'indicateur I_{eso}.

Variable	*GUS*	*DJA*	*lessivage*	*position*
fonction d'appartenance à la classe favorable (F)	0.72	1	0.024	0.41
fonction d'appartenance à la classe défavorable (D)	0.28	0	0.976	0.59

TABLE 5.6 – Les degrés d'appartenance à la classe favorable pour les variables de I_{eso}

5.12.3 Calcul des degrés de vérité

Les degrés de vérité des règles, qui représentent les variables w_i dans les équations (5.13), sont calculés comme suit : pour tout règle i le degré de vérité est le minimum (la conjonction) des fonctions d'appartenance des variables de la règle.

Revenons à la base de règles de l'indicateur I_{eso} montrée sur la Figure 5.4 et considérons les deux règles suivantes :

1. ***Si*** *GUS est favorable et si position est favorable et si lessivage est favorable et si DJA est favorable* ***alors*** *la conclusion est* 10
2. ***Si*** *GUS est favorable et si position est favorable et si lessivage est favorable et si DJA est défavorable* ***alors*** *la conclusion est* 10

Les valeurs de vérité pour ces deux règles représentent les valeurs minimales des fonctions d'appartenance aux classes favorables (et/ou défavorable) de toutes les variables de la règle et on obtient $w_1 = \min(0.72, 1, 0.024, 0.41) = 0.024$ et $w_2 = \min(0.72, 0, 0.024, 0.41) = 0$.

5.12.4 Calcul des indicateurs

En calculant tous les degrés d'appartenance des bases de règles des indicateurs I_{eso}, à partir des arbres de décision présentées sur les figures 5.4, on obtient $I_{eso} = 8.45$. La valeur de l'indicateur I_{eso} est plus grande que 7 (la valeur de seuil de référence acceptable par l'environnement). Par conséquent, les pratiques agricoles ne posent pas de problème pour l'environnement et elles sont durables. Nous pouvons conclure que l'utilisation de la matière active sulcotrione au champ avec un sol moins couvert (10%) ne pose pas de risque pour l'environnement.

5.13 Étude des imperfections sur l'indicateurs phytosanitaires vis-à-vis des eaux souterraines

Dans cette section, nous présentons le calcul des indicateurs en tenant compte de l'imprécision des données d'entrées en particulier les données multi-sources. Les autres variables (caractéristiques de milieu et conditions d'application) sont considérées comme précises (fixes) dans cet exemple ainsi que les conclusions des règles de décision associées au calcul des indicateurs restent inchangées.

Dans un premier temps, nous modélisons l'information fournie par les sources d'informations. Dans un deuxième temps, suivant les caractéristiques des sources d'information et la méta-connaissance disponible sur les sources, nous choisissons une méthode de fusion qui s'adapte à l'exemple, puis nous construisons le treillis et nous calculons les données d'entrées des indicateurs. Enfin, nous propageons l'information résultante dans le mode de calcul des indicateurs. Les méthodes de propagation utilisées sont le calcul d'intervalles et le principe d'extension (voir Annexe B) pour les données d'entrées de I_{eso}. Le résultat final de I_{eso}, représenté par une moyenne pondérée des valeurs de vérité, est calculé en utilisant l'algorithme de Fortin (voir Section B.4.1 de l'Annexe B).

5.13.1 Modélisation des informations fournies par les sources

Les variables $DT50$, koc et DJA (intervenant dans le calcul de I_{eso}) sont fournies par plusieurs sources d'informations sous la forme d'un ensemble des valeurs disjonctifs. Par exemple, les valeurs de $DT50$ données par la base française Agritox sont : 24, 15, 74. Cette information est modélisée par une distribution de possibilités dans $\{0, 1\}$ ce qui se traduit par un intervalle dans $\mathbb{R}$: $[\min, \max]$ où min est la plus petite valeur et max est la plus grande valeur de l'ensemble des valeurs proposées par la source. Dans l'exemple de la base de données Agritox (données calculées dans de laboratoires, noté $AGXl$ dans la Table 5.7) , on obtient l'intervalle $[15, 74]$. D'autres sources d'informations délivrent l'information sous forme de deux valeurs : une valeur minimale et une valeur maximale. Comme par exemple, suivant la base de données française Agritox, les valeurs de $DT50$ calculées au champs appartiennent à l'intervalle $[2, 6]$ (noté $AGXf$ dans la Table 5.7). Les informations fournies par les sources pour la matière active sulcotrione sont présentées dans la Table 5.7.

	DT50 jour	koc L/kg
BUS	$[2, 74]$	?
$PM11$	$[15, 72]$	?
$PM12$	?	$[44, 940]$
$PM13$	?	$[44, 940]$
$INRA$	?	$[1.08, 8.98]$
$Com98$	$[2, 6]$	$[17, 160]$
$AGXf$	$[2, 6]$	$[1.08, 160]$
$AGXl$	$[15, 74]$	$[1.08, 160]$

TABLE 5.7 – Information fournie par les sources pour $DT50$ et koc

La valeur de DJA est 0.00005, cette valeur est proposée par toutes les sources. Notons que les sources d'informations proposent parfois deux intervalles ou un intervalle et une valeur plus vraisemblable que les autres. Ce type d'informations est représenté par des distributions des possibilités dont le support est l'intervalle le plus grand et le noyau représente l'intervalle des valeurs vraisemblables. Considérons l'exemple suivant : *La valeur de* $DT50$ *est entre* 2 *et* 74 *mais les valeurs entre* 15 *et* 38 *sont les plus vraisemblables.*
Cette information est représentée par une distribution de possibilités dont le support est $[2, 74]$ et le noyau est $[15, 38]$ (voir l'exemple de la Section 1.5 du chapitre 1). Les sources d'informations sont décrites dans la Table 5.8. Elles ont des origines différentes (France, Angleterre, Etats-Unis, etc.). Elles sont hétérogènes et indépendantes les unes des autres. Dans le cas où la sources d'informations ne fournit aucune valeur pour la variable, nous considérons ce cas comme une ignorance totale. Cela veut dire qu'on remplace la case vide par le symbole "?" qui représente le plus petit intervalle contenant toutes les valeurs (c-à-d l'intervalle ayant le *minimum* (respectivement *maximum* de toutes les valeurs comme borne inférieure (respectivement supérieure)).

5.13.2 Choix de mode de fusion

Aucune information n'est fournie sur la fiabilité des sources. La méthode de fusion fondée sur la notion des SMC est la plus adaptée, car elle fournit un résultat cohérent avec toutes les sources et elle ne néglige aucune source d'informations. La Table 5.9 présente les résultats de la

Base	Origine
BUS	Base de données anglaise, Angleterre
PM11	Livre, Angleterre, version 11
PM12	Livre, Angleterre, version 12
PM13	Livre, Angleterre, version 13
INRA	fiche de données INRA, France
Com98	Comité de liaison (données en 1996)
AGXf	Base de données Agritox (valeurs calculées au champ), France
AGXl	Base de données Agritox (valeurs calculées au laboratoire), France

TABLE 5.8 – Description des sources d'informations

fusion par SMC des variables liées à la matière active de l'indicateur phytosanitaire. Les SMC obtenus pour chacune des variables sont donnés dans la Table 5.9.

Variable	Résultats de la fusion	Enveloppe convexe du résultat
$DT50$	$[2,6] \cup [15,72]$	$[2,72]$
koc	$[1.08,8.98] \cup [44,160]$	$[1.08,160]$

TABLE 5.9 – Résultats de la fusion des SMC des variables de I_{eso} liées à la matière active sulcotrione

Nous utiliserons les enveloppes convexes des résultats de la fusion par SMC des variables. La variable GUS est une fonction continue de la variable $DT50$ et koc, et nous pouvons calculer son intervalle en utilisant l'extension optimale (voir Annexe B). Les valeurs des variables *position* et *lessivage* sont fixes dans cet exemple.

Pour calculer les bornes inférieure et supérieure de l'indicateur I_{eso}, nous propageons les imperfections des données d'entrée à travers la moyenne pondérée en utilisant l'algorithme de Fortin puisque les conditions d'application sont bien vérifiées. En effet, les fonctions d'appartenance des variables $DT50$ et koc sont des fonctions continues. Donc nous utilisons l'extension optimale pour obtenir les intervalles des fonctions d'appartenance pour chacune des variables GUS, DJA, *lessivage* et *position*. Ensuite, nous calculons, à partir de des valeurs de ces variables, les intervalles de degrés de vérité à l'aide du principe d'extension de la fonction *minimum*. Le résultat de l'indicateur est $[4,10]$. Cette valeur n'est ni plus petite que 7 ni plus grande que 7 puisque les valeurs de l'intervalle ne sont pas toutes inférieures (respectivement supérieures) à 7. Dans telle situation, l'évaluateur ne peut pas prendre une décision concernant les pratiques agricoles.

Nous allons utiliser la méthode détaillée dans le chapitre 3 et nous allons construire le treillis en utilisant le mode de fusion par SMC. D'abord, nous faisons le pré-traitement des données (voir Section 3.3). puis, nous construisons, avec la structure des patrons résultante le treillis des résultats de fusion par SMC des variables $DT50$ et koc.

5.13.3 Construction et interprétation du treillis

La Table 5.10 présente le contexte obtenu à partir du pré-traitement de la Table 5.7. La Figure 5.8 montre le treillis obtenu à partir de la structure de patrons présentée dans la Table 5.10. Le treillis contient 16 concepts. Une extension est présentée avec étiquetage réduit. L'intension des concepts est obtenue à partir des intentions des sous-concepts. Par exemple, l'intension du concept C_1 est $\{(DT50,[15,72]),(koc,[1.08,8.98])\}$. Mais, si les intensions de deux sous-concepts donnent des valeurs différentes pour le même attribut, alors l'union des valeurs est considérée.

Par exemple, l'intension du concept C_2 est $\{(DT50, [2,6] \cup [15,72]), (koc, [1.08, 8.98])\}$ et les intensions de ses sous-concepts sont $\{(DT50, [2,6])\}$, $\{(DT50, [15,72])\}$ et $\{(koc, [1.08, 8.98])\}$.

Les concepts qui subsument directement le concept $\perp$ (appelés aussi atome du treillis) représentent les SMC avec leurs valeurs pour chaque caractéristiques. Par exemple, pour le concept $((BUS, Com98, AGXf), (DT50, [2,6]))$, la valeur $DT50 = [2,6]$ est le SMC des valeurs fournies par les trois sources $BUS, Com98$ et $AGXf$. Autrement dit, chaque intension d'un concept du treillis représente la fusion par SMC des objets dans l'extension. Par exemple, dans le concept $C_3 = (\{(BUS, Com98, AGXf), (BUS, PM11, AGXl)\}, (DT50, [2,6] \cup [15,72]))$, l'intension $(DT50, [2,6] \cup [15,72])$ est le résultat de la fusion par SMC des sources $BUS, Com98, AGXf, PM11$ et $AGXl$. Le concept général $\top$ représente le résultat de la fusion globale de toutes les sources pour toutes les caractéristiques.

	DT50 (days)	koc (L/kg)
$\{BUS, PM12, PM13, INRA, Com98, AGXf\}$	$[2,6]$	$\emptyset$
$\{BUS, PM11, PM12, PM13, INRA, AGXl\}$	$[15,72]$	$\emptyset$
$\{BUS, PM11, INRA, AGXf, AGXl\}$	$\emptyset$	$[1.08, 8.98]$
$\{BUS, PM11, PM12, PM13, Com98, AGXf, AGXl\}$	$\emptyset$	$[44, 160]$

TABLE 5.10 – Table résultant du pré-traitement de la Table 5.7

Dans le treillis résultant, plusieurs concepts permettent de calculer l'indicateur I_{eso}. Nous ne pouvons considérer que les concepts ayant des valeurs pour les deux caractéristiques simultanément. Donc, on obtient 10 concepts valides pour lesquels il est possible de calculer I_{eso}. Par exemple, le concept $\top$ donne une valeur d'indicateur $I_{eso} = [4,10]$ (voir le calcul dans la section précédente). Cette valeur ne permet pas de prendre une décision puisque cet intervalle contient la valeur de référence 7 acceptable par l'environnement. Donc, l'indicateur I_{eso} doit se calculer avec un autre concept valide (ayant des valeurs pour $DT50$ et pour koc). Pour le concept $(\{(BUS, PM11, AGXl), (INRA, AGXf, AGXl)\}, \{(DT50, [15,72]), (koc, [1.08, 8.98])\})$, la valeur de I_{eso} vaut $[4.32, 4.32]$. Cette valeur permet de prendre une décision puisque $4.32 < 7$. Cette valeur montre que l'agriculteur ne suit pas des pratiques durables pour l'environnement (en particulier les eaux souterraines) et qu'il doit les changer.

Cependant, pour le concept $(\{(BUS, Com98, AGXf), (PM12, PM13, Com98, AGXf, AGXl)\}, \{(DT50, [2,6]), (koc, [44, 160])\})$, on obtient $I_{eso} = [9.97, 10]$. Dans ce cas, l'agriculteur ne doit pas changer son pesticide puisque la valeur de I_{eso} est plus grande que 7. Ainsi, ce concept permet de prendre aussi une décision. Notons que les autres concepts du treillis ne permettent pas non plus au décideur d'évaluer les pratiques agricoles utilisées par l'agriculteur. Pour le choix entre plusieurs sous-ensembles de sources, l'utilisateur devra faire appel à des critères extérieures (comme un classement des sources, fiabilité, etc.). Ainsi, pour justifier une décision l'évaluateur peut donner des raisons (arguments). L'agriculteur doit change de pratique quand le pesticide reste plus longtemps au champ (temps de demi-vie au champs représenté par la variable $DT50$) et sa mobilité dans le sol (représentée par la variable koc) est petite. Sinon, si le pesticide ne reste pas longtemps au champ mais il reste bien lié au sol (puisque la valeur de koc est grande) alors l'agriculteur ne doit pas changer ce pesticide.

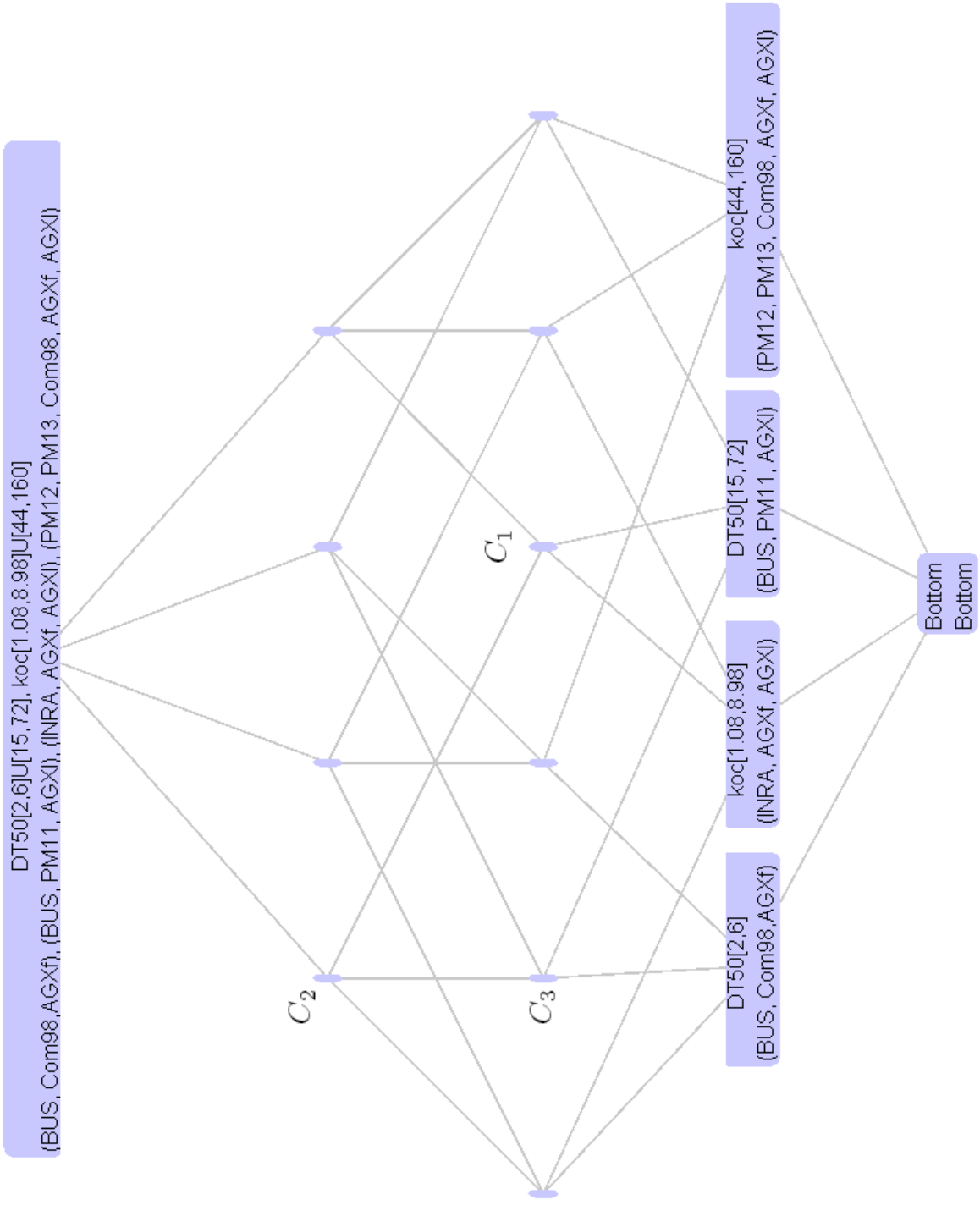


FIGURE 5.8 – Treillis de concepts des résultats de la fusion des SMC pour la variable $DT50$ et koc de la Table 5.10

5.14 Discussion

Prendre en compte l'incertitude dans l'évaluation environnementale permet de donner plus de choix aux agriculteurs dans le choix de leurs pratiques et de garder la qualité scientifique des indicateurs. Dans ce chapitre, nous avons considéré un exemple particulier des indicateurs de la méthode INDIGO qui est l'indicateur "pesticide". Nous avons considéré une matière active et nous avons calculé l'indicateur I_{eso} tel qu'il est calculé à l'INRA de Colmar avec les valeurs choisies par les experts. Nous avons également pris en compte l'incertitude due aux variables multi-sources de cet indicateur. Notons que les indicateurs I_{esu}, I_{air} et I_{phy} se calculent de la même manière que I_{eso}. Les informations imprécises sont représentées par des intervalles. En utilisant le mode de fusion par SMC, nous avons construit le treillis de la fusion par SMC pour I_{eso}. Le treillis montre une divergence entre les sources d'informations utilisées pour le calcul des indicateurs. Il permet les décideurs d'évaluer les pratiques agricoles en considérant des groupements partiels des sources et leurs valeurs. Le regroupement global de toutes les sources ne permet pas de prendre une décision à partir de la valeur associée à $\top$ dans le treillis. Dans ce travail, nous n'avons considéré que les variables provenant de plusieurs sources d'informations. Les imperfections des variables représentant le milieu et les conditions d'application du pesticide, ainsi que les conclusions des règles de décision sont considérées fixes (précises) dans tout ce travail. L'expertise dans le choix des fonctions d'appartenance et des seuils des variables intervenant dans le calcul de l'indicateur n'est pas modélisée. D'autre part, les conclusions des règles de décision des variables de milieu ainsi que les arbres de décision utilisés dans le calcul de l'indicateur ne sont pas construits à partir des jeux de données et l'imperfection due à ce choix n'est pas représentée. Notons qu'aucune information n'est disponible à propos des variables de milieu (potentiel de lessivage, potentiel de ruissellement et potentiel de dérive), ce qui est due à la difficulté de mesurer ces variables et leur donner des marges.

Si l'on dispose de plus d'informations sur les sources, par exemple, un ordre de préférence, ces informations peuvent être introduites au moment de la construction du treillis. Dans l'exemple de la matière active *sulcotrione*, pour l'indicateur I_{eso}, nous considérons deux caractéristiques. Mais si l'on considère l'indicateur I_{phy}, il faut calculer les deux autres modules correspondants aux eaux de surface (représentées par I_{esu} et à l'air I_{air}). La table de données du pesticide sulcotrione n'est pas volumineux. Dans une telle situation, nous appliquons l'approche de la fusion contrainte par un seuil de similarité θ (voir Annexe D), qui permet d'éliminer des groupements dont les résultats de la fusion sont des intervalles de grande taille et qui ne permettent pas aux décideurs d'évaluer les pratiques. Notons que le calcul de l'indicateur pesticide comporte plusieurs niveaux d'agrégation, ce qui augmente parfois l'incertitude en particulier quand l'incertitude sur une variable se propage plusieurs fois.

6

Conclusion et perspectives

Sommaire

6.1 Conclusion

Pour calculer la valeur d'un indicateur, les experts doivent choisir une valeur parmi les valeurs proposées par les sources. Dans cette thèse, nous avons proposé une méthode faisant appel à la théorie des possibilités et à l'analyse formelle de concepts tout comme à par la fusion d'informations et l'analyse d'intervalles pour calculer un indicateur environnemental à partir de données imparfaites.

Dans la première partie de la thèse, nous avons présenté la théorie des possibilités pour la modélisation des informations imparfaites, en particulier les informations fournies par plusieurs sources d'informations différentes. Nous avons utilisé les intervalles (distributions de possibilités dans $\{0, 1\}$), et les distributions de possibilités triangulaires et trapézoïdales (distributions de possibilités dans $[0, 1]$), pour représenter les informations fournies par les sources. Cette représentation permet de prendre en compte l'imprécision des variables intervenant dans le calcul de l'indicateur. De plus, elle est adaptée au mode de calcul de l'indicateur en utilisant une moyenne pondérée.

Après une introduction de l'analyse formelle de concepts et de son extension aux structures de patrons, nous avons présenté des liens existant entre la théorie des possibilités et l'analyse formelle de concepts. Plusieurs connexions de Galois peuvent être définies faisant appel aux mesures de la théorie des possibilités. De plus, la connexion de Galois utilisée dans l'analyse formelle de concepts est fondée sur la mesure de "possibilité garantie" de la théorie des possibilités. Notre contribution porte dans cette partie à étendre ces opérateurs de dérivation à des contextes hétérogènes en utilisant les opérateurs de dérivation de la connexion de Galois introduite dans les structures de patrons et faisant appel aux mesures de la théorie des possibilités. Ces opérations permettent de décomposer un contexte, qu'il soit binaire ou hétérogène, en plusieurs sous-contextes indépendants dont les objets – respectivement les attributs (et leurs descriptions dans les contextes hétérogènes) – sont des ensembles disjoints.

La deuxième contribution de cette thèse porte sur l'utilisation de l'AFC pour combiner les informations provenant de plusieurs sources différentes. Nous avons montré comment un opéra-

teur de fusion peut être considéré comme un infimum dans un inf.-demi-treillis, puis comment construire le treillis permettant d'organiser les sources d'informations et de calculer les résultats de la fusion.

Cette partie comporte deux méthodes s'appuyant chacune sur un opérateur de fusion. Les opérateurs de fusion utilisés sont : l'opérateur conjonctif, l'opérateur disjonctif et l'opérateur de compromis représenté par l'opérateur de fusion fondée sur la recherche des sous-ensembles maximaux cohérents (SMC). La première méthode concerne la fusion des intervalles par treillis. La deuxième méthode est une extension de la première et porte sur la fusion des distributions par treillis.

Concernant la fusion des intervalles, nous avons construit les treillis de concepts des résultats conjonctifs et disjonctifs, puisque ces opérateurs vérifient les propriétés d'un opérateur infimum dans un treillis. Nous avons également proposé une méthode permettant d'appliquer malgré tout des opérateurs de fusion qui ne vérifient pas les propriétés d'un infimum (essentiellement l'associativité), puis de construire le treillis de concepts de la fusion par SMC.

Les treillis de concepts obtenus proposent des classifications des sources d'informations et permettent de calculer les résultats de la fusion des intervalles fournis par les sources. En particulier, chaque concept représenté par une paire (extension, intension) où l'extension correspond au sous-ensemble maximal de sources ayant le résultat de la fusion donnée dans l'intension du concept. Quand le résultat global de la fusion, c-à-d le résultat obtenu en combinant toutes les informations réunies, n'est pas utile pour la décision, le treillis offre plus de choix à l'utilisateur pour le résultat de fusion. Le treillis offre une vue structurée du résultats global et des résultats partiels de la fusion qui sont obtenus à partir des sous-ensembles de sources. Ce choix peut être fait sur la base du label représenté dans l'extension de chaque concept ou bien avec de la méta-information sur les sources (comme la fiabilité, ordre de préférence, etc.).

Nous avons également proposé une méthode de fusion contrainte par un seuil de similarité pour combiner les informations disjonctivement. Cette méthode permet d'éliminer des concepts inutiles dans le treillis quand le nombre de sources augmente. Elle permet de plus de garder les concepts ayant des résultats de fusion suffisamment précis pour respecter le seuil choisi par l'expert.

Concernant la méthode de fusion des distributions : les informations sont représentées par des distributions triangulaires et trapézoïdales. La fusion par treillis est plus riche que dans le cas des intervalles. Outre les résultats de la fusion hierarchisés dans le treillis, le résultat global et l'origine de l'information conservée par les extensions des concepts, des nivaux de confiance sont fournis avec le résultat de la fusion dans les intensions des concepts. Cela permet d'évaluer la cohérence entre les sources d'informations des extensions.

Nous avons exploré également la fusion pour plusieurs variables simultanément. En effet, dans plusieurs applications, les sources fournissent des informations pour plusieurs variables. Nous avons alors proposé une méthode consistant à choisir un opérateur de fusion pour chaque variable à condition que les variables soient indépendantes. Cette méthode permet de trouver les résultats de fusion pour plusieurs variables simultanément.

La dernière partie de ce travail de thèse a porté sur l'évaluation de la méthode fusion par treillis en utilisant l'indicateur phytosanitaire qui mesure la qualité des eaux souterraines pour un pesticide donné. Tout d'abord, nous avons utilisé la théorie des possibilités en particulier les intervalles pour représenter les informations fournies par les sources agronomiques. Les intervalles permettent de prendre en compte les imprécisions des valeurs délivrées par les sources. Ensuite, nous avons choisi l'approche fondée sur les SMC pour construire le treillis de la fusion par SMC des caractéristiques du pesticide. Le treillis fournit plusieurs concepts permettant de calculer l'indicateur. Un algorithme faisant appel aux opérations d'analyse d'intervalles est choisi pour

le calcul de l'indicateur. Le résultat global de la fusion est imprécis et ne permet pas d'évaluer les pratiques agricoles utilisées au champ. Ceci est dû à la divergence entre les sources. Par conséquent, il n'est pas intéressant de prendre en compte globalement toutes les sources mais il faut les étudier séparément ou par blocs. En conséquence, l'approche fondée sur le treillis par SMC permet d'identifier des sources cohérentes ayant un résultat de fusion autorisant de calculer un indicateur évaluant les pratiques agricoles. De plus, elle permet aux experts agronomes de justifier leurs diagnostics.

6.2 Perspectives

Les perspectives de ce travail sont nombreuses. Les liens entre la théorie des possibilités et l'analyse formelle de concepts – en particulier les structures de patrons – sont à explorer. Les relations entre les différents opérateurs de dérivation introduits restent à étudier. La fusion d'informations dans le treillis peut apporter des solutions à différents problèmes rencontrés dans toute application dans diverses disciplines faisant intervenir la fusion de plusieurs sources. En particulier, le treillis des résultats de la fusion forme une base de choix de plusieurs sous-ensembles maximaux et leurs résultats de fusion. Il serait donc intéressant d'étudier les sous-ensembles obtenus et de proposer une méthode de choix entre les concepts. Ainsi une poursuite de ce travail serait de considérer de la méta-information concernant les sources, comme la fiabilité, et de s'intéresser à la façon de l'introduire et la gérer dans le treillis. Un ordre sur les degrés des fiabilités des sources permettrait de choisir et de comparer deux concepts c-à-d deux groupes de sous-ensembles de sources avec leurs résultats de fusion. Cela reviendrait ainsi à comparer des sous-ensembles de degrés de fiabilité. Il faudrait alors comparer la méthode proposée dans ce travail avec d'autres méthodes de fusion et examiner son extension pour d'autres types d'informations. L'idée de considérer les résultats partiels de la fusion et d'observer les sources distinctement est étudiée dans le cadre de la logique possibiliste dans [70]. Ce travail a pour but de combiner les informations dans un cadre possibiliste en intégrant des poids représentant la fiabilité des sources.
Néanmoins, la méthode de la fusion par treillis proposée pour plusieurs variables ne peut être utilisée que si les variables sont indépendantes. Il serait donc intéressant de prendre en compte la dépendance entre les variables.

Une autre ouverture de ce travail peut être l'étude des liens qui existent entre l'argumentation et l'analyse formelle de concepts. En effet, dans ce travail, nous avons utilisé la fusion au sein du treillis dans le cadre de l'analyse formelle de concepts. Nous avons obtenu un treillis composé des paires (sources, résultats) où le résultat représente la combinaison des informations des sources. Il est possible de trouver des liens entre les systèmes d'argumentation et l'analyse formelle de concepts qui permettent d'avoir un treillis d'arguments ainsi que d'évaluer la force d'un argument.

Cependant, la structure de treillis exige l'utilisation d'opérateurs qui vérifient un certain nombre de propriété comme l'idempotence, l'associativité et la commutativité. Dans ce travail, nous avons abordé le problème des opérateurs ne vérifiant pas les propriétés d'un infimum, essentiellement l'associativité. En effet l'associativité est importante en fusion lorsque les informations ne sont pas combinées en même temps. Il serait donc intéressant d'étudier les cas où l'opération de fusion ne vérifie pas toutes ces propriétés.

Néanmoins, l'application de ces approches peut poser quelques problèmes calculatoires. Il est nécessaire de disposer d'un appui algorithmique performant et d'implémentations efficaces pour mettre en œuvre les approches proposées (construction de treillis, calcul des résultats de la fusion avec SMC, etc.).

Enfin, concernant les indicateurs environnementaux, il conviendrait d'estimer l'incertitude lorsque les informations fournies par les sources sont modélisées par des distributions. Dans ce cadre, la méthode proposée permet d'obtenir des sous-ensembles maximaux des sources avec leurs résultats de fusion. Ces résultats sont des ensembles d'intervalles emboîtés, où pour toute α-coupe nous pouvons appliquer la méthode proposée ici (algorithme de Fortin) pour le calcul d'indicateur pour plusieurs niveaux afin de calculer le résultat. Notons que le résultat obtenu n'est pas une distribution.
Dans cette application sur les indicateurs, nous n'avons considéré que l'imperfection liée aux variables multi-sources. L'indicateur est calculé en utilisant des mesures de terrain et un nombre de variables estimées par les experts en utilisant les arbres de décision. Ces mesures de terrain peuvent être aussi imparfaites de même que les conclusions des règles de décision peuvent être imprécises (les conclusions des règles de décision sont considérées comme fixes dans tout notre travail). Les perspectives pour cette partie consistent à chercher des solutions plus générales permettant de représenter l'imperfection due à l'expertise lors du choix des conclusions (c-à-d la conclusion d'une règle de décision serait alors un intervalle et non plus une valeur précise). Ensuite, d'envisager la propagation de ces imperfections à travers le mode de calcul de l'indicateur.

A

Théorie des possibilités

La théorie des possibilités a été proposée par L.Zadeh en 1987 [189] en s'appuyant sur la théorie des sous-ensembles flous qu'il a inventé en 1965 [188]. Cette dernière permet de représenter des classes d'objets dont les critères d'appartenance sont graduels [187]. La théorie des possibilités est ensuite développée par Dubois et Prade [74] pour traiter dans un seul cadre l'incertitude et l'imprécision inhérentes à certaines données.

A.1 Mesure de possibilité

La mesure de possibilité, Π, est une fonction définie sur l'ensemble des parties $P(U)$ de U, à valeurs dans $[0,1]$ telle que :

$$\begin{array}{rcl} \Pi(\emptyset) & = & 0 \\ \Pi(U) & = & 1 \\ \forall i,\ \forall A_i \in P(U)\ ,\ \Pi(\bigcup_{i=1,\ldots,n} A_i) & = & \max_{i=1,\ldots,n} \Pi(A_i) \end{array} \tag{A.1}$$

La mesure de possibilité $\Pi(A)$ quantifie dans quelle mesure l'événement $A \subseteq U$ est possible. Elle est calculée ainsi à partir de la distribution de possibilités, on a :

$$\forall A \in P(U) \quad , \quad \Pi(A) = \sup_{u \in A} \pi(u)$$

De plus, si $\Pi(A) = 1$, la distribution est normalisée. Cette mesure satisfait la relation : $\max\big(\Pi(A), \Pi(\bar{A})\big) = 1$. Cela signifie que l'un des événements au moins est possible. La possibilité de l'un ne veut pas dire l'impossibilité de l'autre. Plus un événement est défini d'une manière plus précise qu'un autre événement, plus la possibilité qu'il se réalise est forte c-à-d $\forall A_1 \subseteq A_2 \Leftrightarrow \Pi(A_1) \leq \Pi(A_2)$.

A.2 Mesure de nécessité

La mesure de nécessité N est une fonction définie sur l'ensemble des parties $P(U)$ de U, à valeurs dans $[0,1]$ telle que :

$$\begin{array}{rcl} N(\emptyset) & = & 0 \\ N(U) & = & 1 \\ \forall i,\ \forall A_i \in P(U)\ ,\ N(\bigcap A_i) & = & \min N(A_i) \end{array} \tag{A.2}$$

La mesure de nécessité $N(A)$ quantifie dans quelle mesure l'événement $A \subseteq U$ est nécessaire. Elle est calculée ainsi à partir de la distribution de possibilité, on a :

$$\forall A \in P(U) \quad , \quad N(A) = 1 - \sup_{u \notin A} \pi(u)$$

De plus, si $N(A) = 1$, l'événement A est certainement vrai. Si $N(A) = 0$ alors l'événement A n'est pas certain du tout.

La mesure de nécessité est la mesure duale de la mesure de possibilité. La normalisation de π garantit la relation $\Pi(A) \geq N(A)$ pour tout événement A (un événement est possible même s'il n'est pas certain). La mesure de nécessité est liée à la mesure de possibilité par la relation $\Pi(A) = 1 - N(\overline{A})$. De plus on a :

- $\Pi(A) = 0$ et $N(A) = 0$ alors A est impossible.
- $\Pi(A) = 1$ et $N(A) = 0$ alors A est possible mais il n'est pas certain.
- $\Pi(A) = 1$ et $N(A) = 1$ alors A est certain

On en déduit les relations suivantes :

$$N(A) > 0 \quad \Leftrightarrow \quad \Pi(A) = 1 \tag{A.3}$$

Ainsi, si $\Pi(A) = 0$ et $N(A) = 1$ alors A est à la fois impossible et certain, ce qui est impossible. L'incertitude d'un événement A, au contraire de la théorie des probabilités, est caractérisée par les deux mesures : sa mesure de possibilité $\Pi(A)$ et sa mesure de nécessité $N(A)$.

A.3 Mesure de possibilité garantie

La mesure de garantie (appelée aussi mesure de suffisance) est une fonction définie sur l'ensemble U à valeurs dans $[0, 1]$, vérifiant les axiomes suivants :

$$\begin{array}{lcl} \Delta(\emptyset) & = & 1 \\ \forall i,\ \forall A_i \in P(U)\ ,\ \Delta(\bigcup A_i) & = & \min \Delta(A_i) \end{array} \tag{A.4}$$

Lorsque la mesure de possibilité est égale à 1, cela signifie qu'il existe au moins un événement élémentaire possible. L'utilisation de la mesure de garantie assure plus de certitude sur la possibilité [77]. Une mesure de garantie signifie que l'événement est certainement possible (on est sûr de la réalisation de l'événement, c-à-d l'événement a été observé). La mesure de suffisance est reliée à la distribution de possibilité par la relation

$$\forall A \subseteq U,\ \Delta(A) \quad = \quad \inf_{u \in A} \pi(u) \tag{A.5}$$

La mesure de suffisance est plus stricte que la mesure de possibilité, $\Delta(A) \leq \Pi(A)$, puisque Π n'évalue que dans quelle mesure un événement est possible (ou l'existence d'une valeur dans A qui est compatible avec l'information disponible), mais Δ évalue dans quelle mesure toutes les valeurs de A sont compatibles avec l'information disponible.

A.4 Mesure de certitude potentielle

La mesure de certitude potentielle évalue la possibilité d'existence d'un élément n'appartenant pas à A (dans $\overline{A}$) qui serait possible. La mesure de certitude ∇ est une fonction de U à valeurs dans $[0, 1]$. Elle est liée avec la distribution de possibilité par la relation :

$$\forall A \in P(U) \quad , \quad \nabla(A) = \sup_{u \notin A} (1 - \pi(u))$$

Cette mesure est la mesure duale de la mesure de suffisance. Ainsi, s'il existe un élément dans U qui n'est pas possible ($\exists x \in U; \pi(x) = 0$), on a $\min(\Delta(A), \Delta(\overline{A})) = 0$ et $\nabla(A) = 1$ puisque $\Delta(A) > 0$. Il est évident que $N \leq \nabla$.

Néanmoins, il n'existe pas de fortes relations entre les quatre mesures de la théorie des possibilités, mais on a :

$$\max(\Pi(A), 1 - N(A)) = \sup_{u \in U} \pi(u) \tag{A.6}$$

$$\min(\Delta(A), 1 - \nabla(A)) = \inf_{u \in U} \pi(u) \tag{A.7}$$

Si la distribution π est normalisée, l'équation (A.6) vaut 1 et si la distribution $1-\pi$ est normalisée alors l'équation (A.7) vaut 0. En utilisant la dualité entre les mesures, on a :

$$\max(N(A), \Delta(A)) \leq \min(\Pi(A), \nabla(A)) \tag{A.8}$$

B

Rappels sur l'analyse d'intervalles

B.1 Arithmétique par intervalles

L'arithmétique par intervalles, comme son nom le suggère, fait intervenir des intervalles avec les opérations arithmétiques. L'idée est d'une part de garantir les résultats en calculant un intervalle dans lequel se trouve le résultat effectif. D'autre part, on cherche à fournir un encadrement satisfaisant de la solution et à avoir un résultat suffisamment précis. L'arithmétique par intervalles est présente très tôt dans la littérature [185, 85]. Brian Hayes propose un historique de l'arithmétique par intervalles [110]. Quelques années plus tard, les principes fondamentaux du calcul d'intervalles ont été étudiée par les trois mathématiciens Mieeczyslaw Warmus en Pologne, Teruo Sunaga au Japon et Ramon E. Moore aux États-Unis. La version de Moore introduite dans [144, 145] propose une alternative aux arithmétiques existantes. En effet, si seule l'arithmétique exacte répond à un besoin de fiabilité, elle est lente et ne permet pas d'évaluer les fonctions mathématiques $(\sin, \exp, \ldots)$ même si elle permet de les manipuler. Le principal point faible de l'arithmétique par intervalles est cet encadrement que l'on obtient de la solution. En effet on peut obtenir un intervalle beaucoup trop large qui entraîne une grande imprécision sur le résultat. Les calculs doivent donc être menés de telle sorte que l'encadrement obtenu soit de largeur raisonnable.

Un des principaux buts de l'analyse d'intervalles est de calculer l'ensemble des valeurs d'une fonction quand ses arguments varient dans des intervalles donnés [144]. Étant donné une fonction à n arguments $f(x_1, x_2, \ldots, x_n)$ de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^m$, et n intervalles $[a_i, b_i], i \in 1, \ldots, n$, l'ensemble des valeurs $y = f(x_1, x_2, \ldots, x_n)$ lorsque chaque x_i varie dans l'intervalle $[a_i, b_i]$, est donné par l'équation suivante :

$$\{y/y = f(x), x \in \times_i [a_i, b_i]\} \tag{B.1}$$

Si f est une fonction continue de $\mathbb{R}^n$ dans $\mathbb{R}$, et que l'on cherche à trouver sa variation sur un produit cartésien d'intervalles, généralement sur un convexe fermé de $\mathbb{R}$, alors le théorème des valeurs intermédiaires permet de savoir que le résultat de de l'équation B.1 est un intervalle de $\mathbb{R}$.

B.2 Fonctions d'intervalles

Définition 11 (Fonction d'intervalles) *Une fonction d'intervalles est une fonction* $g : \mathbb{IR}^n \to \mathbb{IR}^m$ *qui associe à un n-uplet d'intervalles un m-uplet d'intervalles.*

L'extension d'une fonction f doit vérifier la définition suivante :

Définition 12 (extension d'une fonction aux intervalles [149]) *Soient $f : \mathbb{R}^n \to \mathbb{R}^m$ une fonction continue et $g : \mathbb{IR}^n \to \mathbb{IR}^m$. g est une extension aux intervalles de f si et seulement si les deux conditions suivantes sont vérifiées :*
(i) $\forall x \in \mathbb{R}^n, g(x) = f(x)$
(ii) $\forall \mathcal{I} \in \mathbb{IR}^n, \{y/y = f(x), x \in \mathcal{I}\} \subseteq g(\mathcal{I})$

Plusieurs fonctions d'intervalles sont donc possibles pour étendre une fonction réelle aux intervalles. Parmi les extensions possibles d'une fonction f, il existe ce qu'on appelle l'extension *optimale*. La valeur de l'extension optimale d'une fonction f sur un produit d'intervalles $\mathcal{I}$ est le plus petit ensemble de $\mathbb{IR}^m$ contenant $\{y/y = f(x), x \in \mathcal{I}\}$.

Définition 13 (Extension optimale) *Soient $f : \mathbb{R}^n \to \mathbb{R}^m$ une fonction continue et $g : \mathbb{IR}^n \to \mathbb{IR}^m$ une extension de f. L'extension g est optimale en $\mathcal{I} \in \mathbb{IR}^n$ si et seulement si $\forall Z \in \mathbb{IR}^m$*

$$\{y/y = f(x), x \in \mathcal{I}\} \subseteq Z \subseteq g(\mathcal{I}) \Rightarrow Z = g(\mathcal{I})$$

g est dite optimale si et seulement si elle est optimale pour tout $\mathcal{I} \in \mathbb{IR}^n$.

La définition B.2 induit l'existence et l'unicité de l'extension optimale d'une fonction f. Soit $f^{\mathbb{IR}}$ l'unique extension optimale de f. Notons que l'extension optimale d'une fonction f sur un hypercube $\mathcal{I}$ et l'ensemble des valeurs de f atteintes par f sur $\mathcal{I}$ sont différents et on a : $f^{\mathbb{IR}} \neq \{y/y = f(x), x \in \mathcal{I}\}$, ce problème est appelé effet d'enveloppe [151].

Lorsqu'on manipule des fonctions continues de $\mathbb{R}^n$ dans $\mathbb{R}$, l'effet d'enveloppe disparaît. En effet, l'image $\{y/y = f(x), x \in \mathcal{I}\}$ d'un hypercube $\mathcal{I}$ par une fonction f est un intervalle, et la projection d'un intervalle sur l'axe des réels est identique à l'intervalle de départ. Ainsi, l'extension de f est obtenue à partir des extensions de ses composantes :

$$f^{\mathbb{IR}}(\mathcal{I}) \quad = \quad \{y/y = f(x), x \in \mathcal{I}\} \quad = \quad [\min_{x \in \mathcal{I}} f(x), \max_{x \in \mathcal{I}} f(x)] \tag{B.2}$$

B.2.1 Extension optimale des fonctions élémentaires

Les fonctions élémentaires sont étendues de la même façon aux intervalles. Dans beaucoup de cas, on obtient des formules très simples pour calculer l'extension optimale d'une opération en fonction des bornes de ses arguments. Moore [146] définit l'extension optimale de l'ensemble Φ des fonctions élémentaires suivantes :

$$\phi = \{\exp(x), \ln(x), \sin(x), \cos(x), \tan(x), \arccos(x), \arcsin(x), \arctan(x), abs(x), x^n, \sqrt[n]{x} \forall x \in \mathbb{N}^*\}$$

On obtient par exemple les formules suivantes pour les fonctions exponentielle et logarithme appliquées à l'intervalle $[a, b]$:

$$\exp([a, b]) \quad = \quad [\exp(a), \exp(b)] \tag{B.3}$$
$$\ln([a, b]) \quad = \quad [\ln(a), \ln(b)] \tag{B.4}$$

B.2.2 Définitions des opérations arithmétiques

Les opérations arithmétique standards $+, -, \times$ et $\div$ sont étendues aux intervalles en utilisant les extensions optimales.

Soient $A = [a^-, a^+]$ et $B = [b^-, b^+]$ deux intervalles réels, Les opérations d'addition, de soustraction, de multiplication et de division sont définies comme suit :

$$\begin{aligned} A \oplus B &= [a^-, a^+] \oplus [b^-, b^+] \\ &= [a^- + b^-, a^+ + b^+] \end{aligned} \tag{B.5}$$

$$\begin{aligned} A \ominus B &= [a^-, a^+] - [b^-, b^+] \\ &= [a^- - b^+, a^+ - b^-] \end{aligned} \tag{B.6}$$

$$A \otimes B = [a^-, a^+] \otimes [b^-, b^+] \tag{B.7}$$

$$\begin{aligned} &= [\min\{a^-b^-, a^-b^+, a^+b^-, a^+b^+\}, \\ &\quad \max\{a^-b^-, a^-b^+, a^+b^-, a^+b^+\}] \end{aligned} \tag{B.8}$$

$$\begin{aligned} A \oslash B &= [a^-, a^+] \oslash [b^-, b^+] \\ &= [\min\{\frac{a^-}{b^-}, \frac{a^-}{b^+}, \frac{a^+}{b^-}, \frac{a^+}{b^+}\}, \max\{\frac{a^-}{b^-}, \frac{a^-}{b^+}, \frac{a^+}{b^-}, \frac{a^+}{b^+}\}] \end{aligned} \tag{B.9}$$

La dernière équation est définie pour $B \in \mathbb{IR} \backslash [0, 0]$. De nombreuses propriétés découlent de ces opérations [149]. Parmi ces propriétés on peut citer les suivantes, $\forall A$, B et C des intervalles de $\mathbb{IR}$:

– Associativité :

$$(A \oplus B) \oplus C = A \oplus (B \oplus C) \tag{B.10}$$

$$(A \otimes B) \otimes C = A \otimes (B \otimes C) \tag{B.11}$$

– Commutativité :

$$A \oplus B = B \oplus A \tag{B.12}$$

$$A \otimes B = B \otimes A \tag{B.13}$$

– Présence d'un élément neutre

$$A \oplus 0 = A \tag{B.14}$$

$$A \otimes 1 = A \tag{B.15}$$

En revanche, certaines propriétés algébriques fondamentales des réels ne sont plus valides, en particulier la multiplication n'est pas distributive sur l'addition et les intervalles ne possèdent pas d'inverses, ni pour l'addition, ni pour la multiplication :

$$A \otimes (B \oplus C) \subseteq (A \otimes B) \oplus (A \otimes C) \tag{B.16}$$

$$A \ominus A \neq 0 \tag{B.17}$$

$$A \oslash A \neq 1 \tag{B.18}$$

L'intervalle $[0,0]$ est un élément neutre par rapport à l'addition, de même que l'intervalle $[1,1]$ par rapport à la multiplication. Notons que dans la théorie des intervalles, on ne peut pas obtenir un inverse d'un élément pour l'addition et la multiplication. En effet, si A est un intervalle de $\mathbb{R}$ qui n'est pas réduit à un singleton, il n'existe aucun intervalle $B \in \mathbb{IR}$ tel que $A+B=[0,0]$. Ainsi la structure algébrique $(\mathbb{IR}, \oplus)$ n'est pas un groupe et $(\mathbb{IR}, \oplus, \otimes)$ n'est pas un anneau. Ces deux structures étant les structures de bases des nombres réels, et la plupart des outils mathématiques sont inadaptés pour le calcul d'intervalles.

Puisqu'un intervalle est un ensemble d'éléments, la théorie des intervalles manipule les opérations d'union, d'intersection et de complémentation. Un des problèmes de ces opérations est que leur résultat n'est pas nécessairement un intervalle, c-à-d le résultat n'est pas convexe. Ansi l'ensemble vide ($\emptyset$) n'est pas un intervalle. Dans la suite nous verrons que l'intersection de deux intervalles disjoints n'est pas un intervalle, de même que l'union et le complémentaire, puisque les résultats obtenus ne sont pas forcément convexes.

$$A \cap B = \begin{cases} [\max(a^-, b^-), \min(a^+, b^+)] & \text{si } b^- < a^+ \\ \emptyset & \text{sinon} \end{cases} \tag{B.19}$$

$$A \cup B = \begin{cases} [\min(a^-, b^-), \max(a^+, b^+)] & \text{si } b^- < a^+ \\ [a^-, a^+] \cup [b^-, b^+] & \text{sinon} \end{cases} \tag{B.20}$$

Les opérations min et max peuvent être étendues aux intervalles [72], et on a :

$$\min([a^-, a^+], [b^-, b^+]) = [\min(a^-, b^-), \min(a^+, b^+)] \tag{B.21}$$

$$\max([a^-, a^+], [b^-, b^+]) = [\max(a^-, b^-), \max(a^+, b^+)] \tag{B.22}$$

B.2.3 Extension naturelle

L'idée la plus naturelle pour aborder le calcul d'intervalles semble d'évaluer la valeur de la fonction f sur des intervalles en remplaçant chaque opération arithmétique et élémentaire par son extension optimale. Moore [146] a montré que l'on obtient alors bien une extension de f. On appelle cette extension l'extension naturelle.

Théorème 1 (Extension naturelle) *Soient $f : \mathbb{IR}^n \to \mathbb{IR}^m$ une fonction continue et $\mathbf{f}(x_1, \ldots, x_n)$ une expression de f mettant en jeu des opérations arithmétiques $(+, -, \times, \div)$ et des fonctions élémentaires de ϕ. L'expression $\mathbf{f}$ est étendue aux intervalles en remplaçant les opérations réelles par leurs extensions optimales.La fonction d'intervalles g définie par $g(x) = \mathbf{f}(x)$ est alors une extension de f, appelée extension naturelle de f.*

Cette extension est simple, mais plusieurs problèmes apparaissent : il existe notamment une infinité d'expressions pouvant conduire à des extensions différentes.Lorsque l'on utilise des intervalles pour modéliser l'incertitude, plusieurs apparitions d'une même variable dans une expression reviennent à comptabiliser plusieurs fois l'incertitude sur la variable en question, rendant le résultat d'un calcul bien plus imprécis qu'il ne pourrait l'être.

B.2.4 Autres extensions

L'extension optimale et l'extension naturelle ne sont pas les seules extensions aux intervalles possibles d'une fonction f. En effet, il y a une infinité de fonctions d'intervalles possibles vérifiant les deux conditions de la définition B.2. On peut citer par exemple l'extension de la valeur

moyenne (voir par exemple [146]), faisant intervenir l'extension naturelle de la dérivée de f ou l'extension de Taylor fondée sur un développement limité de la fonction f (voir [2]). Le but de ces extensions est de se rapprocher autant que possible des résultats de l'extension optimale, en gardant une complexité de calcul raisonnable, quand l'extension optimale est trop complexe pour être calculée explicitement.

B.3 Analyse d'intervalles flous

L'arithmétique floue a été développée pour étendre les opérations d'addition, de multiplication, de soustraction et de division sur des nombres flous [71, 115] en accord avec le principe d'extension de Zadeh [188], très utilisé dans le calcul des nombres flous [187]. De nombreux auteurs ont essayé d'étendre d'une façon générale les opérations arithmétiques et de généraliser l'extension naturelle des fonctions aux intervalles flous. Ces travaux proposent un calcul fondé sur l'application des méthodes classiques d'analyse d'intervalles sur une partie des α-coupes des intervalles flous considérés. Nguyen [148] a montré sous diverses hypothèses que l'α-coupe d'un résultat flou peut être obtenue grâce aux α-coupes des arguments flous. Les travaux de Dong et al. [64], Yang et al.[184], Anile [3], Hanss [109] proposent des méthodes de calcul efficaces fondées sur l'idée d'appliquer les algorithmes d'analyse d'intervalles classiques sur une sélection d'α-coupes. Ces techniques de calcul présentent des inconvénients : l'algorithme doit être exécuté entièrement pour chaque α-coupe [109] et le résultat obtenu est une approximation du résultat, reconstitué à partir des α-coupes calculées. On ne peut en aucun cas obtenir une formule exacte. D'autres méthodes d'arithmétique d'intervalles sont fondées sur les nombres flous paramétrés $L-R$ [71]. Cette méthode permet d'obtenir des formules exactes pour des opérations comme l'addition, la soustraction, la multiplication et la division des nombres flous positifs [86]. Plus récemment, Fortin et al. [90] ont proposé une méthode (la méthode des profils) pour calculer les valeurs exactes pour tous les degrés de possibilités. Cette approche est fondée sur la notion de nombre graduel [68, 89, 91]. Un intervalle flou est représenté par une paire de bornes floues d'une manière similaire aux intervalles classiques représentés par une paire de deux réels. Les techniques de l'analyse d'intervalles appliquées aux intervalles réels sont alors appliquées aux intervalles flous.

Dans la littérature, un nombre flou $\tilde{M}$ désigne généralement un ensemble flou de réels, dont chaque α-coupe $[\tilde{M}]_\alpha = \{x|\mu_{\tilde{M}}(x) \geq \alpha\}$ est un intervalle. Ainsi, la fonction d'appartenance à l'ensemble flou $\tilde{M}$ a la forme suivante :

- le noyau est un intervalle $[m_1^-, m_1^+]$
- $\mu_{\tilde{M}}$ est non-décroissante sur $(-\infty, m_1^-]$
- $\mu_{\tilde{M}}$ est non-croissante sur $[m_1^+, +\infty)$

Cette structure généralise les intervalles et non les nombres. Le principe d'extension de Zadeh [188] permet de calculer l'image de deux nombres flous $\tilde{M}$ et $\tilde{N}$ par une fonction f :

$$\mu_{f(\tilde{M},\tilde{N})}(z) = \sup_{(x,y)|z=f(x,y)} \min(\mu_{\tilde{M}}(x), \mu_{\tilde{N}}(y)) \tag{B.23}$$

La fonction d'appartenance $\mu_{f(\tilde{M},\tilde{N})}(.)$ peut être construite en appliquant l'extension optimale aux α-coupes [148] :

$$[f(\tilde{M},\tilde{N})]_\alpha = f([\tilde{M}]_\alpha, [\tilde{N}]_\alpha) \tag{B.24}$$

$$= \{f(x,y)|x \in [\tilde{M}]_\alpha, y \in [\tilde{N}]_\alpha\} \tag{B.25}$$

Nous remarquons bien que le principe d'extension appliqué à des nombres flous, plus généralement à des intervalles flous, généralise le calcul d'intervalles. Fortin [89] a utilisé la notion de nombre graduel, en considérant qu'un intervalle flou peut être considéré comme une paire de nombres graduels. À partir de cette notion et de sa similarité avec les nombres réels, les opérations arithmétiques appliquées sur les nombres réels sont applicables sur les nombres graduels.

B.4 Moyenne floue pondérée

La moyenne floue pondérée fut introduite initialement par Baas et Kwakernaak [16] pour traiter des problèmes d'évaluation d'incertitude. Elle est utilisée comme une méthode de fusion de données dans les situations où les décisions et leurs poids peuvent être modélisés par des sous-ensembles flous, les décisions et/ou les poids peuvant être entachés d'incertitude.

La fonction de moyenne floue pondérée est utilisée dans plusieurs domaines d'application comme par exemple l'aide à la décision, l'analyse de risques, etc. Elle est définie par :

$$fwa(\omega_1, \omega_2, \ldots, \omega_n, x_1, x_2, \ldots, x_n) = \frac{\omega_1 x_1 + \omega_2 x_2 + \ldots + \omega_n x_n}{\omega_1 + \omega_2 + \ldots + \omega_n} \tag{B.26}$$

où $(x_1, x_2, \ldots, x_n)$ représente le vecteur de critères dont on veut calculer la moyenne et ω_i les poids positifs affectés aux critères x_i. Le domaine de $fwa(.)$ est $\mathbb{R}^{+n} \times \mathbb{R}^n$ ($\forall i \in \{1, \ldots, n\}$, $\omega_i \in \mathbb{R}^+$ et $x_i \in \mathbb{R}$). Il existe une hiérarchie qu'on peut associer à la fonction définie dans B.4 :

1. Si ω_i et x_i sont représentés par des nombres précis : dans ce cas la valeur de fwa est un nombre précis et c'est la moyenne arithmétique classique.
2. Si ω_i sont représentés par des nombres et x_i sont des critères représentés par des intervalles, c-à-d $x_i \in [a_i, b_i]$ où les bornes a_i et b_i sont précises : dans ce cas la valeur de fwa est un intervalle de nombres (une moyenne pondérée d'intervalles) et $fwa \in [fwa_L, fwa_U]$ où fwa_L et fwa_U sont calculées facilement en utilisant l'analyse d'intervalles classique puisque la fonction ne contient que des intervalles dans son numérateur.
3. Si x_i sont représentés par des nombres précis et ω_i sont représentés par des intervalles de nombres, c-à-d $\omega_i \in [c_i, d_i]$ où les bornes c_i et d_i sont précises, dans ce cas fwa est un cas particulier du problème de la moyenne floue pondérée où on utilise des algorithmes existants pour le calcul des bornes fwa_L et fwa_U de fwa, comme par exemple les algorithmes présentés dans [90, 114, 133, 65]. Ces derniers travaux calculent les valeurs exactes de fwa_L et fwa_U.
4. Si $x_i \in [a_i, b_i]$ et $\omega_i \in [c_i, d_i]$ où a_i, b_i, c_i et d_i sont précis : c'est une autre forme de la moyenne floue pondérée appelée aussi "centroïde généralisé des intervalles flous de type 2" (*generalized centroid of interval type-2 fuzzy sets*) [114, 133]. Comme dans le cas précédent, fwa est un intervalle de nombres pour lequel plusieurs algorithmes peuvent être utilisés pour calculer fwa_L et fwa_U.
5. Si x_i et ω_i sont représentés par des intervalles flous, fwa est un sous-ensemble flou. Ce cas peut être vu comme un cas général des cas précédents. Les bornes inférieure et supérieure de la fonction sont calculées dans la littérature par plusieurs auteurs (voir [129, 114, 133, 65, 43]).

Dans la suite, nous rappellons les algorithmes existants pour le calcul de variation de la fonction fwa en citant les points communs entre certains algorithmes. Leurs complexités respectives sont illustrées par un exemple.

B.4.1 Algorithme de calculs de la moyenne floue pondérée

Algorithme de Dong et Wong (*VFWA :Vertex Fuzzy Weighted Average*)

, Dong et Wong ont montré que les bornes de la fonction fwa sont atteintes à la suite de 2^{2n} permutations distinctes de $2n$ variables $(x_1, \ldots, x_n, \omega_1, \ldots, \omega_n)$ avec $x_i = a_i$ ou b_i et $\omega_i = c_i$ ou d_i, $i = 1, \ldots, n$. L'intervalle de variation de la fonction est $[\min_k\{fwa_k\}, \max_k\{fwa_k\}]$. La procédure de Dong et Wong se résume par les étapes suivantes :

1. Considérer m α-coupes où $\alpha_j \in [0, 1]$.
2. Pour chaque valeur α_j $(j = 1, \ldots, m)$, calculer les intervalles $[a_i, b_i]$ et $[c_i, d_i]$, $i = 1, \ldots, n$.
3. Construire 2^{2n} permutations du vecteur $(x_1, \ldots, x_n, \omega_1, \ldots, \omega_n)$ où $x_i = a_i$ ou b_i, et $\omega_i = c_i$ ou d_i, $i = 1, \ldots, n$
4. Calculer $fwa_k = f(x_{k1}, x_{k1}, \ldots, x_{k1}, \omega_{k1}, \ldots, \omega_{k1})$ où $(x_{k1}, x_{k1}, \ldots, x_{k1}, \omega_{k1}, \ldots, \omega_{k1})$ est la $k^{\text{ème}}$ permutation des 2^{2n}, $k = 1, 2, \ldots, 2n$. L'intervalle de variation de la fonction est $[\min_k fwa_k, \max_k fwa_k]$
5. Répéter les étapes (2)(3)(4) pour tout α_j, $j = 1, 2, \ldots, m$

VFWA se calcule à partir de 2^{2n} permutations. Cet algorithme exige 2^{2n} itérations pour chaque α-coupe, donc $m2^{2n}$ itérations, d'où la difficulté d'utilisation de cet algorithme dans les cas pratiques, même avec un nombre de critères peu élevé.

Algorithme de Liou et Wang (*IFWA :Improved Fuzzy Weighted Average*)

La procédure de l'algorithme de Liou et Wang [132] est fondée sur la définition de deux fonctions fwa_U et fwa_L, en remplaçant les critères x_i par a_i pour la borne inférieure et b_i pour la borne supérieure :

$$fwa_U(\omega_1, \ldots, \omega_n) = \frac{\sum_{i=1}^{n} \omega_i b_i}{\sum_{i=1}^{n} \omega_i} \tag{B.27}$$

$$fwa_L(\omega_1, \ldots, \omega_n) = \frac{\sum_{i=1}^{n} \omega_i a_i}{\sum_{i=1}^{n} \omega_i} \tag{B.28}$$

L'intervalle de variation de la fonction fwa est $[\min_k fwa_{L_k}, \max_k fwa_{U_k}]$ où k est la $k^{\text{ème}}$ permutation de 2^n, $k = 1, \ldots, 2^n$. La borne inférieure de l'intervalle est $L = \min fwa_L$ et la borne supérieure est $U = \max fwa_U$.

1. Calculer $L = f_L(c_1, \ldots, c_n)$ et $U = f_U(c_1, \ldots, c_n)$. Considérer les ensembles des indices I et J tels que $I = \{i, |a_i < L, i = 1, \ldots, n\}$ et $J = \{j, |b_j > U, j = 1, \ldots, n\}$ et $R_L = R_U = \emptyset$.
2. Si $I = \emptyset$, alors la valeur minimale de fwa_L est L et arrêter l'étape 2. Sinon, ω_i prend la valeur d_i pour $i \in R_L$ et c_i pour $i \notin R_L$
   - Déterminer $l_i = fwa_L(\omega_1, \ldots, \omega_{i-1}, d_i, \omega_{i+1}, \ldots, \omega_n)$ pour $i \in I$
   - $L = l_m = \min_{i \in I} l_i$, ajouter m à R_L et retirer m de l'ensemble I ; considérer $T = \{i | a_i > L, i \in I\}$.
   - Retirer i de I où i appartient à T. Si $I \neq \emptyset$ alors répéter l'étape 1 sinon arrêter l'étape 1 et $\min fwa_L = L$
3. Si $J = \emptyset$ alors la valeur maximale de fwa_U est U et arrêter l'étape 3. Sinon, w_i prend la valeur d_i pour $i \in R_U$ et c_i pour $i \notin R_U$
   - Déterminer $u_j = fwa_U(\omega_1, \ldots, \omega_{j-1}, d_j, \omega_{j+1}, \ldots, \omega_n)$ pour $j \in J$

- $U = u_m = \min_{i \in I} u_i$, ajouter m à R_U et retirer m de l'ensemble J ; considérer $T = \{i | a_i > L, i \in I\}$.
- Retirer j de J où j appartient à T. Si $J \neq \emptyset$ alors répéter l'étape 1 sinon arrêter et $\max fwa_U = U$

4. L'intervalle correspondant pour une α-coupe est $[L, U]$.
5. Répéter (2)(3)(4) pour plusieurs α-coupes.

Liou et Wang transforment l'ordre exponentiel de l'algorithme vFWA de Dong et Wong en un ordre polynomial : $2+n(n+1)$ pour chaque α-coupe, ce qui donne $m(2+n(n+1))$ itérations au total. En effet, l'algorithme VFWA exige une multiplication, $2n(n-1)$ additions et une division. En revanche, l'algorithme IFWA de Liou et Wang n'exige qu'une multiplication, une soustraction, 2 additions et une division.

Algorithme de Guh et al. (*PFWA : Programming Fuzzy Weighted Average*)

Cet algorithme est fondé sur la formule introduite par Liou et Wang [132], qui forme également une base pour la plupart des algorithmes FWA. Guh et al. [105] ont proposé une méthode plus rapide pour atteindre la solution exacte. Une procédure d'élimination de paires max-min est proposée. Les intervalles pour chaque α-coupe α_j, $j = 1, \ldots, m$ peuvent se calculer comme suit :

1. Déterminer les plus grands coefficients a_1 tel que $a_1 \geq a_i$ et b_1 tel que $b_1 \geq b_i$ et les plus petits coefficients a_n tel que $a_n \leq a_i$ et b_n tel que $b_n \leq b_i$, pour tout $i = 1, \ldots, n$
2. Pour $\min\{fwa_L\}$, choisir c_1 qui correspond au poids de a_1 et d_n qui correspond au poids de a_n. Pour $\max\{fwa_U\}$, choisir d_1 qui correspond au poids de b_1 et c_n correspondant à b_n
3. Déterminer le coefficient a' (respectivement b') et son poids correspondant ω' (respectivement ω') pour le calcul de $\min\{fwa_L\}$ (respectivement $\max\{fwa_U\}$)

$$a' = \frac{a_1 c_1 + a_n d_n}{c_1 + d_n}$$

$$\omega' = c_1 + d_n \qquad c' = d' = \omega'$$

$$b' = \frac{b_1 d_1 + b_n c_n}{d_1 + c_n}$$

$$\omega' = d_1 + c_n \qquad c' = d' = \omega'$$

4. Éliminer a_1, a_n, c_1, d_n, les remplacer par a' et son poids ω'. Éliminer b_1, b_n, d_1, c_n et les remplacer par b' et son poids ω'.
5. Répéter l'opération $n-1$ fois, l'intervalle résultant $[a', b']$ est la solution pour chaque coupe α_j.
6. Répéter la procédure pour plusieurs coupes.

Algorithme de Lee et Park. (*EFWA : Efficient Fuzzy Weighted Average*)

Lee et Park [129] utilisent aussi les deux fonctions $\min\{fwa_L\}$ et $\max\{fwa_L\}$. La nouveauté de leur algorithme est de ranger par ordre croissant les coefficients a_i et b_i en deux séquences

de coefficients renommées $\{a_i\}$ et $\{b_i\}$. Pour une séquence $S = (e_1, \ldots, e_n)$, les deux seuils sont définis par :

$$\delta_{S_i} = \frac{(a_1 - a_i)e_1 + \ldots + (a_n - a_i)e_n}{e_1 + \ldots + e_n} \tag{B.29}$$

$$\xi_{S_i} = \frac{(b_1 - b_i)e_1 + \ldots + (b_n - b_i)e_n}{e_1 + \ldots + e_n} \tag{B.30}$$

L'algorithme est résumé comme suit :

1. Ranger par ordre croissant les coefficients a_i. Nommer $(a_1, a_2, \ldots, a_n)$ la séquence résultante.
2. Considérer le seuil $\delta = \frac{(premier+dernier)}{2}$ où $premier = 1$ et $dernier = n$. Prendre $e_i = d_i$ pour tout $i = 1, \ldots, \delta$ et $e_i = c_i$ pour $i = \delta+1, \ldots, n$. Pour les n-uplets $S = (e_1, e_2, \ldots, e_n)$, évaluer les valeurs δ_{S_δ} et $\delta_{S_{\delta+1}}$
3. Si $\delta_{S_\delta} > 0$ et $\delta_{S_{\delta+1}} \leq 0$, alors, la valeur de la borne inférieure est $fwa_L(e_1, e_2, \ldots, e_n)$ et aller à l'étape (4) de l'algorithme. Si la condition n'est pas satisfaite exécuter l'étape suivante :
   - $\delta_{S_\delta} > 0$ alors le premier terme de la séquence des coefficients vaut $\delta + 1$; sinon le dernier vaut δ et aller à l'étape (2) de l'algorithme.
4. Ordonner par ordre croissant les coefficients b_i. Nommer $(b_1, b_2, \ldots, b_n)$ la séquence résultante.
5. Considérer le seuil $\xi = \frac{(premier+dernier)}{2}$ où $premier = 1$ et $dernier = n$. Pour $i = 1, \ldots, \xi$, poser $e_i = c_i$ et pour $i = \xi + 1, \ldots, n$, poser $e_i = d_i$. Pour les n-uplets $S' = (e_1, e_2, \ldots, e_n)$, évaluer les valeurs $\xi_{S'_\xi}$ et $\xi_{S'_{\xi+1}}$.
6. Si $\xi_{S'_\xi} > 0$ et $\xi_{S'_{\xi 1}} + \leq 0$, alors la borne supérieure de la fonction fwa vaut $fwa_U(e_1, e_2, \ldots, e_n)$ et arrêter le processus, sinon passer à l'étape suivante.
   - $\xi_{S'_{\xi+1}} > 0$ alors le premier terme de la séquence des coefficients vaut $\xi + 1$; sinon le dernier vaut ξ et aller à l'étape (5) de l'algorithme.

Le nombre d'itérations de l'algorithme EFWA est $(2n \log n)$ pour chaque α-coupe, c-à-d $2mn \log n$ au total pour m α-coupes. En comparant avec IFWA qui est plus efficace que VFWA, l'algorithme EFWA comporte un nombre d'itérations plus significatif que l'algorithme IFWA.

Algorithme de Guh et al. (*LPFWA : Linear Programming Fuzzy Weighted Average*)

Guh et al. [186] exprime les formules des deux fonctions fwa_L et fwa_U sous forme de systèmes linéaires :

$$\min fwa_L = \frac{\omega_1 a_1 + \omega_2 a_2 + \ldots + \omega_n a_n}{\omega_1 + \omega_2 + \ldots + \omega_n} \tag{B.31}$$

$$\max fwa_U = \frac{\omega_1 b_1 + \omega_2 b_2 + \ldots + \omega_n b_n}{\omega_1 + \omega_2 + \ldots + \omega_n} \tag{B.32}$$

où $c_i \leq \omega_i \leq d_i$ pour tout $i = 1, \ldots, n$.On détaillera le calcul de la borne inférieure, le calcul de la borne supérieure étant similaire.

En effet, la formule B.32 peut être exprimée comme un système linéaire simple comme l'indique la formule B.33, où p et q sont deux vecteurs de dimension n, $x \in \mathbb{R}^n$, A désigne une matrice d'ordre $m \times n$ et b est un vecteur de dimension m, fwa est notée f.

$$\begin{aligned} & \min \frac{px}{qx} \\ \text{tel que} \quad & Ax \leq b \\ & x \geq 0 \end{aligned} \tag{B.33}$$

En supposant que $qx \neq 0$ et en considérant les changements de variables, on obtient :

$$z = \frac{1}{qx} \qquad zx = f$$

En multipliant par z, on obtient alors le système linéaire

$$\begin{aligned} & \min pf \\ \text{tel que} \quad & Af \leq bz \\ & qf = 1 \end{aligned} \tag{B.34}$$

$$\begin{aligned} f &\geq 0 \\ z &\geq 0 \end{aligned} \tag{B.35}$$

Les systèmes linéaires obtenus sont résolus à l'aide de l'outil Lindo [17]. En comparant les performances de l'algorithme LPFWA avec l'algorithme EFWA, nous ne trouvons pas de différence significative entre les deux algorithmes.

Algorithme de Kuo-Ping Chiao (*DFWA : Direct Fuzzy Weighted Average*)

Kuo-Ping Chiao [47] a étudié les seuils δ et ξ définis par Lee and Park [129] et trouvé quelques caractéristiques importantes pour lesquelles son algorithme donne les mêmes résultats que l'algorithme de Lee et Park avec une complexité de $O(n \log n)$. L'intervalle de la moyenne pondérée est alors déterminé par ces deux seuils qui représentent comme nous allons le voir ci-dessous le minimum et le maximum de la fonction. Dans son algorithme, Kuo-Ping Chiao utilise deux fonctions τ et σ où $\tau(i) = \delta_i(v_i)$ et $\sigma(i) = \xi_i(u_i)$. v_i (respectivement u_i) représente la séquence S (respectivement S') utilisée dans les travaux de Lee et Park [129]. La procédure correspondante à chaque α-coupe est comme suit :

1. Ordonner les coefficients a_i par ordre croissant et renommer les poids correspondants par $[c_i, d_i]$ en respectant a_i pour la borne inférieure (respectivement, ordonner b_i et renommer les poids en concordance avec b_i pour la borne supérieure).
2. Pour la borne inférieure, construire le vecteur $\tau(i)$ (respectivement $\sigma(i)$ pour la borne supérieure)
3. Pour la borne supérieure, chercher l'indice i tel que $\tau(i) > 0$ et $\tau(i+1) \leq 0$ (respectivement $\sigma(i) > 0$ et $\sigma(i+1) \leq 0$). Ces deux indices correspondent aux deux seuils δ et ξ proposés par Lee et al. [129].

   Par conséquent, $\min\{fwa_L\} = fwa_L(v_i)$ et $\max\{fwa_U\} = fwa_U(u_i)$.
4. L'intervalle calculé pour chaque α-coupe est $[\min\{fwa_L\}, \max\{fwa_U\}]$.
5. Répéter les étapes (2) jusqu'à (4) pour plusieurs valeurs de α.

17. http ://www.lindo.com/

Algorithme de Kao et Liu (*PPFWA : Parametrized Programming Fuzzy Weighted Average*)

Kao et Liu [113] expriment les systèmes linéaires proposés par Guh et al. [186] comme une fonction de paramètre α. Ils ont de plus essayé de trouver une solution générale pour ces systèmes paramétrés. L'idée de Kao et Liu [113] est de remplacer x_i, w_i, a_i, b_i, c_i et d_i par $\tilde{x}_i$, $\tilde{w}_i$, $(x_i)^L_\alpha$, $(x_i)^U_\alpha$, $(w_i)^L_\alpha$ et $(w_i)^U_\alpha$ respectivement. La méthode est décrite comme suit :

- Les bornes inférieure et supérieure de l'α-coupe sont représentées par les deux systèmes linéaires suivants :

$$\begin{cases} \min fwa = \sum_{i=1}^n w_i x_i / \sum_{i=1}^n w_i \\ t.q. \\ (W_i)^L_\alpha \leq w_i \leq (W_i)^U_\alpha, i = 1, \ldots, n \\ (X_i)^L_\alpha \leq x_i \leq (X_i)^U_\alpha, i = 1, \ldots, n \end{cases}$$

$$\begin{cases} \max fwa = \sum_{i=1}^n w_i x_i / \sum_{i=1}^n w_i \\ t.q. \\ (W_i)^L_\alpha \leq w_i \leq (W_i)^U_\alpha, i = 1, \ldots, n \\ (X_i)^L_\alpha \leq x_i \leq (X_i)^U_\alpha, i = 1, \ldots, n \end{cases}$$

Où $(.)^L_\alpha$ et $(.)^U_\alpha$ représentent respectivement les bornes inférieure et supérieure du nombre flou correspondant à la coupe de niveau α. D'une manière similaire à la méthode donnée par Guh et al. [186], ces systèmes linéaires sont transformés en deux systèmes paramétrés par α qui sont résolus à l'aide de Lindo pour une valeur donnée de α. Nous verrons un exemple de transformation et de calcul dans la section B.4.2.

Algorithme de Guu (*MFWA : Median Fuzzy Weighted Average*)

Guu [106] a proposé un algorithme utilisant un système linéaire dont la solution est présentée dans les travaux de Hansen et al. [108]. L'algorithme de Guu est fondé sur le calcul des équations linéaires dont la solution introduit la recherche de la médiane de la séquence des points extrêmes des intervalles. Cet algorithme est aussi similaire à celui proposé dans les travaux [186] et [113], mais il est plus facilement utilisable dans les situations où les outils mathématiques tels que Lindo ne sont pas disponibles.

En effet, pour une α-coupe donnée, les bornes inférieure et supérieure des intervalles de la fonction de moyenne pondérée fwa sont données par les valeurs $\min_{\omega_i \in [c_i, d_i]} fwa_L(\omega_1, \ldots, \omega_n)$ et $\max_{\omega_i \in [c_i, d_i]} fwa_U(\omega_1, \ldots, \omega_n)$. fwa_L atteint son minimum au point extrême de la région $[c_1, d_1] \times \ldots \times [c_n, d_n]$, et fwa_U atteint son maximum au point extrême de la même région. Plus précisément, les auteurs utilisent un changement des variables ω_i et montrent que les relations suivantes restent vraies :

- L'équation (B.32) devient :

$$\min_{\omega_i^* \in \{0,1\}} \frac{\sum_{i=1}^n a_i c_i + \sum_{i=1}^n a_i (d_i - c_i)\omega_i^*}{\sum_{i=1}^n c_i + \sum_{i=1}^n (d_i - c_i)\omega_i^*} \tag{B.36}$$

- l'équation (B.32) devient :

$$\min_{\omega_i^* \in \{0,1\}} \frac{\sum_{i=1}^n (-b_i) c_i + \sum_{i=1}^n (-b_i)(d_i - c_i)\omega_i^*}{\sum_{i=1}^n c_i + \sum_{i=1}^n (d_i - c_i)\omega_i^*} \tag{B.37}$$

L'équation (B.32) est résolue en utilisant l'algorithme introduit par Hansen [108] et résumé par les étapes suivantes :

- Considérer $\bar{a} = \sum_{i=1}^{n} a_i c_i$ et $\bar{b} = \sum_{i=1}^{n} c_i$. $J_0 = \{1, 2, \ldots, n\}$, $J = \{j \in J_0 | a_j < \frac{\bar{a}}{\bar{b}}\}$, et $\omega^* = 0 \; \forall j \in J_0 \backslash J$.
- Recherche de la médiane : déterminer la valeur de l'indice k tel que a_k est la médiane de la séquence $\{a_j | j \in J\}$. Considérer ensuite $J_1 = \{j \in J | a_j \leq a_k\}$, $J_2 = \{j \in J | a_j \geq a_k\}$ et $J_3 = \{j \in J | a_j > a_k\}$.
- Calculer $\bar{a'} = \bar{a} + \sum_{j \in J_1} a_j (d_j - c_j)$, $\bar{b'} = \bar{b} + \sum_{j \in J_1} (d_j - c_j)$ et $\lambda = \frac{\bar{a'}}{\bar{b'}}$. Si $\lambda < a_k$ aller à l'étape 4. Si $J_3 \neq \emptyset$, considérer $a_t = \min a_j, j \in J_3$, alors si $\lambda > a_t$ aller à l'étape 5. Sinon, considérer $\lambda^* = \lambda$ et $\omega_j^* = 1$ pour tout $j \in J_3$ et arrêter.
- Considérer $\omega_j^* = 0$ pour tout $j \in J_2$. $J = J \backslash J_2$ et retourner à l'étape 2.
- Considérer $\omega_j^* = 1$ pour tout $j \in J_1$. $J = J \backslash J_1$, $\bar{a} = \bar{a'}$, $\bar{b} = \bar{b'}$ et retourner à l'étape 2.

Algorithme de Chang et al. (*AFWA : Alternative Fuzzy Weighted Average*)

Partant de la définition de deux points de référence définis ci-après, Chang et al. [43] ont proposé un algorithme pour calculer les deux bornes inférieure et supérieure de la fonction fwa. L'algorithme est similaire aux algorithmes EFWA [129] et MFWA [106]. De plus, il est linéaire. La procédure pour chaque α-coupe est donnée comme suit :

1. Calculer les points de référence l_0 et ρ_0 tels que $l_0 = fwa_L(c_1, \ldots, c_n)$ et $\rho_0 = fwa_U(c_1, \ldots, c_n)$ où fwa_L et fwa_U sont les fonctions définies dans l'algorithme IFWA [132]. Ordonner les coefficients a_i et b_i respectivement en ordre croissant. Considérer $I_0 = \{i | a_i < l_0\}$ et $J_0 = \{i | b_i > \rho_0\}$ pour $i = 1, \ldots, n$ et $p = q = 1$. l_0 et ρ_0 correspondent aux valeurs de L et U calculées dans l'algorithme IFWA.
2. **a)** Pour $\min\{fwa_L\}$
   - Calculer le point de référence $l_p = fwa_L(\omega_1, \ldots, \omega_n)$ où $\omega_i = d_i$ pour $i \in I_{p-1}$ et $\omega_i = c_i$ pour $i \notin I_{p-1}$
   - *test d'optimalité.* Considérer $I_p = \{i \in I_{p-1} | a_i < l_p\}$ et $\Delta I_p = I_{p-1} \backslash I_p$. Si $\Delta I_p = \emptyset$, alors $\min\{fwa_L\} = l_p$ et passer à l'étape 2.2). Sinon, considérer $p = p + 1$ et retourner à l'étape 2.1).

   b) Pour $\max\{fwa_U\}$
   - Calculer le point de référence $\rho_p = fwa_U(\omega_1, \ldots, \omega_n)$ où $\omega_i = d_i$ pour $i \in J_{q-1}$ et $\omega_i = c_i$ pour $i \notin J_{q-1}$
   - *test d'optimalité.* Considérer $J_q = \{i \in J_{q-1} | b_i > \rho_q\}$ et $\Delta J_q = J_{q-1} \backslash J_q$. Si $\Delta J_q = \emptyset$, alors $\max\{fwa_U\} = \rho_q$ et arrêter l'étape 2. Sinon, considérer $q = q - 1$ et retourner à l'étape 2.1).
3. Répéter les étapes (1) et (2) pour plusieurs α-coupes.

Algorithme de Fortin

La fonction de la moyenne pondérée $fwa(.)$ est croissante par rapport aux variables x_i. De plus, si les coefficients x_i sont ordonnés tels que $\forall j < i, \; x_j \leq x_i$, alors il existe $k \in [1, n-1]$ tel que pour tout $i \leq k$, $fwa(.)$ est une fonction croissante par rapport à chaque argument ω_i, et pour tout $i > k$, $fwa(.)$ est décroissante par rapport à chaque argument ω_i. On peut alors appliquer le théorème démontré dans [90], et calculer les deux bornes inférieure et supérieure de la moyenne pondérée. On a alors :

$$\min\{fwa_L\} = \min_{i=1,\ldots,n} \left[\frac{\sum_{j=1}^{i} d_i x_j + \sum_{j=i+1}^{n} c_i x_j}{\sum_{j=1}^{i} d_i + \sum_{j=i+1}^{n} c_i} \right] \tag{B.38}$$

$$\max\{fwa_U\} = \max_{i=1,\ldots,n} \left[\frac{\sum_{j=1}^{i} c_i x_j + \sum_{j=i+1}^{n} d_i x_j}{\sum_{j=1}^{i} c_i + \sum_{j=i+1}^{n} d_i} \right] \tag{B.39}$$

L'algorithme de Fortin, de complexité linéaire, est facile à utiliser dans les cas pratiques où les coefficients x_i sont ordonnés.

Algorithmes de Karnik-Mendel (*KMFWA : Karnik Mendel Fuzzy Weighted Average*)

Karnik et Mendel [114, 133] ont proposé l'algorithme KMFWA pour calculer les bornes de fwa. Récemment, Wu et Mendel [65] ont étendu KMFWA en EKMFWA, qui est la méthode la plus efficace pour le calcul de la moyenne floue pondérée. Les deux algorithmes KMFWA et EKMFWA ont une complexité linéaire. Ils sont fondés sur le calcul de la dérivée partielle de la fonction de la moyenne pondérée comme dans l'algorithme de Fortin et al. [90]. Nous présentons ci-dessous les algorithmes KMFWA puis EKMFWA pour le calcul des deux bornes de la moyenne floue pondérée.

A. L'algorithme KMFWA

Cet algorithme a été introduit en 2001 par Karnik et Mendel. Les étapes de calcul de la borne supérieure et inférieure sont similaires, nous ne détaillons ici que le calcul de la borne supérieure.

0) Dans la formule de la définition de la moyenne floue pondérée $\frac{\sum_{i=1}^{n} \omega_i x_i}{\sum_{i=1}^{n} \omega_i}$, les coefficients x_i sont remplacés par les coefficients b_i ordonnés par ordre croissant ($b_1 \leq b_2 \leq \ldots \leq b_n$) en associant à chaque coefficient son poids correspondant.

1) Initialiser, pour $i = 1, \ldots, n$, les poids ω_i comme suit :

1. $\omega_i = \frac{c_i + d_i}{2}$
2. $\omega_i = c_i$ pour $i \leq \left\lfloor \frac{n+1}{2} \right\rfloor$ et $\omega_i = d_i$ pour $i > \left\lfloor \frac{n+1}{2} \right\rfloor$ où $\left\lfloor \frac{n+1}{2} \right\rfloor$ est le premier entier plus petit ou égal à $\frac{n+1}{2}$.

Calculer ensuite la valeur suivante :

$$f'_{max} = \frac{\sum_{i=1}^{n} \omega_i b_i}{\sum_{i=1}^{n} \omega_i}$$

2) Chercher $k \in 1, \ldots, n-1$, tel que $b_k \leq f'_{max} \leq b_{k+1}$

3) Prendre $\omega_i = c_i$ pour $i \leq k$ et $\omega_i = d_i$ pour $i \geq k+1$ et calculer f''_{max} avec

$$f''_{max} = \frac{\sum_{i=1}^{k} b_i c_i + \sum_{i=k+1}^{n} b_i d_i}{\sum_{i=1}^{k} c_i + \sum_{i=k+1}^{n} d_i}$$

4) Tester si les deux valeurs f'_{max} et f''_{max} sont égales. Si elles le sont alors arrêter le processus et la borne supérieure de la fonction vaut f'_{max}, sinon passer à l'étape 5.

5) Prendre $f'_{max} = f''_{max}$ et aller à l'étape 2.

Pour la borne inférieure, il suffit de remplacer les coefficient b_i par a_i dans toutes les étapes de l'algorithme et de calculer le minimum de la même manière. Les valeurs des bornes inférieure et supérieure calculées dans l'algorithme KMFWA étant calculées pour une α-coupe, il suffit de choisir plusieurs valeurs pour α, de calculer leurs bornes correspondantes puis d'appliquer le minimum et le maximum pour obtenir la plus petite et la plus grande valeur atteintes par la fonction fwa.

B. L'algorithme EKMFWA

Wu et al. dans [65] ont étendu l'algorithme de Karnik et Mendel [114] et ont démontré que leur algorithme EKMFWA est 40% plus rapide que l'algorithme KMFWA, ce qui signifie que cet algorithme est le meilleur parmi tous les algorithmes déjà cités. Nous ne détaillons que le calcul de la borne supérieure de la fonction fwa et nous citons d'une manière similaire le calcul de la borne inférieure de la moyenne floue pondérée.

1) Ordonner et renommer les coefficients b_i pour $i = 1, \ldots, n$ en ordre croissant tels que $b_1 \leq b_2 \leq \ldots b_n$, en affectant les poids respectifs aux coefficients de la séquence obtenue.

2) Prendre $k = \frac{n}{1.7}$ (le plus proche entier) et calculer les valeurs de A, B et Y_{max} données par

$$A = \sum_{i=1}^{k} b_i c_i + \sum_{i=k+1}^{n} b_i d_i$$

$$B = \sum_{i=1}^{k} c_i + \sum_{i=k+1}^{n} d_i$$

$$Y_{max} = \frac{A}{B}$$

3) Chercher $k' \in \{1, \ldots, n-1\}$ tel que $b'_k \leq Y_{max} \leq b_{k'+1}$

4) Tester si k et k' sont égaux. Si oui arrêter et la borne supérieure est égale à $Y_{max} = \frac{A}{B}$. Sinon continuer l'algorithme.

5) Calculer le signe s où $s = signe(k' - k)$ et

$$A' = A - s \sum_{i=\min(k,k')+1}^{\max(k,k')} b_i(d_i - c_i)$$

$$B' = B - s \sum_{i=\min(k,k')+1}^{\max(k,k')} (d_i - c_i)$$

$$Y'_{max} = \frac{A'}{B'}$$

6) Prendre $Y_{max} = Y'_{max}$, $A = A'$, $B = B'$ et $k = k'$ et aller à l'étape 3.

Pour la borne inférieure, il suffit de changer les coefficients b_i par a_i et c_i par d_i dans les valeurs de A et B. Considérer $k = \frac{n}{2.4}$ et les valeurs de A', B' et Y deviennent :

$$A' = A + s \sum_{i=\min(k,k')+1}^{\max(k,k')} a_i(d_i - c_i)$$

$$B' = B + s \sum_{i=\min(k,k')+1}^{\max(k,k')} (d_i - c_i)$$

$$Y'_{min} = \frac{A'}{B'}$$

Les valeurs de k pour les deux bornes sont choisies suite à une étude de comparaison de nombres d'itérations pour atteindre les bornes de fwa [65].

B.4.2 Exemples

Dans cette section, nous utilisons l'exemple de Dong et Wong pour comparer les algorithmes de calcul de la fonction fwa.

– Exemple de Dong et Wong

Pour illustrer l'exemple, supposons que les valeurs des critères A_i et leurs poids W_i sont exprimés par les nombres flous suivants :

$$\mu_{A_1}(x_1) = \begin{cases} x_1 & \text{si } x_1 \in [0,1] \\ 2 - x_1 & \text{si } x_1 \in [1,2] \end{cases} \quad \mu_{A_1}(w_1) = \begin{cases} w_1/3 & \text{si } w_1 \in [0, 0.3] \\ (0.9 - w_1)/0.6 & \text{si } w_1 \in [0.3, 0.9] \end{cases}$$

$$\mu_{A_2}(x_2) = \begin{cases} x_2 - 2 & \text{si } x_2 \in [2,3] \\ 4 - x_2 & \text{si } x_2 \in [3,4] \end{cases} \quad \mu_{A_2}(w_2) = \begin{cases} (w_2 - 0.4)/0.3 & \text{si } w_2 \in [0, 0.7] \\ (1 - x_2)/0.3 & \text{si } w_2 \in [0.7, 1] \end{cases}$$

$$\mu_{A_3}(x_3) = \begin{cases} x_3 - 4 & \text{si } x_3 \in [4,5] \\ 6 - x_3 & \text{si } x_3 \in [5,6] \end{cases} \quad \mu_{A_3}(w_3) = \begin{cases} (w_3 - 0.6)/0.2 & \text{si } w_3 \in [0.6, 0.8] \\ (1 - w_3)/0.2 & \text{si } w_3 \in [0.8, 1] \end{cases}$$

Si on décrit les nombres flous pour une coupe de niveau α, on obtient pour A_i et W_i respectivement, pour $i = \{1, 2, 3\}$, les sous-ensembles suivants : $A_{1\alpha} = [\alpha, 2-\alpha], A_{2\alpha} = [2+\alpha, 4-\alpha], A_{3\alpha} = [4+\alpha, 6-\alpha], W_{1\alpha} = [0.3\alpha, 0.9-0.6\alpha], W_{2\alpha} = [0.4+0.3\alpha, 1-0.3\alpha], W_{3\alpha} = [0.6+0.2\alpha, 1-0.2\alpha]$. Pour la coupe $\alpha = 0.5$, on obtient :

	$i=1$	$i=2$	$i=3$
critère i	$[0.5, 1.5]$	$[2.5, 3.5]$	$[4.5, 5.5]$
poids i	$[0.15, 0.6]$	$[0.55, 0.85]$	$[0.7, 0.9]$

Solution de l'algorithme VFWA

Comme indiqué dans la Section B.4.1, il existe 2^6 permutations pour la variable $(x_1, x_2, x_3, \omega_1, \omega_2, \omega_3)$, par exemple $(0.5, 2.5, 4.5, 0.15, 0.55, 0.85)$ qui donne $fwa = 3.28$. Le calcul de la valeur correspondant à chaque permutation permet de calculer fwa_L et fwa_U, l'intervalle résultant étant alors $[2.59, 4.44]$.

Solution de l'algorithme IFWA

En partant de la définition de Liou et Wong, on a :

$$fwa_L = \frac{0.5\omega_1 + 2.5\omega_2 + 4.5\omega_3}{\omega_1 + \omega_2 + \omega_3} \quad \text{et} \tag{B.40}$$

$$fwa_U = \frac{1.5\omega_1 + 3.5\omega_2 + 5.5\omega_3}{\omega_1 + \omega_2 + \omega_3} \tag{B.41}$$

D'où, les étapes de l'algorithme :

1. Puisque $(c_1, c_2, c_3) = (0.15, 0.55, 0.7)$ alors $L = fwa_L(c_1, c_2, c_3) = fwa_L(0.15, 0.55, 0.7) = 4.6/1.4 = 3.2857$ et $U = fwa_U(c_1, c_2, c_3) = fwa_U(0.15, 0.55, 0.7) = 6/1.4 = 4.2857$
2. $I = \{1, 2\} \neq \emptyset$, $J = \{3\} \neq \emptyset$, $R_L = R_U = \emptyset$ donc $w_1 = c_1, w_2 = c_2, w_3 = c_3$
   – $L_1 = fwa_L(0.6, 0.55, 0.7) = 4.825/1.85 = 2.6081$ et $L_2 = fwa_L(0.15, 0.85, 0.7) = 5.35/1.7 = 3.1471$

- $L = \min\{L_1, L_2\} = L_1$ alors $m = 1$ et $R_L = \{1\}$, $I = \{2\}$ et $T = \emptyset$
- $I = \{2\} \neq \emptyset$, répéter (2). Nous avons $l_2 = fwa_L(0.6, 0.85, 0.7) = 5.575/2.15 = 2.59$, alors $L = l_2 \Rightarrow m = 2$, $R_L = \{1, 2\}$, $I = \emptyset$, et $T = \emptyset$. Nous arrêtons l'étape (2) et obtenons $L = 2.59$

3. $J = \{3\} \neq \emptyset$, $U = 6/1.4$, $R_U = \emptyset$, donc $w_i = c_i$ for $i = 1, \ldots, 3$
   - $u_3 = fwa_U(0.15, 0.55, 0.9) = 7.1/1.6 = 4.44$
   - $U = \max\{4.44\} = 4.44$, c-à-d $m = 3$, par conséquent $R_U = \{3\}$, $J = \emptyset$, et $T = \emptyset$. Nous arrêtons l'étape (3) et $U = 4.44$. L'intervalle obtenu pour $\alpha = 0.5$ est $[2.59, 4.44]$.

Solution de l'algorithme PFWA

Nous avons deux boucles comprenant chacune quatre étapes. En effet :

Boucle 1 : Pour la borne inférieure de la fonction fwa (respectivement la borne supérieure)

1. La plus petite valeur pour les critères est 0.5 (respectivement 1.5) et la plus grande valeur est 4.5 (respectivement 5.5)
2. Les poids correspondants sont $c_1 = 0.6$ et $d_n = 0.7$
3. $a' = 2.65$ (respectivement $b' = 4.929$) et $c' = d' = w' = 1.3$ (respectivement $c' = d' = w' = 1.05$)
4. On élimine les coefficients 0.5 et 4.5 (respectivement 1.5 et 5.5), et leurs poids correspondants $[0.15, 0.6]$ et $[0.7, 0.9]$ (respectivement $[0.15, 0.6]$ et $[0.7, 0.9]$), et on les remplace par $a' = 2.65$ (respectivement $b' = 4.929$) et $c' = d' = 1.3$ (respectivement $c' = d' = 1.05$)

Boucle 2 : Pour la borne inférieure de la fonction fwa (respectivement la borne supérieure)

1. La plus petite valeur pour les critères est 2.5 et 2.65 (respectivement 3.5 et 4.929)
2. Les poids correspondants sont $c_1 = 0.85$ et $d_n = 1.3$
3. $a' = 2.59$ et $c' = d' = w' = 2.15$ (respectivement $b' = 4.44$ et $c' = d' = w' = 1.6$)
4. Éliminons les coefficients 2.5 et 2.65 (respectivement 3.5 et 4.929) et leurs poids correspondants $[0.55, 0.85]$ et $[1.3, 1.3]$ (respectivement $[0.55, 0.85]$ et $[1.05, 1.05]$) ; ensuite considérons $a' = 2.59$ (respectivement $b' = 4.44$) et $c' = d' = 2.15$ (respectivement $c' = d' = 1.6$).

Par conséquent, pour $\alpha = 0.5$, nous avons $fwa \in [2.59, 4.44]$.

Solution de l'algorithme EFWA

1. Ordonner les coefficients a_i pour la borne inférieure de fwa (respectivement b_i pour la borne supérieure), la séquence résultante est $(a_1, a_2, a_3) = (0.5, 2.5, 4.5)$ (respectivement $(b_1, b_2, b_3) = (1.5, 3.5, 5.5)$), considérons $premier = 1$ (respectivement $premier = 1$), et $dernier = 3$ (respectivement $dernier = 3$).
2. $\delta = 2$ (respectivement $\xi = 2$), nous obtenons $\delta_{S_2} = 0.093$ et $\delta_{S_3} = -1.907$ (respectivement $\xi_{S_2} = 0.938$ et $\xi_{S_3} = -1.063$).
3. Nous avons $\delta_{S_2} > 0$ et $\delta_{S_3} < 0$, alors $L = fwa_L(d_1, d_2, c_3) = 2.59$. En ce qui concerne la borne supérieure, nous avons $\xi_{S_2'} > 0$ et $\xi_{S_3'} < 0$. Par conséquent, $U = fwa_U(c_1, c_2, d_3) = 4.44$.

L'intervalle final de fwa est donc $[2.59, 4.44]$ pour $\alpha = 0.5$.

Solution de l'algorithme LPFWA

Appliquons la procédure de l'algorithme LPFWA en considérant $f_i = \omega_i z$ $(fwa = f)$, nous obtenons les systèmes linéaires suivants :

$$\begin{cases} \min 0.5f_1 + 2.5f_2 + 4.5f_3 \\ 0.15z \leq f_1 \leq 0.6z \\ 0.55z \leq f_2 \leq 0.85z \\ 0.7z \leq f_3 \leq 0.9z \\ f_1 + f_2 + f_3 = 1 \\ z \geq 0 \end{cases} \qquad \begin{cases} \max 1.5f_1 + 3.5f_2 + 5.5f_3 \\ 0.15z \leq f_1 \leq 0.6z \\ 0.55z \leq f_2 \leq 0.85z \\ 0.7z \leq f_3 \leq 0.9z \\ f_1 + f_2 + f_3 = 1 \\ z \geq 0 \end{cases}$$

Nous obtenons l'intervalle $[2.59, 4.44]$. Nous avons utilisé le programme LINDO pour résoudre ces systèmes.

Solution de l'algorithme DFWA

1. Ordonner les coefficients a_i et renommer les indices des poids en respectant l'ordre des coefficients a_i.

	$i = 1$	$i = 2$	$i = 3$
$[c_i, d_i]$	$[0.15, 0.6]$	$[0.55, 0.85]$	$[0.7, 0.9]$

2. Déterminer la valeur de la fonction τ : $\tau = (2.1, 0.93, -1.90)$ (respectivement $\sigma = (1.842, 0.938, -1.063)$)
3. Nous avons $\tau(2) > 0$, $\tau(3) < 0$ pour $\min\{f_L\}$ et $\sigma(2) > 0$, $\sigma(3) < 0$ pour $\max\{f_U\}$.
4. $f_L = f(d_1, d_2, c_3) = 2.59$ et $f_U = f(c_1, c_2, d_3) = 4.44$
5. L'intervalle recherché est alors $[2.59, 4.44]$ pour $\alpha = 0.5$.

Solution de l'algorithme PPFWA

Les systèmes linéaires obtenus par la méthode de Kao et Liu sont les suivants pour notre exemple :

$$\begin{cases} \min \alpha f_1 + (2+\alpha) f_2 + (4+\alpha) f_3 \\ (0.3\alpha)t \leq f_1 \leq (0.9 - 0.6\alpha)t \\ (0.4 + 0.3\alpha)t \leq f_2 \leq (1 - 0.3\alpha)t \\ (0.6 + 0.2\alpha)t \leq f_3 \leq (1 - 0.2\alpha)t \\ f_1 + f_2 + f_3 = 1 \\ t \geq 0 \end{cases} \qquad \begin{cases} \max (2-\alpha) f_1 + (4-\alpha) f_2 + (6-\alpha) f_3 \\ (0.3\alpha)t \leq f_1 \leq (0.9 - 0.6\alpha)t \\ (0.4 + 0.3\alpha)t \leq f_2 \leq (1 - 0.3\alpha)t \\ (0.6 + 0.2\alpha)t \leq f_3 \leq (1 - 0.2\alpha)t \\ f_1 + f_2 + f_3 = 1 \\ t \geq 0 \end{cases}$$

Les intervalles obtenus repectivement pour les α-coupes 0, 0.5 et 1 sont $[1.68, 5.43]$, $[2.59, 4.44]$ et $[3.55, 3.55]$ en utilisant l'outil LINDO. De plus Kao et Liu ont transformé ces deux systèmes en deux équations paramétrées par α en utilisant les gradiants des deux bornes.

Solution de l'algorithme MFWA

En considérant l'algorithme MFWA pour notre exemple, nous trouvons deux itérations pour la borne inférieure et une seule itération pour la borne supérieure. L'intervalle obtenu est $[2.59, 4.44]$.

Solution de l'algorithme AFWA

1. Calculer l_0 et ρ_0. Nous obtenons alors $l_0 = fwa_L(c_1, \ldots, c_n) = 3.28$ et $\rho_0 = fwa_U(c_1, \ldots, c_n) = 4.28$. Alors, $I_0 = \{1, 2\}$, $J_0 = \{3\}$.

2. Pour $\min\{fwa_L\}$, nous calculons $l_1 = 2.59$ et $I_1 = \{0, 1\}$ alors $\Delta I_1 = \emptyset$ ce qui donne $\min\{f_L\} = l_1 = 2.59$.
3. Pour $\max\{fwa_U\}$, nous calculons $\rho_1 = 4.44$ et $J_1 = \{3\}$, alors $\Delta J_1 = \emptyset$ ce qui donne $\min\{f_U\} = \rho_1 = 4.44$.
4. L'intervalle résultant pour fwa est $[2.59, 4.44]$

Exemple de KMFWA

Pour l'algorithme KMFWA, il existe deux itérations pour la borne inférieure et deux itérations pour la borne supérieure tandis que pour l'algorithme EKMFWA nous trouvons une seule itération pour la borne inférieure et la borne supérieure est obtenue dès la première comparaison. L'intervalle obtenu est $[2.59, 4.44]$ pour $\alpha = 0.5$.

B.4.3 Discussion

La première solution proposée par Dong et Wong [64] exige 2^{2n} permutations pour un problème à n critères. L'algorithme VFWA de Dong et Wong a ainsi un ordre de complexité de $O(2^n)$. Liou et Wang [132] ont amélioré cet algorithme et ont proposé de ne calculer que $[n(n+1)+2]$ permutations, la complexité étant alors polynomiale en $O(n^2)$. Les algorithmes MFWA de Guu et PFWA de Guh et al. exigent le même nombre d'opérations et ont une complexité linéaire en $O(n)$. En se fondant sur l'efficacité de calcul, il n'y a pas de différence entre l'algorithme EFWA et l'algorithme DFWA. Les deux algorithmes ont une complexité d'ordre $O(n \log n)$. L'algorithme AFWA exige $2n$ opérations arithmétiques et est linéaire. Les algorithmes KMFWA et EKMFWA sont les plus efficaces dans les cas pratiques et avec un grand nombre d'intervalles flous. Nous remarquons que les algorithmes VFWA et IFWA présentent une complexité nettement plus importante que les autres algorithmes. En considérant les outils mathématiques tels que LINDO, nous suggérons la méthode de transformation en systèmes linéaires proposée par Guh et al. en première considération. L'algorithme de Fortin ne peut être utilisé que dans le cas où les coefficients suivent un ordre spécifique, il est linéaire et permet d'avoir une forme exacte pour la fonction fwa.

C

Rappels sur les treillis

C.1 Ensemble ordonné

Définition 14 (Relation binaire) *Une **relation binaire** R entre deux ensembles X et Y est un ensemble de couples d'éléments (x, y) tels que $x \in X$ et $y \in Y$, i.e. un sous-ensemble de $X \times Y$. $(x, y) \in R$ (aussi noté xRy) signifie que l'élément x est en relation par R avec l'élément y. Si $X = Y$, on parle de relation binaire sur X. R^{-1} est la relation inverse de R, i.e. la relation entre Y et X telle que $yR^{-1}x \Leftrightarrow xRy$.*

Définition 15 (Relation d'ordre partiel) *Une relation binaire R sur un ensemble E est dite **relation d'ordre partiel** sur E si elle vérifie les conditions suivantes pour tous $x, y, z \in E$:*

1. *R est reflexive ($(x, x) \in R$)*
2. *R est antisymétrique ($(x, y) \in R$ et $x \neq y$ alors $(y, x) \notin R$)*
3. *R est transitive ($(x, y) \in R$ et $(y, z) \in R$ alors $(x, z) \in R$)*

Une relation d'ordre R est souvent notée $\leq$ (R^{-1} est notée $\geq$) et on dit que "*x est plus petit que y*" lorsque $x \leq y$.

Définition 16 (Ensemble ordonné) *Un **ensemble partiellement ordonné** est un couple $(E, \leq)$ où E est un ensemble et $\leq$ est une relation d'ordre sur E.*

Dans un ensemble ordonné $(E, \leq)$, deux éléments x et y de E sont dits **comparables** lorsque $x \leq y$ ou $y \leq x$, sinon ils sont dits **incomparables**. Pour deux éléments comparables et différents, $x \leq y$ et $x \neq y$, on note $x < y$. Un sous-ensemble de $(E, \leq)$ dans lequel tous les elements sont comparables est appelé **chaîne**. Un sous-ensemble de $(E, \leq)$ dans lequel tous les éléments sont incomparables est appelé **anti-chaîne**.

Définition 17 (Successeur, prédécesseur, couverture) *Soient $(E, \leq)$ un ensemble ordonné et $x, y \in E$. y est dit **successeur** de x lorsque $x < y$ et qu'il n'existe aucun élément $z \in E$ tel que $x < z < y$. Dans ce cas, x est dit **prédécesseur** de y et on note $x \prec y$. Lorsque x est un **prédécesseur** de y on dit que x **couvre** y (et que y est couvert par x). La **couverture** de x est formée par tous ses successeurs.*

Tout ensemble ordonné, $(E, \leq)$, peut être représenté graphiquement par un diagramme appelé "**diagramme de Hasse**" (ou diagramme de couverture) obtenu comme suit :

1. Tout élément de E est représenté par un noeud.

2. Si $x, y \in E$ et $x \prec y$ alors le cercle correspondant à y doit être au-dessus de celui correspondant à x et les deux cercles sont reliés par un segment.

À partir d'un tel diagramme on peut lire la relation d'ordre comme suit : $x < y$ si et seulement s'il existe un chemin ascendant qui relie le cercle correspondant à x à celui correspondant à y.

Définition 18 (Principe de dualité des ensembles ordonnés) *Soit $(E, \leq)$ un ensemble ordonné. La relation inverse $\geq$ de $\leq$ est aussi une relation d'ordre sur E. $\geq$ est appelée **duale** de $\leq$ et $(E, \geq)$ est appelé le dual de l'ensemble ordonné $(E, \leq)$.*

Le diagramme de Hasse de $(E, \geq)$ peut être obtenu à partir de celui de $(E, \leq)$ par une simple réflexion horizontale. De plus, il est possible de dériver les propriétés duales de $(E, \geq)$ à partir des propriétés de $(E, \leq)$.

C.2 Treillis

Définition 19 (Majorant, minorant) *Soient $(E, \leq)$ un ensemble ordonné et S un sous-ensemble de E. Un element $a \in E$ est dit **majorant** de S lorsque $a \geq s \ \forall s \in S$. De façon duale, $a \in E$ est dit **minorant** de S lorsque $a \leq s \ \forall s \in S$.*

Le plus petit majorant (respectivement minorant) de S, s'il existe, est appelé borne supérieure (respectivement borne inférieure) de S et noté $\bigvee S$ (respectivement $\bigwedge S$). Dans le cas où $S = \{x, y\}$, $\bigvee S$ et $\bigwedge S$ sont aussi notés $x \vee y$ et $x \wedge y$ respectivement.

Dans tout ensemble ordonné, lorsque la borne supérieure (respectivement la borne inférieure) existe, elle est unique.

Définition 20 (Treillis, treillis complet) *Un **treillis** est un ensemble partiellement ordonné $(E, \leq)$ tel que $x \vee y$ et $x \wedge y$ existent pout tout couple d'éléments $x, y \in E$. Un treillis est dit **complet** si $\bigvee S$ et $\bigwedge S$ existent pour tout sous-ensemble S de E. En particulier, un treillis complet admet un élément maximal (top) noté $\top$ et un élément minimal (bottom) noté $\bot$.*

Tout treillis fini est un treillis complet.

Définition 21 (Demi-treillis) *Un ensemble ordonné $(E, \leq)$ est un **sup-demi-treillis** (respectivement **inf-demi-treillis**) si tout couple d'éléments $x, y \in E$ admet une borne supérieure $x \vee y$ (respectivement une borne inférieure $x \wedge y$).*

C.3 Connexion de Galois

Définition 22 (Connexion de Galois) *Soient $\varphi : P \to Q$ et $\psi : Q \to P$ deux applications entre deux ensembles ordonnés $(P, \leq_P)$ et $(Q, \leq_Q)$. φ et ψ forment une **connexion de Galois** entre $(P, \leq_P)$ et $(Q, \leq_Q)$ si elles vérifient les conditions suivantes pour tous $p, p_1, p_2 \in P$ et $q, q_1, q_2 \in Q$:*

1. *si $p_1 \leq_P p_2$ alors $\varphi(p_2) \leq_Q \varphi(p_1)$,*
2. *si $q_1 \leq_Q q_2$ alors $\psi(q_2) \leq_P \psi(q_1)$,*
3. *$p \leq_P \psi(\varphi(p))$ et $q \leq_Q \varphi(\psi(q))$.*

Les conditions données dans la définition précédente sont équivalentes à la formule suivante :

$$p \leq_P \psi(q) \Leftrightarrow q \leq_Q \varphi(p)$$

D

Exemple de calcul de l'indicateur pesticide pour un programme de traitement

Dans cette annexe, nous présentons un exemple de calcul pour un traitement. Nous considérons un traitement de cinq matières actives : isoproturon, epoxyconazole, glyphosate, krésoxyme et metsulfuron méthyle. Les variables de milieu sont données suivant leurs règles. On utilise pour cet exemple $ruissellement = 0.4$ et $dérive = 0$. La dose utilisée est 1200. Dans un premier temps, nous choisissons le mode de fusion à utilisé, qui est ici la fusion des SMC puisque les informations de fiabilité et de dépendance entre les sources ne sont pas disponibles. Ensuite, nous utilisons la méthode de pré-traitement présentée dans le chapitre 3 pour construire la structure de patrons. Dans cet exemple, les données sont plus volumineux que l'exemple de la matière active sulcotrione utilisée dans le chapitre 5, par conséquent l'expert choisit des seuils pour éviter un grand nombre de concepts ayant des valeurs imprécises et qui ne sont pas valides pour calculer les indicateurs. A l'aide de l'opérateur de fusion contrainte par un seuil de similarité, nous construisons les treillis correspondants à chaque matière active. Enfin, l'analyste choisit un concept de chaque treillis pour calculer les indices. Nous nous intéressons à l'indicateur pesticide de chacune des matières actives pour pouvoir calculer un indicateur pour le traitement. La table D.1 présente les valeurs des seuils de similarité choisis pour chaque matière active. Ces valeurs sont choisis en respectant les valeurs des variables ainsi que les conditions de partitionnement en classes favorable et défaroble les valeurs des variables intervenant dans le calcul des indicateurs. Enfin, les résultats des indicateurs pesticides pour chacune des matières active est présenté dans la Table D.2. Chaque pesticide représenté par une ligne de la table est un concept choisi dans le treillis obtenu pour chacun des pesticides étudiés. Notons que nous pouvons ainsi utiliser d'autres concepts à condition que le concept soit valide (c-à-d il contient toutes les valeurs pour toutes les variables intervenant dans le calcul de l'indicateur. L'intention du concept ne doit pas contenir l'élément $*$ qui désigne la non-similarité des sources.

L'indicateur phytosanitaire se calcule pour un programme de traitement à partir des indicateurs I_{phy} pour chaque matière active utilisée dans le traitement [176]. En effet, chaque indicateur est transformé entre 0 (absence de risque) et 1 (risque maximum) par translation et pondéré par un coefficient avec le risque. Ces coefficients restent faibles tant que l'indicateur I_{phy} pour chaque matière active est plus grand que 7. Donc, pour une matière active i on a : $k_i = |(-1.7175)\exp(-0.2913 * I_{phy})|$. L'indicateur I_{phy} du traitement se calcule en utilisant

Seuil / Pesticide	θ_{DT50}	θ_{koc}	θ_{DJA}	θ_{KH}	$\theta_{aquatox}$
Isoproturon	100	20	0	90	10^{-9}
Glyphosate	100	2200	0	50	0.001
Métsulfuron méthyle	120	20	0.01	0.01	0
Krésoxyme	0	0	0.01	0.05	10^{-2}
Epoxiconazole	200	1000	0.05	6	0

TABLE D.1 – Seuils choisis pour chacune des matières actives

	$DT50$	koc	DJA	aqu	KH	I_{phy}
Glyphosate	$[18, 47]$	24000	0.3	$[0.64, 15]$	$[2.7 * 10^{-10}, 2.1 * 10^{-7}]$	$[4.9, 6.1]$
Isoproturon	$[12, 28]$	$[80, 105]$	$6 * 10^{-4}$	$[0.03, 9]$	$4.7 * 10^{-9}$	$[4.8, 6.5]$
Krésoxyme	$[25, 34]$	$[219, 372]$	0.4	$[0.02, 0.06]$	$1.45 * 10^{-7}$	$[5.7, 6.7]$
Metsulfuron Méthyle	$[27, 34]$	$[35, 51]$	0.0125	100	$1.1 * 10^{-6}$	$[8.11, 8.15]$
Epoxyconazole	$[62, 187]$	1200	0.04	0.01	$4.7 * 10^{-4}$	$[5.7, 6.3]$

TABLE D.2 – Variables multi-sources intervenant dans le calcul de l'indicateur pesticide avec les résultats pour les matières actives : glyphosate, isoproturon, krésoxyme, metsulfuron méthyle et époxyconazole

l'équation :

$$I_{phy} = \min(I_{phy_i}) - \sum_{i=1}^{n} k_i * \%(\text{surface traitée}) \frac{10 - I_{phy_i}}{10} \tag{D.1}$$

où n est le nombre des matières actives utilisées dans le traitement, I_{phy_i} est l'indicateur phytosanitaire de la matière active i, %(surface traitée) est le pourcentage de la surface traitée par le produit. Dans cet exemple, nous considérons que le produit est appliqué sur le champ entier donc le pourcentage de la surface traitée est 100%.

La variation finale de l'indicateur pesticide d'un traitement se calcule en utilisant les techniques d'analyse d'intervalles présentés dans l'annexe B. Pour cet exemple, l'indicateur global du traitement est $[4.06, 5.68]$. Ce qui indique que l'utilisation du produit phytosanitaire contenant les pesticides isoproturon, glyphosate, krésoxyme, époxiconazole et metsulfuron méthyle puisque la valeur de l'indicateur ne dépasse pas le seuil de référence aceptable par l'environnement.

Bibliographie

[1] A.M. Abidi and R.C. Gonzalez. *Data fusion in robotics and machine intelligence*. Academic Press Professional, Inc., San Diego, CA, USA, 1992.

[2] A.Neumaier. Taylor forms - use and limits, 2002.

[3] A.M. Anile, S. Deodato, and G. Privitera. Implementing fuzzy arithmetic. *Fuzzy Sets Syst.*, 72(2) :239–250, 1995.

[4] Z. Assaghir, P. Girardin, and A. Napoli. Fuzzy logic approach to represent and propagate imprecision in agri-environmental indicator assessment. In *IFSA/EUSFLAT Conf.*, pages 707–712, 2009.

[5] Z. Assaghir, Philippe Girardin, and A. Napoli. Utilisation de la théorie des possibilités pour calculer un indicateur agri-envirennemental. In *Rencontres francophones sur la Logique Floue et ses Applications (LFA)*, 2008.

[6] Z. Assaghir, M. Kaytoue, N. Messai, and A. Napoli. An approach based on formal concept analysis for mining numerical data. In *Seventh Scientific Meeting of the Classification and Data Analysis Group of the Italian Statistical Society*, Septembre 2009.

[7] Z. Assaghir, M. Kaytoue, N. Messai, and A. Napoli. On the mining of numerical data with formal concept analysis and similarity. In *Société Francophone de Classification*, pages 121–124, Septembre 2009.

[8] Z. Assaghir, M. Kaytoue, A. Napoli, and H. Prade. Managing Information Fusion with Formal Concept Analysis. In *Modeling Decisions for Artificial Intelligence, 6th International Conference (MDAI)*, Lecture Notes in Artificial Intelligence, pages 104–115. Springer, 2010.

[9] Z. Assaghir, M. Kaytoue, A. Napoli, and H. Prade. Organisation de la fusion d'information avec l'analyse formelle de concepts. In *Rencontres francophones sur la Logique Floue et ses Applications (LFA)*, 2010. (accepted).

[10] Z. Assaghir, M. Kaytoue, A. Napoli, and H. Prade. Organisation de la fusion d'information avec l'analyse formelle de concepts - Journées Nationales de l'Intelligence Artificielle Fondamentale (IAF), Juin 2010.

[11] Z. Assaghir, M. Kaytoue, and H. Prade. A possibility theory-oriented discussion of conceptual pattern structures. In *Scalable Uncertainty Management, 4th International Conference (SUM)*, Lecture Notes in Artificial Intelligence. Springer, 2010.

[12] Z. Assaghir, M. Kaytoue, H. Prade, and A. Napoli. Fusion de données imparfaites en utilisant l'analyse formelle de concepts. *Fouille de données complexes – Complexité liée aux données multiples, Numéro spécial de la Revue des Nouvelles Technologies de l'Information (RNTI)*, 2010. (In press).

[13] P. Aurousseau, C. Gascuel-Odoux, and H. Squividant. Eléments pour une méthode d'évaluation d'un risque parcellaire de contamination des eaux superficielles par les pesticides. In *Etude et Gestion des Sols*. Association française d'étude des sols, 1998.

[14] A. Ayoun and P. Smets. Data association in multi-target detection using the transferable belief model. *International Journal of Intelligent Systems*, 16 :1167–1182, 2000.

[15] Z. Azmeh, M. Huchard, C. Tibermacine, C. Urtado, and S. Vauttier. Wspab : A tool for automatic classification and selection of web services using formal concept analysis. *ECOWS'08 IEEE Sixth European Conference on Web Services*, pages 31–40, November 2008.

[16] S.M. Baas and H. Kwakernaak. Rating and ranking of multiple-aspect alternatives using fuzzy sets. *Automatica*, 13 :47–58, 1977.

[17] C. Baudrit. *Représentation et propagation de connaissances imprécises et incertaines : Application à l'évaluation des risques liés aux sites et sols pollués.* PhD thesis, Université Paul Sabatier, 2005.

[18] R. Belohlávek. Lattices generated by binary fuzzy relations. *Tatra Mountains Mathematical Publications*, 16 :11 – 19, 1999.

[19] R. Belohlavek. *Fuzzy Relational Systems : Foundations and Principles.* Kluwer Academic Publishers, Norwell, MA, USA, 2002.

[20] R. Belohlávek and V. Vychodil. What is a fuzzy concept lattice ? In *3rd Internationnal Conference on Concept Lattices and Their Applications CLA 2005, September 7-9*, pages 34 – 45, Olomouc, Czech Republic, 2005.

[21] R. Bendaoud, A. Napoli, and Y. Toussaint. Formal concept analysis : A unified framework for building and refining ontologies. In *EKAW*, pages 156–171, 2008.

[22] S. Benferhat, D. Dubois, and H. Prade. Reasoning in inconsistent stratified knowledge bases. In *proc. ISMVL*, pages 184–189, 1996.

[23] G. Birkhoff. *Lattice Theory*, volume 25 of *ASM Colloquium Publications*. AMS, Providence, RI, 3rd edition, 1967. 1st ed., 1940 ; 2nd ed., 1948.

[24] I. Bloch. Fusion de données, ensembles flous et morphologie mathématique en traitement d'images, application à l'imagerie médicale cérébrale et cardiovasculaire multi-modalités. Technical report, ENST 95D007, partie scientifique du rapport présenté pour l'obtention de l'Habilitation à Diriger des Recherches de l'Université René Descartes (Paris 5), U.F.R. de Mathématiques et Informatique, May 1995.

[25] I. Bloch. *Fusion d'informations en traitement du signal et des images.* Hermès, Paris, France, 2003.

[26] I. Bloch, A. Hunter, A. Ayoun, S. Benferhat, P. Besnard, L. Cholvy, R. Cooke, D. Dubois, and H. Fargier. Fusion : general concepts and characteristics. *International Journal of Intelligent Systems*, 16 :1107–1134, 2001.

[27] I. Bloch and H. Maître. Fusion de données en traitement d'images : modèles d'information et décisions. *Traitement du Signal*, 11(6) :435–446, 1994.

[28] H. Bock and E. Diday, editors. *Analysis of Symbolic Data : Exploratory Methods for Extracting Statistical Information from Complex Data*, volume 15 of *Studies in Classification, Data Analysis, and Knowledge Organization*, Secaucus, NJ, USA, February 2000. Springer-Verlag New York, Inc.

[29] C. Bockstaller and P. Girardin. The crop sequence indicator : a tool to evaluate crop rotations in relation to the requirements of integrated arable farming systems. *Aspects of applied biology.*, 47 :405–408, 1996.

[30] C. Bockstaller and P. Girardin. Agro-ecological indicators : instruments to assess sustainability in agriculture. *Nachhaltigkeit in der Landwirtschaft*, pages 69–83, 1999.

[31] C. Bockstaller and P. Girardin. Évaluer les systèmes de culture à l'aide d'indicateurs agri-environnementaux : la méthode indigo. *Les Rencontres Annuelles du CETIOM*, pages 54–58, 2002.

[32] C. Bockstaller and P. Girardin. How to validate environmental indicators. *Agricultural Systems.*, 76 :639–653, 2003.

[33] C. Bockstaller and P. Girardin. *Mode de calcul des indicateurs agro-écologiques.* INRA-ARAA, 2003.

[34] J. Bordat. Calcul pratique du treillis de galois d'une correspondance. *Mathématiques et Sciences Humaines*, 96 :31–47, 1986.

[35] Patrick Bosc and Henri Prade. An introduction to the fuzzy set and possibility theory-based treatment of soft queries and uncertain or imprecise databases, 1994.

[36] A. Bárdossy, A. Bronstert, and B. Merz. 1-, 2- and 3-dimensional modeling of water movement in the unsaturated soil matrix using a fuzzy approach. *Advances in Water Resources*, 18(4) :237 – 251, 1995.

[37] B.G. Buchanan and E.H. Shortliffe. *Rule-Based Expert Systems, The Mycin experiments of the Stanford Heuristic Programming Project.* Addison-Wesley Publishing Company, 1984.

[38] A. Burusco and R. Fuentes-González. Concept lattices defined from implication operators. *Fuzzy Sets Syst.*, 114(3) :431–436, 2000.

[39] A. Burusco and R. Fuentes Gonzalez. The study of the l-fuzzy concept lattice. *Mathware and soft computing*, 1994.

[40] L. De Campos, J.F. Huete, and S. Moral. Probability intervals : A tool for uncertain reasoning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2(2) :167–196, 1994.

[41] C. Carpineto and G. Romano. *Concept Data Analysis : Theory and Applications.* John Wiley and Sons, 2004.

[42] C. Carpineto and G. Romano. Using concept lattices for text retrieval and mining. In *Formal Concept Analysis*, pages 161–179, 2005.

[43] P.T. Chang, K.C. Hung, K.P. Lin, and C.H. Chang. A comparison of discrete algorithms for fuzzy weighted average. *Fuzzy Systems, IEEE Transactions on*, 14(5) :663–675, Oct. 2006.

[44] L. Chaudron and N. Maille. Generalized formal concept analysis. In *ICCS*, pages 357–370, 2000.

[45] L. Chaudron and N. Maille. Le modèle des cubes : représentation algébrique des conjonctions de propriétés. In *RFIA*, 2000.

[46] M. Chein. Algorithme de recherche des sous-matrices premières d'une matrice. *Bull. Math. Soc. Sci. Math. R.S. Roumanie*, 13 :21–25, 1969.

[47] K.P. Chiao. Direct fuzzy weighted average algorithm. *Journal of Mathematical Sciences*, 16 :311–327, 2000.

[48] L. Cholvy. Proving theorems in a multi-source environment. In *IJCAI*, pages 66–73, 1993.

[49] L. Cholvy. Automated reasoning with merged contradictory information whose reliability depends on topics. In *ECSQARU*, pages 125–132, 1995.

[50] L. Cholvy. Reasoning about merged information. In *Belief Change*, volume 3 of *Handbook of defeasible Reasoning and Uncertainty management Systems*, pages 233–263, 1998.

[51] L. Cholvy and S. Moral. Merging databases : Problems and examples. *Int. J. Intell. Syst.*, 16(10) :1193–1221, 2001.

[52] R. Cooke. *Experts in uncertainty.* Oxford University Press, 2001.

[53] R.T. Cox. Probability, frequency and reasonable expectation. *American Journal of Physics*, 14(1) :1–13, 1946.

[54] B. Davey and H. Priestley. *Introduction to Lattices and Order.* Cambridge University Press, 2nd edition edition, April 2002.

[55] S. Destercke. *Uncertainty representation and combination : new results with application to nuclear safety issues.* PhD thesis, Université Paul Sabatier, 2008.

[56] S. Destercke, D. Dubois, and E. Chojnacki. Unifying practical uncertainty representations - i : Generalized p-boxes. *International Journal of Approximate Reasoning*, 49(3) :649 – 663, 2008.

[57] S. Destercke, D. Dubois, and E. Chojnacki. Unifying practical uncertainty representations. ii : Clouds. *International Journal of Approximate Reasoning*, 49(3) :664 – 677, 2008.

[58] S. Destercke, D. Dubois, and E. Chojnacki. Possibilistic information fusion using maximal coherent subsets. *IEEE Transactions on Fuzzy Systems*, 17 :79–92, 2009.

[59] M. Devignes, P. Franiatte, N. Messai, A. Napoli, and M. Smaïl-Tabbone. Bioregistry : automatic extraction of metadata for biological database retrieval and discovery. In *iiWAS*, pages 456–461, 2008.

[60] Y. Djouadi, Dubois D., and Prade H. On the possible meanings of degrees when making formal concept analysis fuzzy. In *Preference Modelling and Decision Analysis Workshop*, pages 253–258, 2009.

[61] Y. Djouadi, D. Dubois, and H. Prade. Différentes extensions floues de l'analyse formelle de concepts. In *Rencontres francophones sur la Logique Floue et ses Applications (LFA)*, pages 141–148, 2009.

[62] Y. Djouadi, D. Dubois, and H. Prade. Possibility theory and formal concept analysis : Context decomposition and uncertainty handling. In Eyke Hüllermeier, Rudolf Kruse, and Frank Hoffmann, editors, *Computational Intelligence for Knowledge-Based Systems Design*, volume 6178 of *Lecture Notes in Computer Science*, pages 260–269. Springer Berlin / Heidelberg, 2010.

[63] Y. Djouadi and H. Prade. Interval-valued fuzzy formal concept analysis. In *ISMIS '09 : Proceedings of the 18th International Symposium on Foundations of Intelligent Systems*, pages 592–601, Berlin, Heidelberg, 2009. Springer-Verlag.

[64] W.M. Dong and F.S. Wong. Fuzzy weighted averages and implementation of the extension principle. *Fuzzy Sets Syst.*, 21(2) :183–199, 1987.

[65] W. Dongrui and J. Mendel. Enhanced karnik–mendel algorithms. *Fuzzy Systems, IEEE Transactions on*, 17(4) :923 –934, 2009.

[66] C. Dou, W. Woldt, I. Bogardi, and M. Dahab. Steady state groundwater flow simulation with imprecise parameters. *Water Resources Res*, 31(11) :2709–2719, 1995.

[67] D. Dubois, F. Dupin de Saint-Cyr, and H. Prade. A possibility-theoretic view of formal concept analysis. *Fundam. Inform.*, 75(1-4) :195–213, 2007.

[68] D. Dubois, H. Fargier, and J. Fortin. A generalized vertex method for computing with fuzzy intervals. In *Fuzzy Systems, 2004. Proceedings. 2004 IEEE International Conference on*, volume 1, pages 541–546, 2004.

[69] D. Dubois, H. Fargier, and H. Prade. Multiple source information fusion : a practical inconsistency tolerant approach. In *IPMU*, pages 1047–1054, 2000.

[70] D. Dubois, J. Lang, and H. Prade. Dealing with multi-source information in possibilistic logic. In *ECAI*, pages 38–42, 1992.

[71] D. Dubois and H. Prade. Operations on fuzzy numbers. *International Journal of Systems Science*, 9(6) :613–626, 1978.

[72] D. Dubois and H. Prade. *Théorie des possibilités*. Masson, Paris, 1985.

[73] D. Dubois and H. Prade. *Théorie des possibilités- Applications à la représentation des connaissances en informatique*. Masson,Paris, 1987.

[74] D. Dubois and H. Prade. *Possibility theory : an approach to computerized processing of uncertainty*. Plenum Press, 1988.

[75] D. Dubois and H. Prade. Possibility theory and data fusion in poorly informed environments. *Control Eng. Practice*, 2 :811–823, 1994.

[76] D. Dubois and H. Prade. Introduction : Revising, updating and combining knowledge. In *Belief Change*, volume 3 of *Handbook of defeasible Reasoning and Uncertainty management Systems*, pages 1–15, 1998.

[77] D. Dubois and H. Prade. Possibility theory : Qualitative and quantitative aspects. In *Handbook on Defeasible Reasoning and Uncertainty Management Systems,*, pages 21–42. Kluwer Academic Publishers, 1998.

[78] D. Dubois and H. Prade. *Fundamentals of fuzzy sets*. The Handbooks of Fuzzy Sets, Etats-Unis, 2000.

[79] D. Dubois and H. Prade. Possibility theory in information fusion. In *Proc. FUSION 2000*, volume 1, pages PS6–P19, 2000.

[80] D. Dubois and H. Prade. La problématique scientifique du traitement de l'information. *Revue I3 : Information - Interaction - Intelligence*, 1 :9–34, 2001.

[81] D. Dubois and H. Prade. Possibility theory in information fusion. In *Data fusion and Perception*, pages 53–76, 2001.

[82] D. Dubois and H. Prade. Représentations formelles de l'incertain et de l'imprécision. In *Concepts et méthodes pour l'aide à la décision - Volume 1*, pages 111–165. Hermès - Lavoisier, 2006.

[83] D. Dubois and H. Prade. A set-theoretic view of belief functions. In Roland Yager and Liping Liu, editors, *Classic Works of the Dempster-Shafer Theory of Belief Functions*, volume 219 of *Studies in Fuzziness and Soft Computing*, pages 375–410. Springer Berlin / Heidelberg, 2008.

[84] D. Dubois and H. Prade. Possibility theory and formal concept analysis in information systems. In *Proceedings of the Joint 2009 International Fuzzy Systems Association World Congress and 2009 European Society of Fuzzy Logic and Technology Conference, Lisbon, Portugal, July 20-24, 2009*, pages 1021–1026, 2009.

[85] P. Dwyer. *Linear computations.* Wiley, New York, 1951.

[86] D. Dubois et H. Prade. Possibility theory : An approach to computerized processing of uncertainty. *SIAM Review*, 34(1) :147–148, 1992.

[87] S. Ferré. *Systèmes d'information logiques : un paradigme logico-contextuel pour interroger, naviguer et apprendre.* Thèse d'université, Université de Rennes 1, Octobre 2002.

[88] B. De Finetti. La prévision : ses lois logiques, ses sources subjectives. Technical document, Institut Poincarré, 1937.

[89] J. Fortin. *Analyse d'intervalles flous, application à l'ordonnancement dans l'incertain.* PhD thesis, Université Paul Sabatier, 2006.

[90] J. Fortin, D. Dubois, and H. Fargier. Gradual numbers and their application to fuzzy interval analysis. *IEEE T. Fuzzy Systems*, 16(2) :388–402, 2008.

[91] J. Fortin, D. Dubois, and H. Fargier. Gradual numbers and their application to fuzzy interval analysis. *IEEE T. Fuzzy Systems*, 16(2) :388–402, 2008.

[92] C. Freissinet, M. Erlich, and M. Vauclin. A fuzzy logic-based approach to assess imprecisions of soil water contamination modelling. *Soil and Tillage Research*, 47(1-2) :11 – 17, 1998.

[93] B. Ganter. Two basic algorithms in concept analysis. FB4-Preprint 831, Technische Hochschule Darmstadt, 1984.

[94] B. Ganter and S. O. Kuznetsov. Pattern Structures and Their Projections. In *Int. Conf. on Conceptual Structures*, pages 129–142, 2001.

[95] B. Ganter and K. Reuter. Finding all closed sets : A general approach. *Order*, 8(3) :283–290, September 1991.

[96] B. Ganter and R. Wille. *Formal Concept Analysis.* Springer, mathematical foundations edition, 1999.

[97] J. Gebhardt and R. Kruse. Parallel combination of information sources. In *Belief Change*, volume 3 of *Handbook of defeasible Reasoning and Uncertainty managemen Systems*, pages 393–439, 1998.

[98] G. Georgescu and A. Popescu. Non-dual fuzzy connections. *Archive for Mathematical Logic*, 43 :1009–1039(31), November 2004.

[99] P. Girardin, C. Bockstaller, and H. Van Der Werf. Use of agro-ecological indicators for the evaluation of farming systems. *European Journal of Agronomy*, 7 :261–271, 1997.

[100] P. Girardin, C. Bockstaller, and H. Van Der Werf. Indicators : tools to evaluate the environmental impacts of farming systems. *Journal of Sustainable Agriculture.*, 13 :5–21, 1999a.

[101] P. Girardin, C. Bockstaller, and H. Van Der Werf. A method to assess the environmental impact of farming systems by means of agri-ecological indicators. *environmental indices : systems analysis approach*, 1 :297–312, 1999b.

[102] P. Girardin, C. Bockstaller, and H. Van Der Werf. Assessment of potential impact of agricultural practices on the environment the agro*eco method. *Environmental Impact Assessment*, 20 :227–239, 2000.

[103] R. Godin and P. Valtchev. Formal concept analysis-based class hierarchy design in object-oriented software development. In *Formal Concept Analysis*, pages 304–323, 2005.

[104] R. Gras, M. Benoit, J.P Deffontaines, M. Duru, M. Lafarge, A. Langlet, and P.M. Osty. *Le fait technique en agronomie. Activité agricole. Concepts et méthodes d'étude.* L'hamarttan/Ed. Institut National de la Recherche Agronomique, 1998.

[105] Y.Y. Guh, C.C. Hong, K.M. Wang, and E.S. Lee. Fuzzy weighted average : a max-min paired elimination method. *Comput. Math. Appl.*, 32 :115–123, 1996.

[106] S.M. Guu. Fuzzy weighted averages revisited. *Fuzzy Sets and Systems*, 126(3) :411 – 414, 2002.

[107] D. Guyonnet, B. Bourgine, D. Dubois, H. Fargier, B. Côme, and J. Chilès. Hybrid approach for addressing uncertainty in risk assessments. *Journal of Environmental Engineering*, 129(1) :68–78, 2003.

[108] P. Hansen, M. V. Poggi de Aragao, and C.C. Ribeiro. Hyperbolic 0-1 programming and query optimization in information retrieval. *Math. Program.*, 52(2) :255–263, 1991.

[109] M. Hanss. *Introduction to fuzzy arithmetic. Theory and application.* Springer, Berlin, 2004.

[110] B. Hayes. A lucid interval. *American Scientist*, 2003.

[111] A. Jaoua and S. Elloumi. Galois connection, formal concepts and galois lattice in real relations : application in a real classifier. *Journal of Systems and Software*, 60(2) :149–163, 2002.

[112] H. Jeffreys. *Theory of probability.* Oxford University Press, 1961.

[113] C. Kao and S.T Liu. Fractional programming approach to fuzzy weighted average. *Fuzzy Sets and Systems*, 120 :435–444, 2001.

[114] N. Karnik and J. Mendel. Centroid of a type-2 fuzzy set. *Inf. Sci.*, 132(1-4) :195–220, 2001.

[115] A. Kaufmann and M. Gupta. *Applied Fuzzy Arithmetic.* Van nostrand Reinhold, New York, 1985.

[116] M. Kaytoue, Z. Assaghir, N. Messai, and A. Napoli. Classification de données numériques par treillis de concepts basée sur une similarité. In *Société Francophone de Classification (SFC)*, 2010.

[117] M. Kaytoue, Z. Assaghir, N. Messai, and A. Napoli. Complémentarité de deux méthodes de classification pour la construction de treillis de concepts à partir de données numériques. In *17ème conférence en Reconnaissance des Formes et Intelligence Artificielle*, 2010.

[118] M. Kaytoue, Z. Assaghir, N. Messai, and A. Napoli. Two complementary classification methods for designing a concept lattice from interval data. In *Foundations of Information and Knowledge Systems*, volume 5956 of *Lecture Notes in Artificial Intelligence*, pages 345–362. Springer-Verlag Berlin Heidelberg, 2010.

[119] M. Kaytoue, Z. Assaghir, A. Napoli, and S. O. Kuznetsov. Embedding tolerance relations in formal concept analysis for classifying numerical data. In *19th Conference on Information and Knowledge Management (CIKM).* ACM, 2010.

[120] M. Kaytoue, S. Duplessis, S. O. Kuznetsov, and A. Napoli. Two FCA-Based Methods for Mining Gene Expression Data. In *Formal Concept Analysis*, LNCS 5548, 2009.

[121] M. Kaytoue, S. Duplessis, S. O. Kuznetsov, and A. Napoli. Two fca-based methods for mining gene expression data. In S. Ferré and S. Rudolph, editors, *International Conference on Formal Concept Analysis*, volume 5548 of *Lecture Notes in Computer Science*, pages 251–266. Springer, 2009.

[122] M. Kaytoue, S. O. Kuznetsov, Z. Assaghir, and A. Napoli. Embedding Tolerance Relations in Concept Lattices - An application in Information Fusion. Research Report RR-7353, INRIA, 2010.

[123] M. Kaytoue, S. O. Kuznetsov, A. Napoli, and S. Duplessis. Mining gene expression data with pattern structures in formal concept analysis. *Information Sciences*, In Press, 2010.

[124] S. Kuznetsov. A fast algorithm for computing all intersections of objects in a finite semi-lattice. *Automatic Documentation and Mathematical Linguistics*, 27(5) :11–21, 1993.

[125] S. Kuznetsov. Galois connections in data analysis : Contributions from the soviet era and modern russian research. In *Formal Concept Analysis*, pages 196–225, 2005.

[126] S. O. Kuznetsov. Pattern structures for analyzing complex data. In H. Sakai, M. K. Chakraborty, A. E. Hassanien, D. Slezak, and W. Zhu, editors, *Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, 12th International Conference, RSFDGrC 2009, Delhi, India, Dec.15-18, 2009. Proceedings*, volume 5908 of *LNCS*, pages 33–44. Springer, 2009.

[127] S.O. Kuznetsov and S.A. Obiedkov. Comparing Performance of Algorithms for Generating Concept Lattices. *Journal of Experimental and Theoretical Artificial Intelligence*, 14 :189–216, 2002.

[128] L. Lakhal and G. Stumme. Efficient mining of association rules based on formal concept analysis. In *Formal Concept Analysis*, pages 180–195, 2005.

[129] D.H. Lee and D. Park. An efficient algorithm for fuzzy weighted average. *Fuzzy Sets and Systems*, 87(1) :39 – 45, 1997.

[130] R.A. Leonard. movement of pesticides into surface waters. In *Pesticides in the soil environment : processes, impacts, and modeling*, volume 2, pages 303–349, 1990.

[131] M. Lerond, C. Larrue, P. Michel, B. Roudier, and C. Sanson. *L'évaluation environnementale des politiques, plans et programmes. Objectifs, méthodologie et cas pratiques.* Tec-Doc/Ed., 2003.

[132] T.S. Liou and M.J. Wang. Fuzzy weighted average : An improved algorithm. *Fuzzy Sets and Systems*, 49 :307–315, 1992.

[133] F. Liu and J. Mendel. Aggregation using the fuzzy weighted average as computed by the karnik-mendel algorithms. *IEEE T. Fuzzy Systems*, 16(1) :1–12, 2008.

[134] N. Maille. *Modèle logico-algébrique pour la fusion symbolique et l'analyse formelle.* PhD thesis, Ecole Nationale Supérieure de l'Aéronautique et de l'espace, 1999.

[135] R. Malouf. Maximal consistent subsets. *Comput. Linguist.*, 33(2) :153–160, 2007.

[136] J. Medina, M. Ojeda-Aciego, and J. Ruiz-Calviño. Formal concept analysis via multi-adjoint concept lattices. *Fuzzy Sets and Systems*, 160(2) :130 – 144, 2009.

[137] N. Messai. *Analyse de concepts formels guidée par des connaissances de domaine : Application à la découverte de ressources génomiques sur le Web.* PhD thesis, Université Henri Poincaré- Nancy 1, 2009.

[138] N. Messai, M. Devignes, A. Napoli, and M. Smaïl-Tabbone. Querying a bioinformatic data sources registry with concept lattices. In *ICCS*, pages 323–336, 2005.

[139] N. Messai, M. Devignes, A. Napoli, and M. Smaïl-Tabbone. Treillis de concepts et ontologies pour interroger l'annuaire de sources de données biologiques bioregistry. *Ingénierie des Systèmes d'Information*, 11(1) :39–60, 2006.

[140] N. Messai, M. Devignes, A. Napoli, and M. Smail-Tabbone. Many-valued concept lattices for conceptual clustering and information retrieval. In *Proceeding of the 2008 conference on ECAI 2008*, pages 127–131, Amsterdam, The Netherlands, The Netherlands, 2008. IOS Press.

[141] B. Meunier. *Logique floue, principes, aide à la décision.* Hermès, 2003.

[142] G. Mitchell, A. May, and A. McDonald. Picabue : a methodological framework for the development of indicators of sustainable development. *Int. J. sustain. dev. world ecol.*, 2 :104–123, 1995.

[143] I.S. Molchanov. *Theory of random sets.* Springer,London, 2005.

[144] R. E. Moore. *Interval Analysis.* Prentice-Hall, Englewood Cliffs, NJ, New York, 1963.

[145] R. E. Moore and C. T. Yang. Interval analysis I. Technical document, Lockheed Missiles and Space Division, Sunnyvale, CA, USA, 1959.

[146] R.E. Moore. *Interval Analysis.* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1966.

[147] S. Moral and J. Sagrado. Aggregation of imprecise probabilities, 1997.

[148] S. Nahmias. Fuzzy variables. *Fuzzy Sets and Systems*, 1(2) :97 – 110, 1978.

[149] A. Neumaier. *Interval Methods for Systems of Equations.* Cambridge University Press, Cambridge, 1990.

[150] A. Neumaier. Clouds, fuzzy sets, and probability intervals. *Reliable Computing*, 10 :249–272, 2004.

[151] Arnold Neumaier. The wrapping effect, ellipsoid arithmetic, stability and confidence region. *Computing Supplementum*, 9 :175–190, 1993.

[152] E. Mephu Nguifo and P. Njiwoua. *Using Lattice-based Feamework as a Tool for Feature Extraction*, chapter 13, pages 205–216. Kluwer Academic Publishers, 1998.

[153] E.M. Norris. An algorithm for computing the maximal rectangles in a binary relation. *Revue Roumaine de Mathématiques Pures et Appliquées*, 23(2) :243–250, 1978.

[154] M. Oussalah. Study of some algebrical properties of adaptive combination rules. *Fuzzy Sets and Systems*, 114 :391–409, 2000.

[155] M. Oussalah. On the use of hamacher's t-norms family for information aggregation. *Information Sciences*, 153 :107 – 154, 2003.

[156] M. Oussalah, H. Maaref, C. Barret, and H. Prade. From adaptive to progressive combination of possibility distributions. *Fuzzy Sets and Systems*, 139(3) :559–582, 2003.

[157] Z. Pawlak. *Rough Sets : Theoretical Aspects of Reasoning about Data.* Kluwer Academic, Dordrecht, 1991.

[158] F. Pervanchon and A. Blouet. De la durabilité de l'agriculture raisonnée. *Natures Sciences Sociétés*, 10 :36–39, 2002.

[159] T. Pham. *Fusion de l'information géographique hiérarchisée.* PhD thesis, Université de Provence Aix-Marseille, 2005.

[160] V. Phan-Luong. A framework for integrating information sources under lattice structure. *Information Fusion*, 9(2) :278 – 292, 2008.

[161] G. Polaillon. *Organisation et interprétation par les treillis de Glois de données de type multivalué, intervalle ou histogramme.* Thèse de doctorat en informatique, Université Paris IX-Dauphine, Décembre 1998.

[162] U. Priss. Linguistic applications of formal concept analysis. In *Formal Concept Analysis*, pages 149–160, 2005.

[163] F.P. Ramsey. Truth and probability. In R. B. Braithwaite, editor, *The Foundations of Mathematics and other Logical Essays*, Histoy of Economic Thought Chapters, chapter 7, pages 156–198. McMaster University Archive for the History of Economic Thought, 1926.

[164] N. Rescher and R. Manor. On inference from inconsistent premisses. *Theory and Decision*, 1 :179–217, 1970.

[165] J. Reus, P. Leendertse, C. Bockstaller, I. Fomsgaard, V. Gutsche, K. Lewis, C. Nilsson, L. Pussemier, M. Trevisan, H. van der Werf, F. Alfarroba, S. Blümel, J. Isart, D. McGrath, and T. Seppälä. Comparison and evaluation of eight pesticide environmental risk indicators developed in europe and recommendations for future use. *Agriculture, Ecosystems and Environment*, 90(2) :177–187, 2002.

[166] S.A. Sandri, D. Dubois, and H.W. Kalfsbeek. Elicitation, assessment and pooling of expert judgements using possibility theory, 1990.

[167] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, New Jersey, 1976.

[168] G. Shafer. The combination of evidence. *International Journal of Intelligent Systems*, 1(3) :155–179, 1986.

[169] P. Smets. Imperfect information : Imprecision and uncertainty. In *Uncertainty Management in Information Systems*, pages 225–254. Kluwer Academic Publishers, Boston, 1996.

[170] M. Thiollet-Scholtus. *Construction d'un indicateur de qualité des eaux de surface vis-à-vis des produits phytosanitaires à l'échelle du bassin versant viticole*. PhD thesis, Institut National Polytechnique de Lorraine, 2004.

[171] T. Tilley, R. Cole, P. Becker, and P. W. Eklund. A survey of formal concept analysis support for software engineering activities. In *Formal Concept Analysis*, pages 250–271, 2005.

[172] M. Troffaes. Generalizing the conjunction rule for aggregating conflicting expert opinions. *International Journal of Intelligent Systems*, 21(3) :361 – 380, 2006.

[173] P. Walley. The elicitation and aggregation of beliefs. Technical document, University of Warwick, 1982.

[174] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, New York, 1991.

[175] M. Ward and R. Dilworth. Residuated lattices. *Transactions of the American Mathematical Society*, 45(3) :335–354, 1939.

[176] H. Van Der Werf and C. Zimmer. An indicator of pesticide environemental impact based on fuzzy expert system. *Chemosphere.*, 36 :2225–2249, 1998.

[177] R. Wille. Restructuring lattice theory : an approach based on hierarchies of concepts. *Ordered sets*, pages 445–470, 1982.

[178] J. Wohlfahrt. *Développement d'un indicateur d'exposition des eaux de surface aux pertes de pesticides à l'échelle du bassin versant*. PhD thesis, Institut National Polytechnique de Lorraine, 2009.

[179] K. Wolff. Concepts in fuzzy scaling theory : order and granularity. *Fuzzy Sets and Systems*, 132(1) :63 – 75, 2002.

[180] R. Yager. On ordered weighted averaging aggregation operators in multicriteria decision-making. *IEEE Trans. Syst. Man Cybern.*, 18(1) :183–190, 1988.

[181] R. Yager. New modes of owa information fusion. *International Journal of Intelligent Systems*, 13(7) :661–681, 1998.

[182] R. Yager. Induced aggregation operators. *Fuzzy Sets Syst.*, 137(1) :59–69, 2003.

[183] S. Yahia and A. Jaoua. Discovering knowledge from fuzzy concept lattice. *Data mining and computational intelligence*, pages 167–190, 2001.

[184] H. Yang, H. Yao, and D. Jones. Calculating functions of fuzzy numbers. *Fuzzy Sets Syst.*, 55(3) :273–283, 1993.

[185] R. Young. The algebra of many-valued quantities. In *Mathematische Annalen*, pages 260–290. Springer Berlin / Heidelberg, 1931.

[186] C.C. Hong Y.Y. Guh and E.S. Lee. Fuzzy weighted average : The linear programming approach via charnes and cooper's rule. *Fuzzy Sets and Systems*, 117 :157–160, 2001.

[187] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning - ii. *Inf. Sci.*, 8(4) :301–357, 1975.

[188] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8(3) :338 – 353, 1965.

[189] L.A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Syst.*, 100 :9–34, 1999.

INSTITUT NATIONAL
POLYTECHNIQUE
DE LORRAINE

AUTORISATION DE SOUTENANCE DE THESE
DU DOCTORAT DE L'INSTITUT NATIONAL
POLYTECHNIQUE DE LORRAINE

o0o

VU LES RAPPORTS ETABLIS PAR :
Monsieur Didier DUBOIS, Directeur de Recherche, Université Paul Sabatier, IRIT, Toulouse
Monsieur David MAKOWSKI, Chargé de Recherche, AgroParisTech, Thiverval-Grignon

Le Président de l'Institut National Polytechnique de Lorraine, autorise :

Madame ASSAGHIR Zainab

à soutenir devant un jury de l'INSTITUT NATIONAL POLYTECHNIQUE DE LORRAINE, une thèse intitulée :

"Analyse formelle de concepts et fusion d'informations : application à l'estimation et au contrôle d'incertitude des indicateurs agri-environnementaux"

en vue de l'obtention du titre de :

DOCTEUR DE L'INSTITUT NATIONAL POLYTECHNIQUE DE LORRAINE

Spécialité : **« Sciences Agronomiques »**

Fait à Vandoeuvre, le 02 novembre 2010
Le Président de l'I.N.P.L.,
F. LAURENT

INSTITUT NATIONAL POLYTECHNIQUE DE LORRAINE - NANCY -

NANCY BRABOIS
2, AVENUE DE LA
FORET-DE-HAYE
BOITE POSTALE 3
F - 54501
VANDŒUVRE CEDEX

TEL. 33/03.83.59.59.59
FAX. 33/03.83.59.59.55

Titre : ANALYSE FORMELLE DE CONCEPTS ET FUSION D'INFORMATIONS : APPLICATION A L'ESTIMATION ET AU CONTROLE D'INCERTITUDE DES INDICATEURS AGRI-ENVIRONNEMENTAUX

Résumé : La fusion d'informations consiste à résumer plusieurs informations provenant des différentes sources en une information exploitable et utile pour l'utilisateur.Le problème de la fusion est délicat surtout quand les informations délivrées sont incohérentes et hétérogènes. Les résultats de la fusion ne sont pas souvent exploitable et utilisables pour prendre une décision, quand ils sont imprécis. C'est généralement due au fait que les informations sont incohérentes. Plusieurs méthodes de fusion sont proposées pour combiner les informations imparfaites et elles appliquent l'opérateur de fusion sur l'ensemble de toutes les sources et considèrent le résultat tel qu'il est. Dans ce travail, nous proposons une méthode de fusion fondée sur l'Analyse Formelle de Concepts, en particulier son extension pour les données numériques : les structures de patrons. Cette méthode permet d'associer chaque sous-ensemble de sources avec son résultat de fusion. Toutefois l'opérateur de fusion est choisi, alors un treillis de concept est construit. Ce treillis fournit une classification intéressante des sources et leurs résultats de fusion. De plus, le treillis garde l'origine de l'information. Quand le résultat global de la fusion est imprécis, la méthode permet à l'utilisateur d'identifier les sous-ensemble maximaux de sources qui supportent une bonne décision. La méthode fournit une vue structurée de la fusion globale appliquée à l'ensemble de toutes les sources et des résultats partiels de la fusion marqués d'un sous-ensemble de sources. Dans ce travail, nous avons considéré les informations numériques représentées dans le cadre de la théorie des possibilités et nous avons utilisé trois sortes d'opérateurs pour construire le treillis de concepts. Une application dans le monde agricole, où la question de l'expert est d'estimer des valeurs des caractéristiques de pesticide provenant de plusieurs sources, pour calculer des indices environnementaux est détaillée pour évaluer la méthode de fusion proposée.

Mots clés : Imprécision, Théorie des Possibilités, Fusion d'informations, Analyse Formelle de concepts, Structure de patrons, Indicateur

Title : FORMAL CONCEPT ANALYSIS AND INFORMATION FUSION : APPLICATION ON THE UNCERTAINTY ESTIMATION OF ENVIRONMENTAL INDICATOR

Abstract : Merging pieces of information into an interpretable and useful format is a tricky task even when an information fusion method is chosen. Fusion results may not be in suitable form for being used in decision analysis. This is generally due to the fact that information sources are heterogeneous and provide inconsistent information, which may lead to imprecise results. Several fusion operators have been proposed for combining uncertain information and they apply the fusion operator on the set of all sources and provide the resulting information. In this work, we studied and proposed a method to combine information using Formal Concept Analysis in particular Pattern Structures. This method allows us to associate any subset of sources with its information fusion result. Then once a fusion operator is chosen, a concept lattice is built. The concept lattice gives an interesting classification of fusion results and it keeps a track of the information origin. When the fusion global result is too imprecise, the method enables the users to identify what maximal subset of sources would support a more precise and useful result. Instead of providing a unique fusion result, the method yields a structured view of partial results labeled by subsets of sources. In this thesis, we studied the numerical information represented in the framework of possibility theory and we used three fusion operators to built the concept lattice. We applied this method in the context of agronomy when experts have to estimate several characteristics values coming from several sources for computing an environmental risk.

Keywords : Imprecision, Possibility Theory, Information Fusion, Formal Concept Analysis, Pattern Structure, Indicator.