Year : 2017

Intégration des méthodes computationnelles en renseignement criminel. Application sur la détection de problèmes à travers les tendances dans les activités criminelles

Grossrieder Lionel

Grossrieder Lionel, 2017, Intégration des méthodes computationnelles en renseignement criminel. Application sur la détection de problèmes à travers les tendances dans les activités criminelles

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive http://serval.unil.ch
Document URN : urn:nbn:ch:serval-BIB_30451BB9537B7

UNIL | Université de Lausanne
Faculté de droit, des sciences criminelle et d'administration publique
École des sciences criminelles

Intégration des méthodes computationnelles en renseignement criminel

Application sur la détection de problèmes à travers les tendances dans les activités criminelles

THÈSE DE DOCTORAT

présentée à la
Faculté de droit, des sciences criminelles et d'administration publique
de l'Université de Lausanne

pour l'obtention du grade de
Docteur ès Sciences en science forensique
par

Lionel Grossrieder

Directeur de thèse :

Prof. Olivier Ribaux

LAUSANNE

2017

UNIL | Université de Lausanne
Ecole des sciences criminelles
bâtiment Batochime
CH-1015 Lausanne

IMPRIMATUR

A l'issue de la soutenance de thèse, le Jury autorise l'impression de la thèse de M. Lionel Grossrieder, candidat au doctorat en science forensique, intitulée

« Intégration des méthodes computationnelles en renseignement criminel : application sur la détection de problèmes à travers les tendances dans les activités criminelles »

Le Président du Jury

Professeur Pierre Esseiva

Lausanne, le 5 mai 2017

Lionel Grossrieder

Intégration des méthodes computationnelles en renseignement criminel
Application sur la détection de problèmes à travers les tendances dans les activités criminelles

Président du jury
Prof. Pierre Esseiva

Membres du jury
Prof. Kilian Stoffel
Dr. Bertrand Schnetz
Prof. Quentin Rossy
Prof. Stefano Caneppele

Directeur de thèse
Prof. Olivier Ribaux

Université de Lausanne
École des sciences criminelles
Faculté de droit, des sciences criminelles et d'administration publique
1015 Lausanne

« *Toutes nos erreurs sont des jugements téméraires, et toutes nos vérités, sans exception, sont des erreurs redressées* » Alain (1868-1951)

REMERCIEMENTS

Ce travail de thèse a été réalisé à l'École des sciences criminelles (ESC) de l'Université de Lausanne sous la direction du Professeur Olivier Ribaux, directeur de l'ESC. Le jury était composé du Professeur Pierre Esseiva, vice-directeur de l'ESC, président du jury, du Professeur Kilian Stoffel, recteur de l'Université de Neuchâtel, du Docteur Bertrand Schnetz, chef de la Police judiciaire de la Police cantonale jurassienne, du Professeur Quentin Rossy, professeur à l'ESC et du Professeur Stefano Caneppele, professeur à l'ESC. Je remercie l'ensemble du jury pour l'intérêt porté à mes travaux, leurs précieux conseils et le temps qu'ils m'ont aimablement accordé.

Cette thèse n'aurait jamais vu le jour sans l'implication et le soutien tant intellectuel qu'émotionnel des nombreuses personnes qui l'ont nourri, chacun et chacune à leur manière. En arpentant cette aventure doctorale, j'ai eu l'occasion de faire des rencontres tant insolites qu'inattendues et, bien plus que l'aboutissement de ces nombreuses années d'efforts, c'est bel et bien ces rencontres qui constituent à mes yeux le succès le plus précieux de cette thèse. Je rends ainsi hommage à toutes ces personnes et tiens à remercier tout particulièrement :

Le Professeur Olivier Ribaux pour la confiance qu'il m'a accordée, sa passion communicative de la recherche et son expérience qu'il a partagée sans modération avec moi. Je le remercie infiniment pour tout ce qu'il m'a appris durant toutes ces années.

L'équipe de l'Institut du Management de l'Information de l'Université de Neuchâtel, le Professeur Kilian Stoffel, le Docteur Fabrizio Albertetti et le Docteur Paul Cotofrei qui, à travers une solide collaboration, m'ont permis de véritablement saisir les enjeux et l'importance de valoriser l'interdisciplinarité dans la recherche.

La Division Coordination et Renseignement Judiciaire de la Police cantonale vaudoise pour m'avoir accueilli en leurs locaux et participé activement à ma recherche. Un grand merci notamment à Messieurs Sylvain Ioset et Damien Dessimoz qui m'ont permis de découvrir l'action de la Police cantonale vaudoise en matière de renseignement criminel opérationnel.

Les Polices cantonales de Fribourg, Genève, Jura, Neuchâtel, Valais et Vaud qui, à travers le CICOP, ont accepté de collaborer à ce projet doctoral en offrant l'accès à la base de données PICAR.

Le Fonds National Suisse pour la Recherche Scientifique pour le soutien financier qui a permis de réaliser ce travail dans les meilleures conditions possible (N° 135236 et 156287), ainsi que la Fondation pour l'Université de Lausanne dont le soutien a permis d'en disséminer les résultats au sein de conférences internationales.

Tous les amis et collègues de l'ESC qui ont rendu la traversée de ces années de thèse aussi agréable et stimulante que possible, notamment :

- mes fidèles compagnons d'(in)fortune qui ont ramé avec moi dans les tumultes de la thèse. Stéphanie Loup pour avoir été là tant dans les moments de joie que de doute. Ton amitié et tes conseils furent à n'en pas douter un phare infaillible dans cet océan doctoral, merci infiniment. Julien Chopin pour ta bonne humeur et ton aide inestimable. Une belle amitié initiée au master, confirmée au doctorat et, je l'espère, enrichie à l'avenir. Sonja Bitzer pour toutes les réflexions stimulantes, les rires et les voyages récurrents en Australie. Natalia Delgrande pour ton soutien, ta franchise et nos moments partagés. Et Aurélie Stoll qui, malgré une brève interruption, m'a rejoint sur le chemin ardu de la thèse ;
- les Professeurs Manon Jendly et Quentin Rossy pour m'avoir permis d'élargir mes horizons de chercheur, pour votre enthousiasme, vos encouragements sans faille et les discussions passionnantes et passionnées que j'espère encore partager longtemps avec vous ;
- la Biocafet pour l'ambiance incroyable qui a bercé les bureaux au 5ème. Elénore, Maëlig, Adrien, Thomas, Lorène, Ludo, Marion, Élodie, Oriana, que ce soit autour d'un café, d'un apéro ou d'une bataille de nerfs, vous avez rythmé cette thèse au son de la convivialité et de l'humour. Mille mercis. J'en profite également pour remercier solennellement le petit truc bleu ;
- le groupe analyse criminelle pour toutes les discussions stimulantes qui ont alimenté ce travail. Un grand merci entre autres à Simon, Julien, David, Mélanie, Thibault, Amélie, Denise, Samuel et Killian ;
- les étudiants, et en particulier Margarida, Juliette et Myriam qui, grâce à leurs travaux, ont posé leur indispensable pierre dans la réalisation de cette thèse ;
- et tous les autres collègues, d'ici et d'ailleurs, qui ont un jour croisé mon chemin au détour d'un café, d'un couloir, d'une réunion ou d'une bière. J'aimerai pouvoir tous vous citer, mais faute d'y arriver, j'ai une pensée pour chacun d'entre vous.

Mes amis du bout du lac, Christian, David D., Tino, Ben, Sandra, David T. et Virginie, pour tous les moments de détentes et de rigolades partagés ensemble.

Ma famille pour leur soutien inconditionnel et mes parents qui ont toujours cru en moi et qui m'ont donné les moyens de concrétiser mes projets en toutes circonstances.

À toutes et à tous,

Merci.

RÉSUMÉ

Cette thèse a été réalisée en parallèle avec l'émergence de la police prédictive (*predictive policing)*. Ce mouvement naturellement issu de l'ère du *big data* a stimulé et questionné l'utilisation de plus en plus intensive des modèles computationnels et des technologies en analyse et renseignement criminel. Les différents acteurs de la sécurité, dont notamment les services de police, se retrouvent désormais confrontés à des quantités croissantes de données de criminalité, quand elles ne sont tout simplement pas de nature nouvelle (p. ex. les traces numériques). Un des défis de l'analyse criminelle est de faire face à ces nouvelles et grandes quantités de données dans le but de les détecter, les collecter, les traiter, les analyser et les exploiter en informations utiles tant à l'investigation qu'au renseignement criminel.

L'élaboration d'une méthodologie réaliste, correctement formalisée et transparente s'impose alors comme un défi prioritaire. Cette problématique soulève des enjeux liés aux libertés individuelles, mais également au besoin pour la prise de décisions en matière d'action de sécurité, d'être fondés sur des données probantes.

Afin de répondre à ces enjeux, nous nous intéressons à la question suivante : comment intégrer de manière pragmatique les méthodes computationnelles dans les processus d'analyse criminelle préexistants en considérant un cadre de travail interdisciplinaire puisant à la fois dans la criminologie et la science forensique et orienté sur la résolution de problème ?

Cette question étant évidemment vaste, cette thèse se restreint à proposer des pistes pour une telle méthodologie en considérant 2 objectifs principaux :

- L'expression d'une approche méthodologique interdisciplinaire en renseignement criminel en se fondant sur le développement d'une unité d'analyse criminelle particulière qui a itérativement intégré et harmonisé ses méthodes et outils ;

- L'amélioration du système opérationnel existant via l'intégration d'un composant computationnel dans la détection de tendances des activités criminelles au sein des processus de renseignement de l'unité d'analyse considérée.

Considérant ces délimitations, l'hypothèse principale de ce travail est la suivante :

- Il est possible de détecter des changements dans des patterns d'activités (nouveaux patterns, évolution de patterns déjà connus, disparition d'un pattern connu) dans la distribution spatio-temporelle des données de la criminalité en détectant les patterns de rupture par des méthodes computationnelles appliquées systématiquement dans l'environnement particulier considéré (unité d'analyse).

Cette proposition est complétée par 3 hypothèses spécifiques, à savoir :

- Les patterns dans les données reflètent les patterns dans les activités criminelles.

- Une détection automatique est susceptible de rendre plus complète, plus rapide et plus précise, la détection de problèmes par les analystes de l'unité d'analyse considérée

- Une classification situationnelle des événements alimentée par la criminologie environnementale est appropriée pour encadrer et guider cette détection.

En prenant la méthodologie d'une unité de renseignement criminel en Suisse romande comme fondement, une approche intégrative et itérative dans l'application des méthodes computationnelles en renseignement criminel est exprimée. L'approche proposée est basée sur un postulat fondamental en analyse criminelle : les activités litigieuses suivent des patterns susceptibles d'être détectés et analysés à l'aide des données disponibles. Le raisonnement est basé sur la plus élémentaire de ces données : la trace, vestige physique (et numérique) de l'activité litigieuse, qui a été reconnue et collectée sur les scènes de crime. L'approche développée est ensuite appliquée sur deux objectifs majeurs de l'unité d'analyse considérée : le traitement des données à l'aide de classification automatique et la détection de problèmes à l'aide des tendances dans les données de la criminalité. Cette application vise à intégrer quelques méthodes computationnelles simples de classification et de détection de ruptures dans les processus opérationnels.

Les résultats obtenus corroborent l'hypothèse principale de ce travail qui soutient la possibilité de détecter des problèmes dans la distribution spatio-temporelle des données de la criminalité. Cependant, l'hypothèse d'une détection automatique plus complète, rapide et précise que la détection humaine n'est que partiellement corroborée. En effet, il apparaît que les capacités de l'algorithme varient selon le type de criminalité étudié. L'hypothèse spécifique postulant que ces patterns de données

reflètent les patterns d'activités criminelles se trouve corroborée par les différentes illustrations empiriques proposées. L'arrivée de nouveaux groupes d'auteurs, les changements d'environnement avec le passage à l'heure d'hiver ou encore l'activité d'un auteur sériel sont autant d'exemples imprimant des patterns particuliers dans les données collectées. Concernant la dernière hypothèse spécifique, une classification situationnelle des événements semble appropriée pour effectuer le processus de détection et l'hypothèse est corroborée.

Nous avons observé que les méthodes computationnelles semblent être adéquates pour soutenir l'analyse criminelle et ses dispositifs de veille, et ce, tout particulièrement concernant la détection de patterns dans les tendances des activités criminelles. Néanmoins, la formalisation de l'approche et des processus démontre que la production de renseignement criminel ne peut être réduite à une baguette magique qui serait capable d'extraire mystérieusement des connaissances pertinentes à partir des données à disposition. Cet idéal véhiculé par certaines approches en data mining n'apparaît pas réaliste en analyse criminelle. Un mouvement particulier, plus pragmatique, appelé data mining guidé par le domaine (D3M) (Cao, 2008) a retenu notre attention : il recommande d'injecter des connaissances *a priori* dans les processus, tout en restant ouvert à la découverte de nouveautés. Cela nous mène à l'importance de construire un cadre de travail interdisciplinaire plus ambitieux en sciences criminelles, qui sera susceptible de structurer l'approche de manière plus approfondie. Un tel cadre de travail, que l'on qualifiera de *Criminologie Forensique Computationnelle* (CFC) vise à délivrer l'analyse et le renseignement criminel, en se basant sur les données de la criminalité générées par les traces, analysées avec les méthodes computationnelles, et expliquées/supportées par les théories criminologiques.

TABLE DES MATIÈRES

AVANT-PROPOS

Le travail effectué dans cette thèse s'inscrit dans le contexte général opérationnel de l'analyse criminelle. Aussi lorsqu'il est fait mention du terme « analyse criminelle », il est entendu au sens large en englobant l'analyse criminelle opérationnelle et le renseignement criminel opérationnel, même si les deux notions ne sont pas précisées à chaque fois dans un souci de lisibilité. De même, lorsque le terme « renseignement » est indiqué, il est entendu « renseignement criminel », sauf mention contraire.

Les termes «auteur», «délinquant» et «criminel» font indifféremment référence à la notion d'individu susceptible d'être impliqué dans un acte litigieux.

Par ailleurs, les termes «crime», « criminalité », « activité criminelle », «acte litigieux » ou «délit» ne sont pas représentatif d'une catégorisation légale ou théorique, mais renvoient systématiquement à une notion plus large de l'objet « crime » entendu comme un comportement déviant problématique.

Enfin, dans un souci de lisibilité, il a été décidé d'arrondir tous les nombres à maximum deux décimales après la virgule.

LISTE DES ABRÉVIATIONS

AFNOR : Association Française de Normalisation

BPMN : Business Process Model and Notation

CFC : Criminologie Forensique Computationnelle

CH : Confédération helvétique

CICOP : Concept Intercantonal de Coopération Opérationnelle et Préventive

D3M : Domain Driven Data Mining

DAB : Distributeur automatique de billets

DCRJ : Division Coordination et Renseignement Judiciaire

FCPD : Fuzzy Change Points Detection

JEP: Journal d'événements police

KDD : Knowledge Discovery Databases

PICAR : Plateforme d'Information du CICOP pour l'Analyse et le Renseignement

SARA : Scanning, Analysis, Response, Assessment

VD : Vaud

INTRODUCTION

Parallèlement à l'essor des nouvelles technologies de l'information et de la communication, les différents acteurs de la sécurité, dont notamment les services de police, se retrouvent désormais confrontés à des quantités croissantes de données de criminalité, quand elles ne sont tout simplement pas de nature nouvelle (p. ex. les traces numériques). Un des défis de l'analyse criminelle est de faire face à ces nouvelles et grandes quantités de données dans le but de les détecter, les collecter, les traiter, les analyser et les exploiter en informations utiles tant à l'investigation qu'au renseignement criminel. Le traitement des données à l'aide de classification automatique et la détection de problèmes à l'aide des tendances dans les données de la criminalité représentent notamment des objectifs majeurs des unités d'analyse. Les sciences de l'information semblent pouvoir contribuer à cette adaptation en mettant à disposition des méthodes et techniques computationnelles susceptibles de soutenir ces processus de détection en analyse criminelle. Les méthodes computationnelles de détection de ruptures dans des tendances ou de classification automatique sont devenues classiques et leur potentiel a déjà été testé sur des données de criminalité. Le sujet semble donc épuisé et représente peu de nouveauté et d'intérêt pour les chercheurs explorant ces champs. Il faut toutefois bien constater qu'elles sont paradoxalement très peu utilisées dans les pratiques de l'analyse criminelle. Quelles sont alors les raisons de ce décalage ?

Nous postulons qu'il manque une démarche d'intégration de ces méthodes dans une approche interdisciplinaire mobilisant des théories criminologiques et forensiques. Le questionnement sur la nature de cette interdisciplinarité et l'expression d'une telle approche limitée à l'étude de quelques questions spécifiques (classification automatique et détection de ruptures dans les tendances de la criminalité) constituent les contributions fondamentales de cette thèse.

Dans une première partie, le contexte et la problématique qui entourent ces interrogations sont exposés. Pourquoi est-ce que la question de l'intégration de méthodes computationnelles en analyse criminelle se pose ? Est-ce raisonnable de l'envisager ? Et si oui, quelle approche adopter ?

Dans la deuxième partie, la démarche méthodologique sélectionnée et la réflexion l'entourant sont présentées. Quels sont les fondements théoriques de ce processus ? Comment s'intègrent les différentes disciplines ? La considération d'une unité d'analyse au sein d'un service de police comme objet d'étude permet de guider la réflexion et d'apporter des éléments de réponses.

La troisième partie confronte l'approche développée avec une application opérationnelle en présentant la manière dont la composante computationnelle est intégrée au sein du processus. L'emphase est posée sur une problématique récurrente en analyse criminelle : les fréquents changements ou ruptures dans les patterns d'activités criminelles répétitives, autrement dit, les tendances dans les données de la criminalité. Une expérimentation conduite au sein de l'unité d'analyse criminelle étudiée est présentée. La combinaison de modèles, méthodes ou inférences provenant de différents domaines ouvre la voie à un développement prometteur et pragmatique de l'automatisation du processus.

Finalement, dans la dernière partie de cette thèse, une discussion générale est menée sur la manière dont les différents champs de recherche peuvent s'engager dans une approche commune. Un cadre de travail intégratif et interdisciplinaire est proposé et discuté. Les limites et les perspectives soulevées par ce travail sont également considérées.

Partie I : PROBLÉMATIQUE

Les mutations de la criminalité (p. ex. mobilité accrue, développement dans de nouveaux espaces numériques et économiques) et la nouvelle traçabilité des activités humaines par l'usage des technologies de l'information et de la communication appellent à renforcer le traitement des informations et le renseignement qui pilotent l'action de sécurité. La constante augmentation des données à traiter et la nécessité de célérité quant au renseignement produit rendent les techniques computationnelles, et particulièrement les techniques de data mining, attractives pour l'analyse criminelle. Elles suscitent également une inquiétude vis-à-vis de son application dans ce domaine sensible, notamment à cause des atteintes perçues ou possibles aux libertés individuelles. On comprend dès lors l'importance de maîtriser l'intégration des sciences de l'information au sein de l'action de sécurité.

Cette première partie permet de situer la problématique générale de cette recherche et se décompose en trois temps. En premier lieu, elle propose d'apporter un éclairage sur la place du renseignement criminel au sein de l'action de sécurité et des sciences de l'information. Elle s'interroge tant sur le contexte des productions scientifiques en la matière que sur le contexte professionnel qui voit l'émergence du *predictive policing* parmi les services de police (chapitre 1). Dans un second temps, une revue de littérature recense les diverses tentatives d'intégration des méthodes computationnelles en renseignement criminel (chapitre 2). Finalement, cette première partie s'achève sur la formulation des questions de recherche qui diligentent la réalisation de cette thèse (chapitre 3).

1. Action de sécurité et sciences de l'information

1.1. Des données en quantités, des données à traiter

Hilbert et López (2011) ont estimé que la capacité mondiale en termes de calcul computationnel présente une croissance annuelle de 58% et que la capacité de stockage de l'information augmenterait de 23% par année. Cette croissance est visible notamment depuis le début des années 2000 et l'avènement de l'ère digitale. C'est dans ce contexte que la notion de *big data* est apparue. Une définition sommaire se résume à parler de larges volumes de données informatiques ou mégadonnées selon le terme français consacré[1]. Une définition plus précise permet de dégager trois dimensions caractérisant le concept de *big data* (Laney, 2001). Il y a bien sûr l'idée de volume, car ce sont de grandes quantités de données qui caractérisent ce courant. Mais il y a également la variété qui caractérise le *big data*. Les données sont certes plus nombreuses, mais également de différents types, ceci étant rendu possible par la traçabilité frénétique des activités humaines qui génère ces données et la diversification des méthodes de collecte. Ces données présentent alors toute une diversité de caractéristiques propres qui ne pourront pas être appréhendées de la même façon par des analystes chargés d'interpréter et d'exploiter ces informations. La dernière dimension est la vélocité. Cet élément caractérise le rythme auquel les données sont générées, traitées et utilisées. De manière générale, la flexibilité d'un système dépend beaucoup de cette dimension. De plus en plus de processus nécessitent une analyse en temps réel, par exemple dans le domaine de la finance, où le suivi quotidien de la bourse requiert une attention et une flexibilité particulières pour s'adapter à l'évolution des marchés financiers.

L'augmentation du volume, la diversification et la rapidité de traitement des données se posent donc comme les fondements de l'avènement du *big data*. Selon Cukier et Mayer-Schönberger (2013), cette émergence change profondément la façon dont est utilisée l'information. Premièrement, la quantité de données utilisée à des fins d'analyse est beaucoup plus importante qu'autrefois, où les analystes ne collectaient que de petits échantillons à cause des limites liées à la collecte et au traitement de ces données. Désormais, l'information est facilement disponible d'un point de vue

[1] Recommandé par la délégation générale à la langue française et aux langues de France (https://www.legifrance.gouv.fr/affichTexte.do;jsessionid=95C98A08EB169EEA13DBBD68B92C83E6.tpdila18v_1?cidTexte=JORFTEXT000029388087&dateTexte=&oldAction=rechJO&categorieLien=id&idJO=JORFCONT000029387119, consulté le 08.03.2017).

technique (mais pas aussi facilement exploitable, spécialement dans un contexte judiciaire) et il n'y a plus besoin de se contenter d'extraits choisis. Conséquence directe de cela, le second changement concerne la pertinence de ces données. Là où de petits échantillons contrôlés limitaient l'incertitude, l'augmentation considérable des volumes de données l'accroît. Les analystes tolèrent ainsi des données plus incertaines au nom d'un calcul rationnel qui voit le potentiel d'utilisation d'un vaste jeu de données surpasser les limites liées à leur incertitude.

Transposées dans un contexte policier, les caractéristiques évoquées du *big data* se retrouvent également par analogie, dans une mesure moindre néanmoins[2]. Les conséquences du développement de la capacité de stockage se répercutent sur le nombre de données que peuvent enregistrer les services de police, augmentant ainsi leur volume. L'omniprésence des téléphones portables et des réseaux sociaux sur internet permet aux délinquants de générer de nouvelles sortes de traces de leurs activités illicites qui sont susceptibles d'être détectées et collectées. Des données de plus en plus variées sont ainsi enregistrées. La valeur de l'information prend cependant une signification différente et les limites qu'impose le système judiciaire en matière de libertés individuelles restreignent l'accessibilité aux données. Finalement, la vélocité, c'est-à-dire le rythme auquel les données sont générées, traitées et utilisées, est une caractéristique essentielle du travail policier, puisque la production de renseignement criminel est liée à sa flexibilité pour s'adapter à l'évolution de la criminalité.

1.2. Du *big data* au *data mining*

Une des principales problématiques soulevées par le *big data* se traduit par la difficulté de traiter les données de manière exhaustive. Des méthodes automatisées pour extraire les schémas pertinents des données sont susceptibles d'apporter une réponse adéquate et si les services de police, et l'action de sécurité en général, doivent gérer cette problématique, une des solutions envisagées est le recours aux techniques dites de *data mining*. Le *data mining* qui peut être traduit par « fouille de données », ou encore « forage de données » est un ensemble de méthodes et techniques dont le but est l'exploration et l'analyse de grandes bases de données afin de détecter des règles et des patterns inconnus ou dissimulés (Tufféry, 2007). Ces techniques peuvent être issues de méthodes statistiques, informatiques ou encore relevant de l'intelligence artificielle.

[2] Les quantités de données traitées par les services de police restent restreintes comparées au gigantesque volume de données traité dans le domaine financier ou marketing. Cependant, la criminalité financière peut requérir des traitements de données qui relèvent du *big data*.

Ces techniques, auparavant reléguées au domaine théorique, connaissent depuis près d'une dizaine d'année un essor fulgurant dans les milieux pratiques, notamment dans le domaine marketing. Cela est dû notamment à la convergence de différents facteurs selon Berry et Linoff (2004) :

- De larges volumes de données sont produits.
- Les données peuvent être entreposées dans une mémoire.
- La puissance de calcul est devenue abordable.
- L'intérêt pour la gestion des relations clients est fort.
- Les logiciels commerciaux de data mining deviennent disponibles.

La définition du *data mining* reste cependant floue dans la mesure où les critères d'appartenance (grande quantité de données, exploration, détection de patterns) sont sujets à interprétation selon le domaine d'application. Afin de rester suffisamment exhaustifs, nous préférons employer le terme de « méthodes computationnelles » que nous définissons dans le cadre de cette thèse comme les méthodes et techniques réalisées à l'aide d'un programme informatique et dont l'objectif est de traiter des données en vue de produire de l'information. Ce concept s'inscrit de manière plus large dans le domaine des sciences de l'information et permet ainsi d'inclure les techniques de *data mining*, sans se limiter à celles-ci.

1.3. Le monopole technique

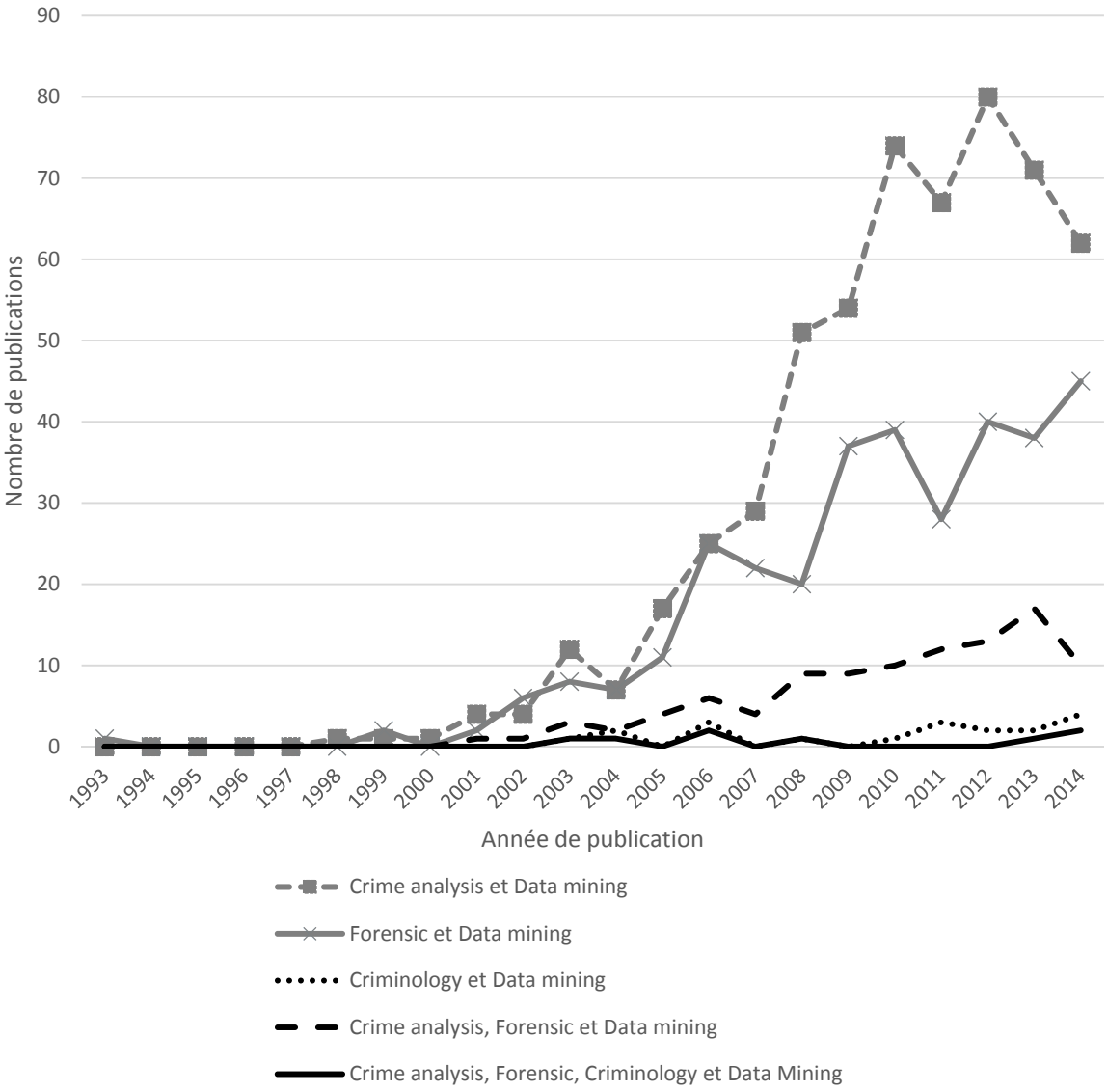
Dès lors qu'il est question de combiner des méthodologies, techniques, ou méthodes de deux domaines différents, respectivement l'analyse criminelle et les sciences de l'information, il est important de s'interroger sur la production scientifique qui les concerne. Cette situation n'est pas nouvelle en analyse criminelle, car ce domaine a longtemps attiré des chercheurs issus de nombreuses disciplines. Lorsque l'on parle de l'objet « crime », celles qui paraissent les plus pertinentes sont la criminologie, la science forensique et le droit pénal. Mais comment ces différentes disciplines se combinent-elles dans le paysage scientifique actuel ? Quel profil de chercheur s'est emparé de cette question ?

Loin d'être une analyse documentaire exhaustive, un bref point de vue sur les publications scientifiques répertoriées sur la base de données en ligne SCOPUS[3] fournit des éléments de réponse. Différents termes ont été introduits dans le moteur

[3] www.scopus.com

de recherche de SCOPUS qui ensuite a retourné le nombre d'occurrences apparaissant dans le titre, les mots-clés et le résumé des publications référencées dans la base de données. Le terme « méthodes computationnelles » étant trop générique, l'utilisation du terme « *data mining* » a été préférée. Ce type de techniques reflète bien l'idée que l'on peut se faire de l'application de méthodes computationnelles sur des grands volumes de données. Différentes combinaisons ont été testées en combinant le terme « *data mining* » avec, non seulement le terme « *crime analysis* », mais également les termes « *criminology* » et « *forensic* ».

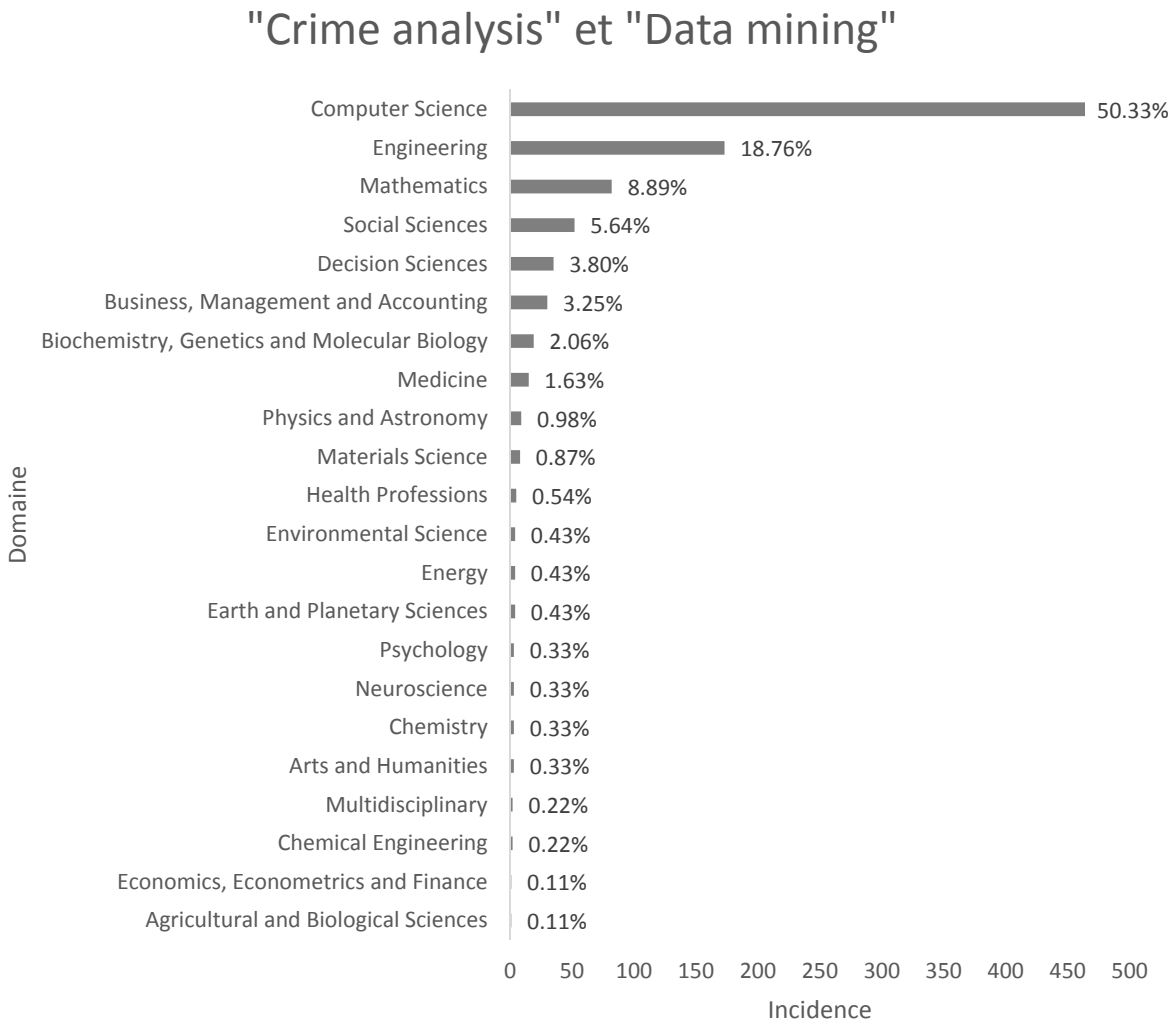
Un premier aperçu longitudinal montre très clairement une émergence du data mining en analyse criminelle, criminologie et science forensique au début des années 2000 (Graphique 1). En revanche, il y a relativement peu de publications intégrant la science forensique à l'analyse criminelle et au data mining. La science forensique, comme discipline, semble ne pas avoir été invitée dans ce débat. L'intégration de celle-ci avec des méthodes computationnelles est perçue comme technologiquement pointue et susceptible de rajouter des difficultés potentielles pour les praticiens. De plus, la science forensique est souvent considérée afin d'aider une cour de justice à prendre des décisions fondées sur des informations véhiculées par les traces (preuves), ce qui requiert une approche radicalement différente du traitement de l'information par rapport aux objectifs et méthodes de l'analyse criminelle. Néanmoins, certaines applications pragmatiques ont montré la solidité et le potentiel informatif de l'apport de la science forensique en analyse criminelle (Braga & Pierce, 2004; Grossrieder, Albertetti, Stoffel, & Ribaux, 2013), et le développement de la forensique computationnelle souligne les changements qui sont en train de se produire. Cette dernière approche se définit comme un domaine de recherche émergent dont l'objectif est de munir la science forensique de méthodes, modèles et outils computationnels pour faciliter son adaptation au récent contexte du *big data* (Franke & Srihari, 2008). Pourtant, cette forensique computationnelle est encore trop restrictive dès lors qu'elle ne fait presque aucune référence explicite aux connaissances criminologiques pertinentes en analyse criminelle. Ce constat est également visible sur le Graphique 1 où les publications sur la banque de données SCOPUS regroupant la criminologie, le *data mining*, l'analyse criminelle et la science forensique sont faiblement représentées. Il faut toutefois nuancer ces résultats, car les méthodes de collecte de SCOPUS ont pu évoluer avec le temps, notamment avec l'apparition du mot « *data mining* » dans les mots-clés disponibles.



***Graphique 1** : Distribution des publications scientifiques par mots-clés en fonction des années sur SCOPUS (n= 1'019). Source : www.scopus.com, état au 04.08.2015.*

Lorsqu'on observe la distribution des occurrences de termes en fonction du domaine dans lequel est classée la publication, le constat suit une tendance très nette (Graphique 2). Près de 80% des publications où figurent les termes « *crime analysis* » et « *data mining* » sont considérées comme relevant de l'informatique, de l'ingénierie et des mathématiques. Les sciences sociales ne concernent que 5.64% des publications et seulement 0.22% prennent l'étiquette « multidisciplinaire ». Un constat similaire peut être tiré avec les publications sur le data mining, la science forensique et la criminologie[4].

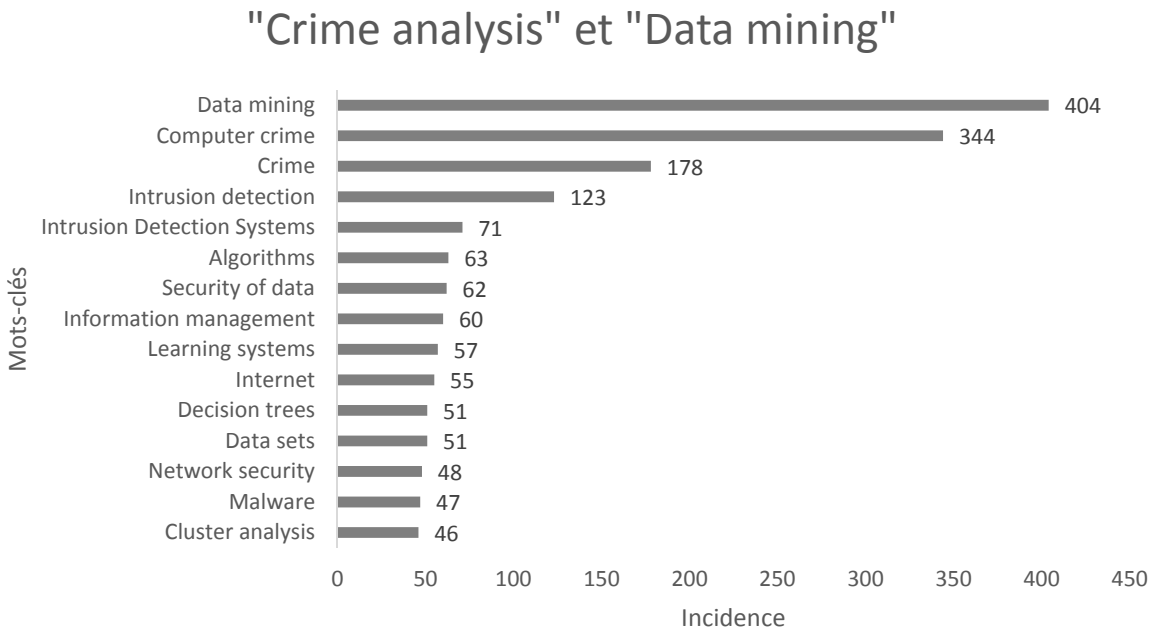
[4] Pour les graphiques détaillés, voir Annexe A



***Graphique 2** : Taux d'incidence des termes « crime analysis » et « data mining » dans les titres, mots-clés et résumé des publications scientifiques répertoriées sur SCOPUS en fonction du domaine de publication (n= 922). Source : www.scopus.com, état au 04.08.2015.*

De même, lorsqu'on observe les mots-clés les plus utilisés pour labelliser les publications correspondantes, on note une forte majorité de mots-clés liés à des techniques d'analyse computationnelles et à la criminalité informatique (Graphique 3). Comme précédemment, le même constat s'impose avec les publications sur le *data mining*, la science forensique et la criminologie[5].

[5] Pour les graphiques détaillés, voir Annexe A



***Graphique 3** : Taux d'incidence des termes « crime analysis » et « data mining » dans les titres, mots-clés et résumé des publications scientifiques répertoriées sur SCOPUS en fonction des mots-clés (n= 3'875). Classement des 15 scores les plus élevés (n= 1'660 ; 42.84%).*

Bien que conscient du caractère exploratoire et non exhaustif de cette rapide illustration, il en ressort une tendance à la monopolisation technique de la question d'utilisation des méthodes computationnelles en analyse criminelle. Cette problématique est majoritairement empoignée par des académiciens basant leur approche sur le cœur de leur discipline, principalement issue des sciences de l'information en général. L'analyse criminelle n'est alors à leurs yeux qu'un champ d'application parmi d'autres. Leur contribution relative au sein de la méthodologie d'analyse criminelle en devient difficile à situer.

1.4. Lacunes dans l'expression des modèles

Parallèlement au positionnement du monde académique sur l'application de méthodes computationnelles en analyse criminelle, la question de l'implémentation de ces innovations technologiques se pose également. La planification et la gestion de projets informatiques sont susceptibles de fournir certaines indications sur la méthodologie à adopter en vue de réfléchir à une intégration pragmatique en analyse criminelle. Le paysage fédéraliste suisse offre un terrain d'essai particulièrement pertinent lorsqu'il s'agit d'illustrer les risques liés à la mise en place de projets informatiques. Une série d'exemples de projets conduits au niveau fédéral durant les années 2000 et ayant connu quelques déboires est susceptible d'apporter un éclairage complémentaire dans

notre démarche[6]. Cet éclairage demeure cependant illustratif et n'a pas pour vocation d'analyser en profondeur les facteurs ayant causé ces différents échecs. L'étude des méthodologies de travail employées nécessiterait un travail à part entière pour en saisir les enjeux détaillés. Bien que les raisons de ces échecs soient très souvent attribuées à une absence de planification générale tant technique qu'organisationnelle, nous sommes d'avis que les causes sont plutôt à chercher dans une absence d'expression des modèles et processus sous-jacents au domaine, ce qui engendre des difficultés dans le développement des projets, ainsi que des retards et des dépassements de budgets. Dans le contexte de l'analyse criminelle, cela se traduit notamment par le fait que les activités policières en générale, et en police judiciaire en particulier, sont bien codifiées par les procédures, mais mal exprimées du point de vue des problèmes et de leur résolution. Il s'ensuit une confusion entre les éléments administratifs et procéduraux et les raisonnements qui conduisent à la résolution de problèmes. Ce dernier aspect met en jeu l'incertitude, les contours flous de certaines formes de criminalité, et la difficulté à saisir cette réalité dans des modèles. Cette dimension est souvent sous-estimée par les professionnels eux-mêmes, peu entraînés et intéressés à exprimer formellement les dimensions fondamentales de leur propre activité (Kind, 1987).

Bien que la volonté de conduire ces projets à l'échelle fédérale soit louable, il devient difficile de s'adapter aux spécificités locales pourtant primordiales dans un état fédéraliste comme la Suisse. Le fait de faire appel à des partenaires externes présente également le risque de confier la réalisation du projet à des techniciens spécialistes dans leur domaine, mais qui ne saisiront pas forcément la volonté du mandant, celle-ci n'étant pas toujours explicite. Se pose alors la question d'un malentendu systématique quant à l'expression des modèles qui fondent le système informatisé. Les sommes considérables investies par les décideurs politiques sont également une source de problèmes potentielle. Portant souvent sur des millions de francs à l'échelle de la Suisse, ces sommes génèrent des pressions importantes sur les personnes qui portent la responsabilité de ces projets, celles-ci étant souvent limogés et remplacés en cas d'évaluation négative de l'avancée du projet en termes de délai ou de prestation. Ainsi,

[6] On peut citer parmi ces derniers les projets suivants :
INSIEME en 2001 (https://www.news.admin.ch/message/index.html?lang=fr&msg-id=46038);
ISS en 2013 (http://www.ejpd.admin.ch/ejpd/fr/home/aktuell/news/2013/2013-09-20.html);
novoSIPAC en 2015 (https://www.admin.ch/gov/fr/accueil/documentation/communiques.msg-id-58255.html);
et Mistra en 2014 (https://www.news.admin.ch/message/index.html?lang=fr&msg-id=50513).

il peut y avoir un fort tournus parmi les personnes impliquées dans la réalisation du projet, impliquant de fait une gestion et une planification plus complexe. La durée souvent longue de ces projets nourrit ce tournus naturellement par les mutations et les évolutions diverses de carrières des individus. Finalement, l'idée de rebâtir sur du neuf avec une tendance à balayer l'existant empêche une réflexion continue sur la problématique et ralentit considérablement les efforts employés dans la résolution du problème. Cela complexifie également le développement, car la solution de remplacement est dans la majeure partie des cas terminée avant de pouvoir remplacer l'ancien système, ce qui ne permet que rarement des tests en conditions réelles.

Les différentes caractéristiques communes de ces grands projets informatiques sont ainsi susceptibles de générer un certain nombre d'inconvénients :

- Une difficulté d'exprimer des modèles suffisamment bien formalisés pour être traduits en systèmes informatisés.
- Un tournus important des personnes impliquées dans le projet qui ne permet pas de choisir les personnes adéquates capables de construire des modèles.
- Une négligence des solutions et des réflexions déjà existantes.

La combinaison de ces différents inconvénients augmente les risques d'une perte de maîtrise du projet. Celle-ci est susceptible de se traduire par des retards, un manque de transparence de la part du mandataire, une prestation non conforme aux volontés du mandant, et des dépassements de budget. L'importance et la gravité de ces éléments peuvent amener à une réévaluation partielle ou complète du projet, voire à son annulation totale, entraînant ainsi une perte financière importante et réduisant les chances de parvenir à une solution viable dans l'avenir.

Ces grands projets informatiques ont ainsi souvent trouvé une issue défavorable. Dans le cadre de l'analyse et du renseignement criminel, le risque est moindre, car l'action de sécurité est principalement conduite à l'échelle cantonale. Néanmoins, dans le contexte actuel cherchant à renforcer la collaboration intercantonale (CLDJP, 2013), il n'est pas impossible de se retrouver dans une situation similaire aux exemples mentionnés plus haut.

L'enjeu est donc d'anticiper ces possibles développements et d'éviter de tomber dans les mêmes écueils que les grands projets informatiques. Un développement de méthodes computationnelles en vue d'une application en renseignement criminel

augmenterait ses chances de succès s'il adopte une méthodologie itérative, à une échelle réduite dans un premier temps, qui construit, bloc après bloc, sur les solutions déjà implémentées. Une telle démarche permet de réfléchir en termes de sous-projets, plus rapides à implémenter, moins coûteux, et moins complexe à produire. Il est également important que la réflexion se fasse conjointement avec des professionnels sensibles aux spécificités du domaine, qu'ils soient présents en interne ou non, afin de garantir une solution adaptée aux contraintes du renseignement criminel. Cette approche dite « située » en élaboration de logiciel (*software design*) est basée sur le fait que les individus sont impliqués dans des interactions continues au sein d'un environnement en constante évolution (Hofmann, Pfeifer, & Vinkhuyzen, 1993).

Apprendre des erreurs passées en matière de conduite de projet informatique permet d'orienter un peu mieux la réflexion autour de l'intégration de méthodes computationnelles en analyse criminelle. Ces échecs sont susceptibles de constituer un indice d'une expression trop faible des théories sous-jacentes à l'analyse criminelle et de l'entêtement à ne pas reconnaître ces faiblesses. Nous postulons que l'analyse des données de criminalité ne peut pas se réaliser sans faire explicitement référence à des connaissances qui intègrent des éléments de criminologie et de science forensique. Des éléments méthodologiques semblent donc se dégager de nos observations, mais au-delà de ces remarques en matière de gestion de projet, il existe d'importantes considérations portant sur les objectifs mêmes des méthodes computationnelles.

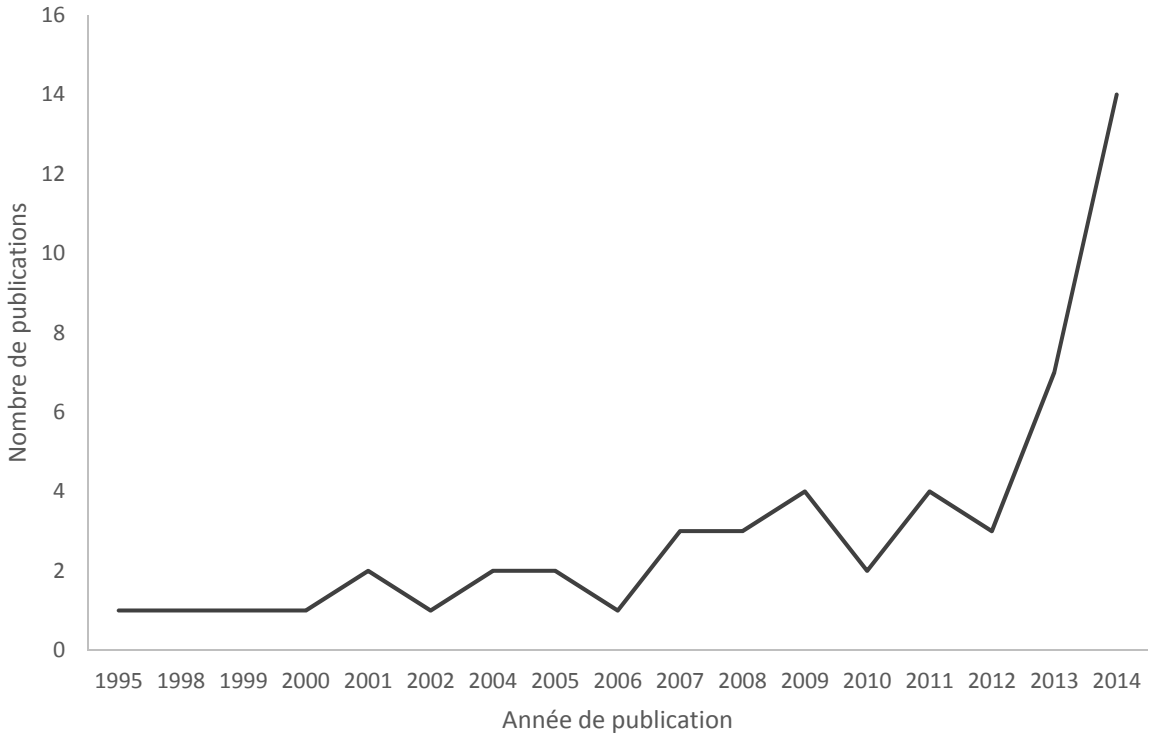
Le besoin d'appliquer ces techniques en renseignement criminel est difficilement réfutable, mais que demande-t-on réellement à ces algorithmes ? Quels sont leurs objectifs ? Leurs limites ? Et quels enjeux sont-elles susceptibles d'alimenter ? Dans quelle méthodologie ou cadre formel prennent-ils leur place ?

Comprendre comment les méthodes computationnelles sont appliquées en renseignement criminel dans le contexte actuel et comment a été pensée leur intégration peut fournir des éléments de réponse à ces interrogations.

1.5. Les « mythes » des promesses algorithmiques

Le monopole académique des sciences de l'information et les risques liés à la réalisation de vastes projets informatiques n'ont cependant pas freiné l'implémentation de méthodes computationnelles supposées soutenir les méthodes de l'analyse et du renseignement criminel. Dans la poursuite d'un idéal d'anticipation et

de prévention, un nouveau courant a récemment émergé en matière d'action de sécurité. Appelé *predictive policing* en anglais, que l'on pourrait traduire par police prédictive, ce modèle est rapidement devenu populaire parmi les forces de police (Bratton & Malinowski, 2008; Friend, 2013; Pearsall, 2010; Perry, McInnis, Price, Smith, & Hollywood, 2013; Vlahos, 2012). Un bref point de vue sur la littérature scientifique tend à confirmer cette tendance émergente à partir des années 2010 comme l'indique le Graphique 4. Selon Perry et ses collègues (2013), le *predictive policing* peut se définir comme étant l'application de techniques d'analyse, généralement de nature quantitative, en vue d'identifier les cibles potentielles d'interventions policières et de prévenir le crime à l'aide de prédictions statistiques.



***Graphique 4** : Distribution par année des publications sur le predictive policing référencées dans la base de données SCOPUS (www.scopus.com, état au 04.08.2015).*

Plusieurs outils commerciaux ont dérivé de ce modèle dont, parmi les plus connus, figurent Blue CRUSH[7], opération initiée en 2005 à Memphis (Perry et al., 2013) et dédiée à l'analyse de patterns spatio-temporels en matière de crimes violents et sérieux ; PredPol[8], lancé en 2011 et qui cherche à prédire quand et où un type de crime

[7] http://www-03.ibm.com/press/us/en/photo/32175.wss
[8] https://www.predpol.com/

est susceptible de se produire ; et Precobs[9], dernier né dans la famille des logiciels de prédiction du crime avec un développement en 2013 et qui se spécialise dans la prédiction des cambriolages. Le fonctionnement de ces solutions est souvent comparable à des boîtes noires et il devient difficile pour les utilisateurs potentiels d'appréhender leurs bases mathématiques sous-jacentes. Les outils cités en sont un parfait exemple. L'opération Blue CRUSH utilise le logiciel *IBM SPSS predictive analytics*[10] qui propose une sélection d'extensions destinées à « prédire avec confiance ce qu'il va se passer pour ainsi prendre des décisions plus intelligentes »[11]. Il n'est en revanche pas possible de connaître précisément comment fonctionnent les différents algorithmes qui constituent le logiciel. La plupart des descriptions disponibles en sources ouvertes restent opaques et imprécises. Un exemple type peut être observé dans un article de présentation de l'opération Blue CRUSH :

« Les analystes compilent des données chaque jour en utilisant *IBM predictive analytics software*, lequel mixe des quantités d'information et montre des rapports d'incident en quelques secondes, tout en incluant des sources de données issues des patrouilles, concernant les types d'infractions, le moment de la journée, le jour de la semaine ou diverses caractéristiques de la victime et de l'auteur. L'agence est capable d'intégrer IBM SPSS et un système d'information géographique pour à la fois analyser et visualiser les données sous forme de graphiques, cartes géographiques, et rapports. »[12](Phelps, 2010).

On retrouve en substance cette idée de boîte noire dans laquelle on introduit des données de toutes sortes. Elles sont ensuite traitées de manière opaque pour produire des résultats sous la forme séduisante de cartes et de graphiques dont il devient difficile d'évaluer la pertinence.

En plus de l'opacité mathématique qui accompagne ces solutions de prédictions du crime, s'ajoute la relative complexité de ces modèles. Le logiciel PredPol est basé sur

---

[9] http://www.ifmpt.de/
[10] http://www-01.ibm.com/software/analytics/spss/11/na/cpp/
[11] Citation originale: "[…] *predict with confidence what will happen next so that you can make smarter decisions* […]".
Source: http://www-01.ibm.com/software/analytics/spss/?source=homepage&hpzone=nav_bar
[12] Citation originale : "*Analysts pull data each day using the IBM predictive analytics software, which crunches volumes of information showing incident reports in seconds, including incoming data sources from patrols, pertaining to type of criminal offense, time of day, day of week or various victim/offender characteristics. The agency is able to integrate IBM SPSS and a geographic information systems tool to both analyze and visualize data in the form of charts, geographical maps, and reports.*"

les travaux de Mohler et ses collègues qui ont adapté un processus de « points auto-excitants » (*self-exciting point process*) en sismologie, afin de modéliser les cambriolages d'habitations (Mohler, Short, Brantingham, Schoenberg, & Tita, 2011). Malgré une transparence louable avec la mise à disposition de leur méthodologie détaillée, il apparaît difficile et peu probable qu'un praticien impliqué dans les politiques d'action de sécurité soit à même de comprendre et d'apprécier les équations lui permettant de paramétrer de manière adéquate le modèle (Figure 1).

$$\lambda_{lm}(t) = \mu_{lm} + \sum_{\{t_k<t:\ i(k)=l,\ j(k)=m\}} g(t-t_k).$$

$$l(K_0,\omega) = \sum_{lm}\left\{ -\mu_{lm} + \sum_{\{k:\ i(k)=l,\ j(k)=m\}}\left[\log\left\{\lambda_{lm}(t_k;K_0,\omega)\right\} - \int_{t_k}^{T} g(t-t_k;K_0,\omega)dt\right]\right\}$$

***Figure 1** : Principales équations de Mohler et al. (2013). En haut, fonction de densité conditionnelle. En bas, fonction de vraisemblance. Au-delà de la signification de chaque paramètre, c'est la complexité inhérente de l'approche qui est soulevée ici.*

Loin d'être inutile et s'appuyant sur des fondamentaux solides (la criminalité suit des schémas susceptibles de se reproduire), les techniques de *predictive policing* sont néanmoins accompagnées d'idées reçues ou de « mythes » véhiculés entre autres par les campagnes commerciales agressives de logiciels de prédiction. Dans un rapport du *National Institute of Justice,* Perry et ses collègues (2013) expriment ces différents mythes entourant le *predictive policing*.

Une première idée reçue consiste à croire que l'ordinateur connaît réellement le futur. Souvent présenté comme des « boules de cristal » capable d'anticiper le crime, c'est en réalité de probabilités qu'il s'agit. C'est donc plus une évaluation des risques qu'une réelle prédiction. Cette évaluation est extrapolée à partir du passé ce qui signifie que la qualité du résultat obtenu dépend de la qualité des données utilisées en entrée. Nous sommes donc loin des promesses commerciales assurant un résultat concluant à partir de toutes sortes de données possibles.

Un autre grand mythe entourant le *predictive policing* est la croyance que l'ordinateur fera tout pour nous. Ici encore les stratégies commerciales mettent en avant une simplicité et une autonomisation presque complète. On peut lire sur la page internet du logiciel PredPol :

« Pour les analystes, l'outil PredPol peut être installé en quelques jours et génère ses prédictions exploitables en un clic de souris »[13].

Cette idée reçue occulte l'élément le plus important du processus, l'être humain. Perry et ses collègues (2013) notent bien les tâches essentielles qui dépendent encore presque exclusivement du facteur humain, comme par exemple, identifier des données pertinentes, prétraiter les données pour les adapter à l'analyse, réviser et interpréter les résultats ou encore mettre en place des actions qui exploitent les résultats et évaluer leurs impacts. Les connaissances sur l'environnement criminel doivent être maintenues régulièrement à jour et cela passe nécessairement par des interactions sociales. Les algorithmes gagnent alors leur utilité comme support aux différentes étapes nécessaires à la production de renseignement.

Le besoin d'un modèle à la pointe de la technologie et supposé onéreux est également une préconception erronée au sujet du *predictive policing*. La complexité et la puissance croissante des logiciels de prédiction ne sont pas nécessairement utiles dans les tâches qu'ils sont supposés accomplir. Souvent présentée comme un argument commercial, la puissance de calcul d'un outil de prédiction peut rapidement être assimilée à la qualité de ses résultats.

Finalement, le dernier mythe mis en avant par Perry et ses collègues (2013) laisse supposer que des prédictions précises mènent automatiquement à une réduction majeure de la criminalité. Bien que ce résultat soit, en l'occurrence, l'objectif final escompté, l'algorithme de prédiction n'est qu'une étape d'un processus plus vaste. Ce sont les choix d'actions appropriées qui seront entreprises en aval qui auront un impact réel sur les niveaux de criminalité. Ce constat peut être réalisé également à l'échelle de l'analyse criminelle, car celle-ci ne réduit pas le crime à elle seule. L'analyse criminelle ne fournit qu'un produit à partir duquel il s'agit d'élaborer, puis de déployer une réponse appropriée parmi un éventail de possibilités. La réduction du crime dépend donc d'une série de choix à différents niveaux et ne découle pas uniquement de l'utilisation d'un outil informatisé (Santos, 2014). L'ensemble du processus d'analyse et renseignement criminel doit être pris en considération.

[13] Citation originale : "*For analysts, PredPol's tool can be set up within days and generates its actionable predictions in one click of a mouse.*" Source: https://www.predpol.com/technology/

Ce processus se concentre sur l'utilité tactique au lieu de la précision des prédictions, se repose sur des données valides et fiables, et permet la compréhension des facteurs derrière lesquels se repose la prédiction. De même, négliger l'évaluation des techniques, ainsi que la protection des données individuelles est susceptible de nuire à l'application de techniques de *predictive policing*.

À travers ce questionnement sur les mythes et les écueils à éviter en *predictive policing*, se révèle une problématique cruciale dans un contexte policier qui considère plus sérieusement le changement de nature et de quantité des données traitées et qui s'apparente toujours plus à ce que l'on appelle le *big data*. Lorsque les ordres de grandeur changent, c'est toute la méthodologie qui est susceptible d'être reconsidérée. La manière la plus simple de procéder pour les décideurs est d'acquérir des logiciels de prédiction du crime, souvent présentés comme une solution miracle, afin de montrer une préoccupation à la modernisation de la police (Dupont, 2016). Cependant, derrière ces boîtes noires, les mécanismes demeurent obscurs et les effets bénéfiques de cette stratégie ne sont pas évidents dans cette opacité. Une des principales lacunes est l'approche dite descendante ou *top-down*, qui vise à concevoir les techniques et les processus de manière générale pour les appliquer sur un problème, en occultant les contraintes et spécificités de l'analyse criminelle. Les outils de prédictions, et plus généralement les méthodes computationnelles, sont vus comme la pièce centrale du processus (Figure 2).

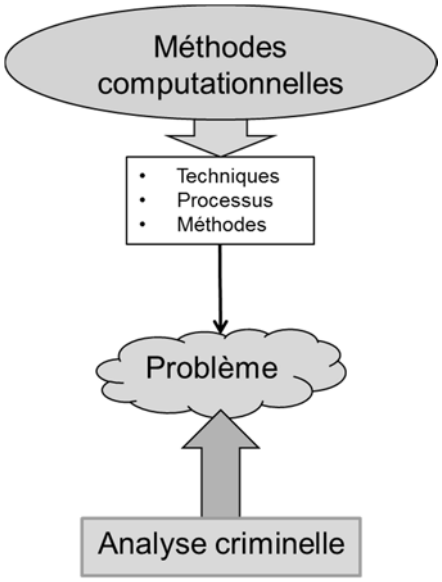


***Figure 2** : Approche top-down de l'application des méthodes computationnelles en analyse criminelle. Les techniques relevant du domaine informatique sont utilisées sur des problèmes communs à l'analyse criminelle. Ces derniers sont vus comme un champ d'application parmi d'autres pour l'utilisation de méthodes computationnelles.*

2. Méthodes computationnelles en renseignement criminel : où en est-on ?

Dans leur ambition extrême, les méthodes computationnelles permettraient de traiter de vastes ensembles de données et nous diraient tout sur le phénomène criminel. Par exemple, toute la variété des formes de répétitions criminelles, ainsi que les schémas criminels prédictibles devraient se dégager d'eux-mêmes par ces traitements, sans devoir les guider par des connaissances *a priori* sur le crime (Grossrieder et al., 2013). Cependant, cette approche idéaliste de méthodes computationnelles sans injection de connaissance en analyse criminelle est difficile à confirmer dans la littérature. Certaines études insistent au contraire sur la manière dont les problématiques liées à la criminalité peuvent être résolues en se basant sur une exploitation intelligente des données dirigée par les connaissances du domaine, p. ex. les connaissances forensiques et criminologiques dans notre cas. Par exemple, certaines recherches ont montré que l'analyse de lien dans des grands jeux de données d'événements criminelles est facilitée par l'utilisation de connaissance *a priori* (Schroeder, Xu, & Chen, 2003; Schroeder, Xu, Chen, & Chau, 2007) ; d'autres chercheurs soulignent que la détection des menaces d'intrusions informatiques est essentiellement basée sur les connaissances du domaine (Young, Goldberg, Memory, Sartain, & Senator, 2013) ; et finalement, l'apport de l'analyse de concept formel dans la détection et le suivi du crime organisé est mis en exergue par Andrews et ses collègues (Andrews, Akhgar, Yates, Stedmon, & Hirsch, 2013). Basée sur les concepts d'entrepôt de données (*data warehousing*), une méthodologie destinée à structurer des données forensiques dans un objectif de renseignement a été proposée (Albertetti & Stoffel, 2012). Cette méthodologie détaille un processus en 5 étapes pour structurer les données issues des journaux d'événements de la police afin de produire des magasins de données (*data marts*) spécifiquement conçus en accord avec les exigences d'une technique analytique.

La plupart des études sur l'utilisation des méthodes computationnelles en analyse criminelle varient tant sur le type de criminalité que sur la diversité des techniques utilisées. Il existe des exemples dans la veille en criminalité financière, et spécialement dans la détection des fraudes à la carte de crédit (Filipov, Mukhanov, & Shchukin, 2008; Whitrow, Hand, Juszczak, Weston, & Adams, 2009). Le même genre de méthodes a également été appliqué sur d'autres types de crimes, tels que certaines formes de crime organisé et de cybercriminalité (Chen et al., 2003; Nissan, 2012) ou le trafic de stupéfiants (Ratle et al., 2008; Terrettaz-Zufferey, Ratle, Ribaux, Esseiva, &

Kanevski, 2006). Les techniques de *clustering* qui servent à regrouper des éléments semblables ont été appliquées dans la détection de séries de cambriolages (Borg, Boldt, Lavesson, Melander, & Boeva, 2014), mais en sous-estimant le potentiel des traces dans le processus. Les arbres de décision figurent également parmi les techniques utilisées en science forensique, par exemple dans le cas des incendies en vue de déterminer la relation entre les conditions climatiques et les feux de cheminée (Holmes, Wang, & Ziedins, 2009), ou encore dans le processus d'investigation avec les facteurs influençant la décision d'analyser une trace (Bitzer, Ribaux, Albertini, & Delémont, 2016).

La logique floue offre une alternative pour représenter le caractère incertain, incomplet et multidimensionnel des données de la criminalité. Plusieurs études ont testé et démontré la capacité des approches floues à gérer des problématiques liées à l'analyse criminelle (Stoffel, Cotofrei, & Han, 2010) et particulièrement la découverte de règle concernant les cambriolages (Cotofrei & Stoffel, 2011). D'autres exemples de l'utilisation de ces approches se traduisent par une cartographie auto-organisatrice floue en prévention de la criminalité (Li, Kuo, & Tsai, 2010), des fonctions d'association floue dans la mise en relation de crimes (Albertetti, Cotofrei, Grossrieder, Ribaux, & Stoffel, 2013a, 2013b), ou encore par un *clustering* flou destiné à détecter les points chauds de la criminalité (Grubesic, 2006).

Au-delà de l'incontestable intérêt de ces modèles, il se dégage une approche à deux vitesses : les chercheurs et praticiens en analyse criminelle considèrent les techniques computationnelles comme un moyen prometteur pour soutenir le traitement des données de criminalités. D'un autre côté, les théoriciens en sciences de l'information considèrent l'analyse criminelle comme un champ d'application qui fournit des données intéressantes pour tester la génération de connaissance à l'aide de modèles computationnels. Les deux visions ne se combinent pas aisément.

C'est peut-être une des raisons pour lesquelles, il y a encore peu d'émergences de réelles approches interdisciplinaires qui intégrerait tant les considérations forensiques, criminologiques et computationnelles. Ce constat peut s'expliquer par la difficulté de situer l'équilibre subtil entre l'injection de connaissance *a priori* (p. ex. des informations sur le type de patterns détectable qui est attendu dans les données) et l'objectif souvent irréaliste d'utiliser des techniques computationnelles totalement non supervisées (Grossrieder et al., 2013). À cela s'ajoute la difficulté de les intégrer

dans un ensemble cohérent en adéquation avec la réalité quotidienne de l'environnement professionnel des analystes.

Certains chercheurs ont réalisé de sérieux efforts afin de traduire les théories disponibles en cadres de travail pragmatiques et en boîtes à outils pratiques. Un exemple type est le manuel d'analyse criminelle en 60 étapes qui trouve ses fondations théoriques dans la criminologie environnementale (Clarke & Eck, 2005). Facilement accessible, du matériel pratique est disponible sur internet sous forme de guide à destination des praticiens (ex. : www.popcenter.org). Plus récemment, les enjeux pratiques sont considérés à travers un état des lieux du *predictive policing* dans le contexte d'autres formes d'action de sécurité proactives (Perry et al., 2013). Des distinctions utiles des différentes formes de répétitions criminelles prévisibles clarifient les potentiels et les limites de cette approche (Cusson, 2008).

Malgré ces efforts, un fort décalage demeure présent. L'analyse criminelle reste largement pratiquée à l'aide de méthodes heuristiques et d'outils informatiques basiques (ex. : simples calculs sur des tableurs ou cartographie criminelle élémentaire), alors que d'un autre côté, les technologies de pointe se développent séparément dans les cercles académiques produisant un ensemble d'outils computationnels sophistiqués, publicisés, commercialisés de manière agressive, et qui font l'objet de l'essentiel des publications scientifiques (Perry et al., 2013).

L'importance, le rôle et le potentiel des méthodes computationnelles en analyse criminelle ne sont pas contestables en regard de la nouvelle traçabilité des activités humaines. Nous avons vu que la valeur ajoutée concrète de ces méthodes est cependant loin d'être évidente. Les approches proposées montrent des limites, et se révèlent quelques fois irréalistes. Les innovations sont difficiles à situer technologiquement ou méthodologiquement (Perry et al., 2013). L'examen préalable des tentatives existantes d'application de méthodes computationnelles en analyse criminelle permet d'en relever les principales limites et lacunes, mais également de distinguer leur potentiel. Il alimente une réflexion constructive sur les méthodes à concevoir, en particulier :

- Le cloisonnement entre les disciplines sur les problématiques concernées par l'analyse et le renseignement criminel.
- Le manque d'expression des modèles et processus.
- La dissonance entre les objectifs opérationnels et les outils d'analyse.

- L’illusion d’une prédiction parfaite du crime par les méthodes computationnelles.
- L’escalade dans la complexité et la puissance de calcul des algorithmes.
- Les préoccupations des décideurs sur la nécessité de moderniser les organisations policières.
- La négligence du cycle complet en analyse et renseignement criminel en ne se focalisant que sur l’étape d’analyse.

3. Questions de recherche

Les différents points de vue investigués précédemment révèlent qu'un large éventail de modèles, méthodes, instruments et techniques sont désormais disponibles pour les analystes criminels, mais ceux-ci sont accompagnés d'autant de difficultés et de spécificités à prendre en compte.

Dans l'optique d'entamer une réflexion sur l'intégration des méthodes computationnelles en analyse criminelle, et à la vue de nos précédentes observations, nous postulons que l'approche envisagée doit satisfaire un certain nombre de conditions :

- Une approche **interdisciplinaire** qui considère non seulement les méthodes computationnelles, mais également la criminologie et la science forensique.
- Une approche **intégrative** qui convoque les connaissances théoriques, les processus méthodologiques et les techniques d'analyse de différents domaines en vue d'un objectif commun et ne considérant pas le problème comme un simple champ d'application propice à valoriser chaque discipline.
- Une approche dite ascendante ou ***bottom-up*** qui se fonde sur le problème plutôt qu'une approche dite descendante ou *top-down* qui part des outils.
- Une approche **itérative** qui construit pas à pas sa méthodologie et qui intègre les nouveaux composants tout en conservant l'acquis ;
- Une approche **concrète** qui nourrit sa réflexion à l'aune de la réalité du terrain et des pratiques opérationnelles en matière d'action de sécurité et qui évalue empiriquement l'efficacité du dispositif.

3.1. L'interdisciplinarité au service de la résolution de problème

Un raisonnement structuré et interdisciplinaire est susceptible de mieux situer et intégrer ces innovations. En effet, il est reconnu dans plusieurs domaines que le besoin de résoudre des problèmes sociaux intrinsèquement complexes et l'influence grandissante des nouvelles technologiques sont des moteurs puissants de l'interdisciplinarité (Committee on Facilitating Interdisciplinary Research, 2004). Contrairement à une démarche pluri- ou multidisciplinaire qui convoque plusieurs disciplines fournissant chacune une perspective différente sur un problème, une intégration interdisciplinaire rassemble des parties indépendantes de connaissance en vue de produire des relations harmonieuses entre elles (Stember, 1991). Cette

orientation est parfaitement inscrite dans la voie qu'emprunte un mouvement en criminologie qui inscrit son action sous le label de « *crime science* » (sciences criminelles ou sciences du crime), comme le note Pease (2010, p. 3) : « […] l'étude du crime devrait être repensée en vue d'engager un plus large panel de disciplines et d'identifier les synergies de ces disciplines dans l'étude du crime et sa prévention »[14].

À travers son orientation interdisciplinaire, une telle approche n'exclut pas les sciences de l'information du débat. Bien au contraire, l'intégration des méthodes computationnelles en analyse criminelle est susceptible d'être optimisée en coopérant avec des chercheurs en sciences de l'information tout en étant guidée par un cadre de travail interdisciplinaire. C'est à cette fin que nous avons étroitement collaboré avec une équipe de recherche de l'Institut du Management de l'Information de l'Université de Neuchâtel, spécialisée dans le *data mining*. Cette collaboration s'est révélée particulièrement fertile. En immergeant ce projet dans une culture interdisciplinaire, la fertilisation croisée des équipes de recherches a permis de garantir l'accès aux connaissances nécessaires aux questionnements de cette thèse, de guider les nombreuses décisions qui les jalonnent et surtout d'intégrer les points de vue adoptés sur un même problème. Ces choix qu'il a fallu prendre tout au long de ce travail ferment volontairement certains pans de la recherche, qui auraient pu avoir également leur place, pour mettre l'emphase sur des points spécifiques jugés pertinents par les deux équipes de recherche. Chaque partenaire s'est engagé dans la recherche de solutions en transcendant les frontières de chaque discipline, en développant des réseaux avec de nombreux chercheurs et en établissant de fortes connexions avec les experts du terrain décrits dans le prochain chapitre.

3.2. Une approche *bottom-up* : l'étude d'une unité de renseignement criminel

Cette thèse esquisse donc une approche intégrative centrée sur le problème (Figure 3) qui a été testée dans une unité régionale d'analyse. Elle s'inscrit dans la durée en accompagnant l'évolution itérative de cette unité de renseignement criminel en Suisse romande. Ce point de départ permet de développer une méthodologie globale et intégrée qui prend en compte les contraintes et les spécificités inhérentes au renseignement et à l'analyse criminelle.

[14] Citation originale: "[…] *the study of crime should be rebadged in an attempt to engage a wider range of disciplines and identify between discipline synergies in the study of crime and its prevention*".

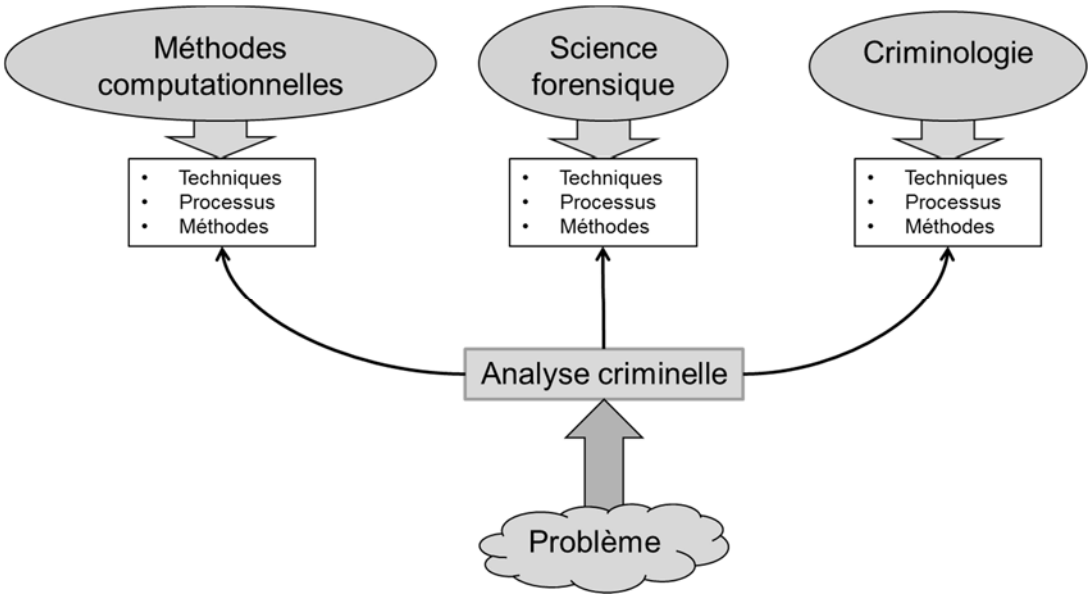


***Figure 3** : Approche intégrative centrée sur le problème. Les techniques et méthodes de diverses disciplines telles que les méthodes computationnelles, la science forensique et la criminologie nourrissent les processus en analyse criminelle avec la résolution de problèmes comme moteur commun.*

3.2.1. Le CICOP et PICAR

L'unité d'analyse considérée dans le développement de la méthodologie opère au sein d'un des quatre centres régionaux d'analyse criminelle en Suisse (Figure 4). Le Concept Intercantonal de Coopération Opérationnelle et Préventive (CICOP) est un réseau englobant 6 services de police en Suisse romande. Ce centre a été conçu au début des années 90, au moment où de nouvelles formes de criminalité sérielle et itinérante émergèrent en Europe et en Suisse, notamment après la chute du communisme et l'ouverture des frontières.

***Figure 4** : Zone d'action des quatre centres régionaux d'analyse criminelle en Suisse (source : CICOP)*

Fort d'environ 25 analystes criminels travaillant en réseau dans cette structure, son objectif est d'opérer un suivi des problèmes répétitifs liés à la criminalité à travers les juridictions et spécialement la criminalité que les Anglo-saxons appellent criminalité de haut volume (*high volume crime*) (Ribaux & Birrer, 2010). Cette mise en réseau est particulièrement active au sein des organismes locaux, à travers notamment les connexions avec les services forensiques, les enquêteurs, les services de prévention et les agents de terrain.

Le fonctionnement du CICOP est soutenu en partie par une banque de données partagée entre les différents cantons. La Plateforme d'Information du CICOP pour l'Analyse et le Renseignement (PICAR) rassemble tous les événements criminels jugés relevants[15] et les codifie en accord avec une méthodologie simple élaborée avec le temps (voir chapitre suivant). La base de données intègre plusieurs autres types d'information (p. ex. rapports externes, véhicules, relations entre scènes de crime à l'aide des traces) et traite également les images (Dessimoz & Champod, 2016). L'une des particularités de cette banque de données est sa classification des événements. PICAR utilise une classification d'inspiration situationnelle (Birrer, 2010) qui permet de codifier des types de situations à l'aide de codes phénomènes. Ces codes sont définis en fonction des caractères circonstanciels de l'événement, comme le mode opératoire, la voie d'entrée, le moment de la journée ou le type de cible. Par exemple, le code GIORNO CILINDRO définit un type particulier de cambriolage d'habitation qui a eu lieu durant la journée, et où le malfaiteur a arraché le cylindre de la porte palière pour s'introduire dans l'appartement. Contrairement à une classification légale classique, les codes phénomènes sont adaptés à la production du renseignement criminel car ils ont été conçus dans l'objectif de faciliter la détection des répétitions criminelles et leur analyse, orientant ainsi naturellement les décideurs vers un éventail circonscrit de réponses appropriées. La manière dont est opérationnalisée cette classification situationnelle est explicitée en détail à travers la thèse de Birrer (2010).

3.2.2. Un développement itératif et intégré

Ces approches pragmatiques en matière de gestion et d'analyse de données furent adoptées afin de progressivement développer une méthodologie intégrée et harmonisée fondée sur les théories des opportunités en criminologie et soutenues par

[15] Principalement les événements de nature sérielle ou itinérante, p. ex. les cambriolages, les vols à l'astuce, ou encore, les agressions.

une base de données intercantonale (Birrer, 2010; Ribaux & Birrer, 2010). La cartographie criminelle fut intégrée sous plusieurs formes et l'usage systématique des traces pour lier des crimes est une des originalités du système (Ribaux & Birrer, 2008). Le tout est intégré dans un système de prise de décision se fondant sur le renseignement produit par l'analyse quotidienne des données issues d'activités criminelles (Aepli, Ribaux, & Summerfield, 2011).

Grâce à ces développements, le cadre de travail intègre maintenant plusieurs dimensions en bénéficiant des progrès de la recherche dans plusieurs disciplines (la criminologie environnementale, le renseignement forensique ou les méthodes computationnelles). Il a été développé étape par étape, en évitant de sauter d'une nouveauté technologique à une autre, mais en s'ouvrant à de nouvelles perspectives. La croissante complexité du dispositif et des processus qui a résulté de cette approche a été progressivement absorbée par l'organisation et ses acteurs, des managers aux analystes. Cette complexité ne résulte pas d'une exploitation de technologies sophistiquées, mais plutôt d'un assemblage de composants simples et faciles à adapter. Le champ d'application du dispositif n'a pas cessé de croître depuis son initialisation. De même, son exploitation dans une zone géographique plus large est maintenant sérieusement examinée.

Mais de nouveaux défis se présentent avec l'émergence du *predictive policing*. La question d'informatiser un composant prédictif du système est désormais concrètement posée à l'aune de fortes pressions commerciales, publiques et politiques. Face à la multitude de produits et approches, les praticiens se sentent une fois encore submergés. Cependant, au lieu de considérer le remplacement de l'ensemble de la méthodologie déployée auparavant, notre approche vise l'intégration réaliste de méthodes computationnelles à l'ensemble de la méthodologie déjà existante, sans en détruire ses composants efficaces.

3.3. La détection de patterns dans les tendances des activités criminelles

Le fonctionnement du CICOP s'inscrit dans une approche de l'action de sécurité, qui souhaite évoluer vers davantage de proactivité. Le modèle policier le plus abouti qui exprime cette volonté est celui de la police guidée par le renseignement (en anglais : *intelligence-led policing*) dont les facettes essentielles ont été synthétisées par Ratcliffe (2016). Développé à l'aune de cette approche, le modèle des 4P de Ratcliffe (2011) situe le *pattern* comme la brique élémentaire sur laquelle se fonde l'analyse des formes

répétitives de crimes. Selon ce modèle, l'objectif idéal est de *Prévenir* le crime en interprétant et exploitant les informations pertinentes. Toutefois, cela n'est possible qu'en développant une attitude *Proactive* qui ne peut être mise en œuvre que si le phénomène criminel considéré présente une certaine *Prédictibilité*. Cette dernière repose entièrement sur l'existence de *Patterns,* que l'on peut définir sommairement par des schémas détectables reflétant les comportements criminels dont on peut supposer une réplication ultérieure[16].

La détection de patterns encore inconnus ou la reconnaissance d'un type de pattern connu constitue des enjeux centraux de l'analyse des crimes répétitifs. C'est souvent une fois qu'un pattern est découvert que les analystes se rendent compte que l'activité se déroulait déjà bien avant en recherchant rétrospectivement les cas similaires. Cette détection est probablement l'opération la plus difficile de l'ensemble du processus de résolution de problèmes. Afin d'attirer l'attention sur ces changements dans la structure des données qui peuvent signifier l'apparition de (nouvelles) formes de criminalité ou la disparition d'autres, la recherche systématique de ruptures est susceptible de permettre plus facilement la détection de ces patterns.

Un des objectifs définis par l'unité d'analyse du CICOP est précisément la détection de patterns de ruptures dans les tendances des activités criminelles en utilisant les méthodes computationnelles et en les combinant avec la méthodologie existante. Le but de cette opération est de détecter par des ruptures de « □tendances□ » un nouveau problème ou une évolution significative d'un problème déjà connu. Ces ruptures sont supposées résulter en particulier de l'engagement ou du désistement de groupes d'auteurs, des changements dans leur mobilité, de leur passage dans une région, des types d'opportunités qui les intéressent, de leurs modes opératoires, ou des types de traces qu'ils transfèrent. Cette opération n'est pas triviale, car les données agrégées ne rendent pas immédiatement visibles ces changements.

Il est important ici de distinguer les patterns en deux niveaux, puisque les patterns de rupture qui sont détectés sont susceptibles d'alerter sur la possibilité de détecter de nouveaux patterns criminels ou une modification de patterns existants. Les augmentations et les diminutions dans les données provenant de différentes sources indiqueraient de telles nouveautés (Figure 5).

[16] Pour plus de précision sur la définition du pattern, voir le chapitre 5.

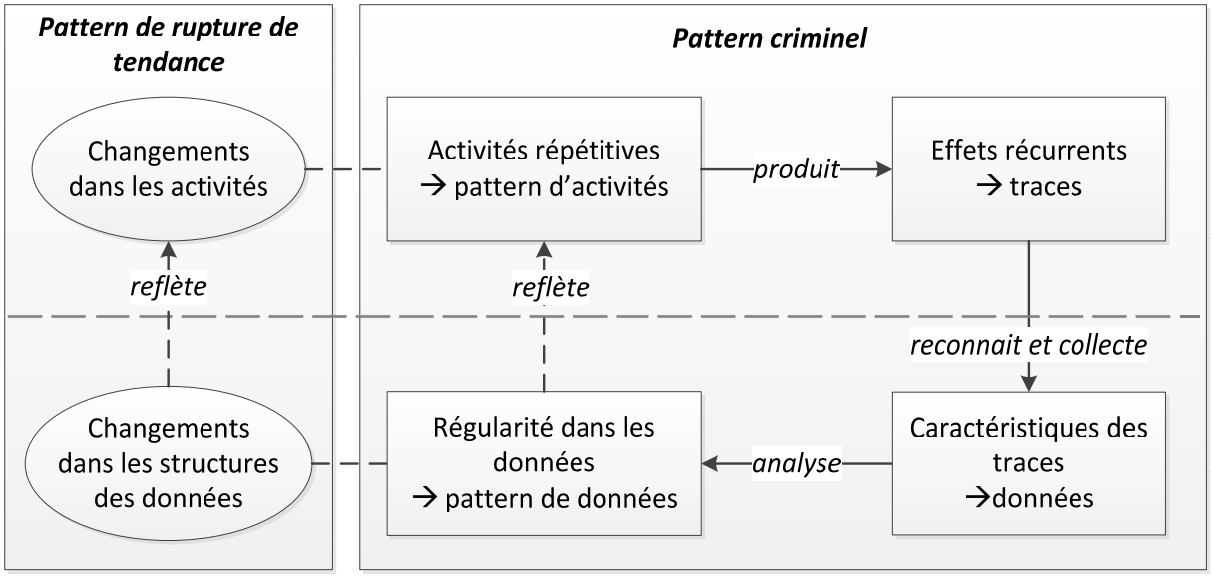


***Figure 5** : Hypothèse du processus de détection des ruptures dans les tendances des activités criminelles : de la même manière que les patterns identifiés dans les données reflètent les patterns des activités criminelles, les changements dans les structures des données reflètent potentiellement des changements dans les activités.*

3.4. Hypothèses de recherche

Comment intégrer de manière pragmatique les méthodes computationnelles dans les processus d'analyse criminelle préexistants en considérant un cadre de travail interdisciplinaire puisant à la fois dans la criminologie et la science forensique et orienté sur la résolution de problème ?

Cette question étant évidemment vaste, cette thèse se restreint à proposer des pistes pour une telle méthodologie en considérant 2 objectifs principaux :

- L'expression d'une approche méthodologique interdisciplinaire en renseignement criminel en se fondant sur le développement d'une unité d'analyse criminelle particulière qui a itérativement intégré et harmonisé ses méthodes et outils ;

- L'amélioration du système opérationnel existant via l'intégration d'un composant computationnel dans la détection de tendances des activités criminelles au sein des processus de renseignement de l'unité d'analyse considérée.

Considérant ces délimitations, l'hypothèse principale de ce travail est la suivante :

- Il est possible de détecter des changements dans des patterns d'activités (nouveaux patterns, évolution de patterns déjà connus, disparition d'un pattern connu) dans la distribution spatio-temporelle des données de la criminalité en détectant les patterns de rupture par des méthodes computationnelles appliquées systématiquement dans l'environnement particulier considéré (unité d'analyse).

Cette proposition est complétée par 3 hypothèses spécifiques, à savoir :

- Les patterns dans les données reflètent les patterns dans les activités criminelles.

- Une détection automatique est susceptible de rendre plus complète, plus rapide et plus précise, la détection de problèmes par les analystes de l'unité d'analyse considérée.

- Une classification situationnelle des événements inspirée par les approches environnementales en criminologie est appropriée pour encadrer et guider cette détection.

C'est à travers ce focus sur l'intégration d'un composant computationnel dans les processus existants qu'est esquissé un cadre de travail interdisciplinaire combinant la science forensique, la criminologie et les sciences de l'information susceptible de guider plus généralement l'implantation d'un système d'analyse criminelle.

Afin d'éprouver ces hypothèses, le cadre de travail théorique qui a influencé l'évolution progressive de l'unité d'analyse du CICOP est exposé à la partie suivante. Son niveau d'expression actuel reste néanmoins insuffisant. Malgré plusieurs travaux théoriques portant sur, d'une part, l'engagement de la criminologie environnementale pour justifier le modèle de classification des données utilisé (Birrer, 2010), et d'autre part, sur l'expression formelle d'une série d'aspects forensiques (Ribaux & Margot, 2003; Ribaux, Walsh, & Margot, 2006), l'intégration des théories criminologiques, forensiques et computationnelles pour fonder une méthodologie holistique n'est pas encore achevée.

La partie suivante vise ainsi à faire avancer le travail d'expression et d'intégration des processus en analyse criminelle tout en se concentrant sur un élément charnière qui ouvre une connexion vers les méthodes computationnelles : le pattern. Cet effort restera quoi qu'il en soit encore insuffisant pour atteindre notre objectif : une série de processus appliqués quotidiennement dans l'unité d'analyse restent encore largement tacites et demanderont à être mieux exprimés afin de pouvoir situer l'apport potentiel des méthodes computationnelles et isoler quelques tâches prometteuses où les appliquer.

PARTIE II : VERS UNE MÉTHODOLOGIE INTÉGRATIVE ET ITÉRATIVE

Afin d'aborder le premier objectif de cette thèse, à savoir l'expression des processus d'analyse et renseignement au sein d'une démarche méthodologique interdisciplinaire, il convient de s'interroger sur le rôle des différentes disciplines impliquées dans ces processus. La science forensique étudie les traces, vestiges d'un crime. Leur mise en relation permet de détecter des répétitions : par exemple, le même profil ADN extrait de traces biologiques recueillies sur des scènes d'événements différents indique l'activité d'un même auteur. Plus généralement, la comparaison systématique de la variété des traces recueillies sur chaque événement s'intègre dans un processus de découvertes de relations entre des activités criminelles, que nous appelons « veille opérationnelle ». Parallèlement, les approches dites « situationnelles » en criminologie expliquent l'existence de crimes répétitifs et permettent d'en distinguer les formes et spécificités. Elles servent de fondements pour des méthodologies qui visent à détecter ces répétitions dans des grands jeux de données. L'intégration de ces approches criminologiques avec les dimensions forensiques précise alors la méthodologie existante. Cette dernière est effectivement utilisée dans l'unité d'analyse que nous ciblons (le CICOP), mais elle est exprimée de manière incomplète en regard de notre objectif : identifier le potentiel d'un troisième composant, à savoir les méthodes computationnelles. Cette partie contribue ainsi à exprimer cette articulation au travers de l'expression de la méthodologie utilisée par le CICOP (chapitre 4) et d'une définition de la notion de pattern (chapitre 5) qui jouera le rôle de charnière entre les trois approches (computationnelle, forensique, et criminologique) (chapitre 6).

4. Fondements théoriques et méthodologiques : de la trace au pattern

L'unité de renseignement du CICOP opère en se fondant sur un postulat fondamental en analyse criminelle : les activités litigieuses suivent des patterns susceptibles d'être détectés et analysés à l'aide des données disponibles. Ce postulat n'est pas aussi évident qu'il semble l'être au premier coup d'œil. Afin de comprendre comment les patterns peuvent être considérés comme l'objet pivot rassemblant les différentes approches en analyse criminelle, il est pertinent de décomposer le raisonnement et sa méthodologie globale.

Un point de départ s'impose en considérant la plus élémentaire des pièces disponibles en termes de données. Il s'agit des résultats produits par un crime ou un comportement problématique et se déposant sur un substrat physique et/ou numérique. Ce vestige physique et numérique d'une activité litigieuse, reconnu et collecté sur les scènes de crime, est la trace (Ribaux, 2014). La trace étant l'objet d'étude de la science forensique (Margot, 2011), nous pouvons alors nous demander dans quelle mesure les modèles en science forensique sont susceptibles de soutenir l'analyse criminelle.

4.1. Science forensique et analyse criminelle : de l'utilisation de la trace

Un crime, ou une activité litigieuse, perturbe l'environnement physique immédiat de manière inhabituelle et cause ainsi la création de traces de différentes natures selon le principe de Locard (1920). Souvent réduit à la formulation « tout contact laisse une trace », ce postulat présente en réalité une portée plus large (Crispino, 2006). Ainsi, la logique de l'investigation débute par la reconnaissance de traces pertinentes sur la scène de crime comme des signes de l'activité inhabituelle. Ces traces sont ensuite collectées, observées et analysées. Enfin, l'activité est finalement reconstruite à l'aide d'un processus d'interprétation (Inman & Rudin, 2000; Ribaux, 2014). Ce processus est cependant imparfait car il se fonde sur des données incomplètes et incertaines. Tous les événements ne sont pas rapportés ou considérés sous une loupe forensique. De plus, le processus de collecte de données ne garantit pas, si une trace existe, qu'elle sera nécessairement reconnue et collectée. Il est alors important de bien distinguer cette transition qui permet de passer de la réalité à la reconstruction, car c'est de ce passage que dépendent en grande partie tous les raisonnements et hypothèses élaborés par la suite. La scène de crime qui représente l'étendue spatiale et temporelle réelle de la distribution des traces laissées après une activité litigieuse n'est pas connue entièrement par les investigateurs de scène de crime. Ces derniers délimitent un champ

d'investigation qui par hypothèse est supposé refléter une scène de crime (Ribaux, 2014). De manière analogue, les traces reconnues et détectées ne sont pas forcément connues dans leur entièreté par les investigateurs. Les caractéristiques de la trace qui sont collectées et traitées sous forme de données constituent également un modèle qui est supposé refléter la trace. Par exemple, en détectant une trace de sang, l'investigateur sera susceptible de la collecter en la traduisant sous forme de données de son choix. Il peut la photographier ou l'envoyer au laboratoire pour extraire un profil ADN. Les données ainsi obtenues sont dépendantes d'un choix raisonné de la part des investigateurs qui se veut le plus plausible en fonction de la réalité observée. Cette articulation est exprimée à travers la Figure 6 où une distinction est effectuée entre les entités relevant des faits (la réalité) et celle relevant des raisonnements (la cognition). Cette première étape dans une tentative de formalisation du processus général intégré sera par la suite complétée à chaque nouvel argument apporté.

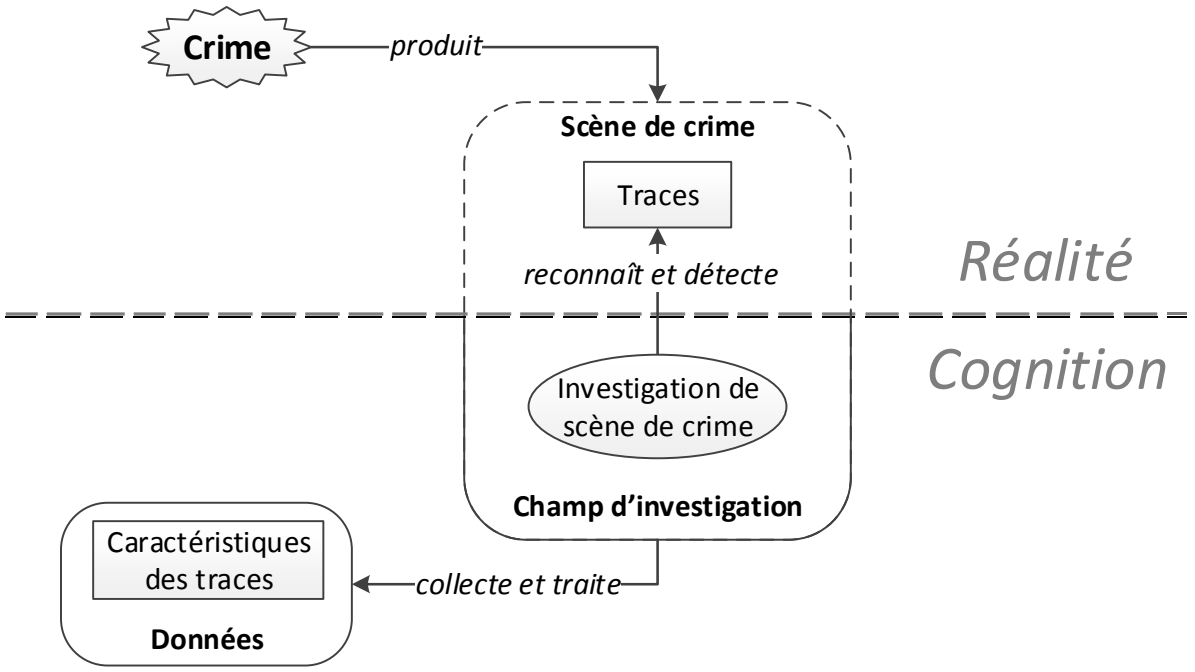


***Figure 6** : 1ère étape de formalisation : la trace comme transition. Un crime, ou une activité, produit des effets sous forme de traces, lesquelles sont reconnues et détectées par l'investigation de scène de crime. Les caractéristiques des traces sont alors collectées et traitées sous forme de données.*

La trace, souvent fragmentaire, peut tout aussi bien être pertinente ou non vis-à-vis de l'activité (p. ex. une trace de soulier causée par le lésé d'un appartement cambriolé). La reconstruction signifie rechercher la cause la plus probable susceptible d'être responsable des effets observés (les traces telles qu'elles ont été observées). Plusieurs hypothèses peuvent ainsi expliquer les mêmes effets observés. Ce processus de

raisonnement est appelé abduction, en référence à la logique de Peirce (Margot, 2011; Peirce, 1935). La Figure 7 schématise l'abduction sous forme de triangle (Crispino, 2008; Margot, 2003; Ribaux, 2014). Ce type de raisonnement est caractéristique des sciences historiques qui, à l'opposé des sciences dites expérimentales, ne permettent que de raisonner sur des événements passés en vue notamment de déterminer la « preuve tangible »[17](Cleland, 2011).

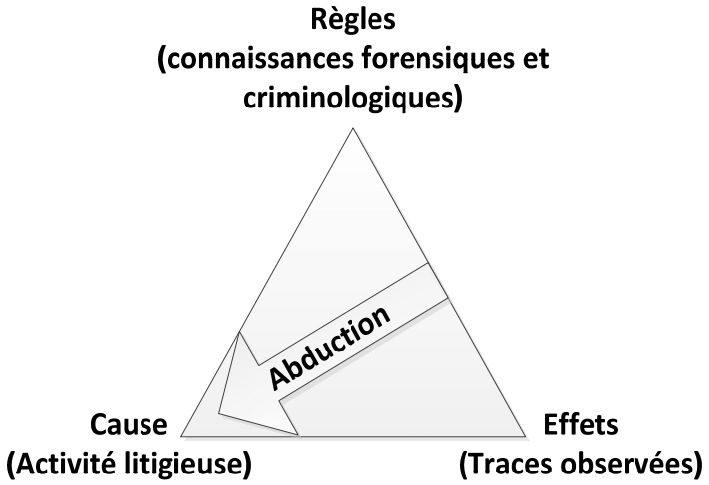


***Figure 7** : La logique d'abduction en science forensique : l'interprétation des traces observées à l'aide de connaissances forensiques et criminologiques permet d'identifier la cause la plus probable et ainsi d'éventuellement reconstruire l'activité litigieuse (inspiré par Crispino, 2008; et Margot, 2003).*

Traditionnellement en science forensique, ce processus de reconstruction vise à fournir des informations susceptibles d'aider une cour de justice à établir les faits, voir à qualifier l'infraction, et à prendre une décision. L'information transmise par ce type de données peut cependant être utilisée au sein de différents processus, notamment en analyse criminelle.

Lorsque des auteurs répètent leurs activités litigieuses, les diverses traces qu'ils transfèrent peuvent présenter des similarités qui indiquent une source commune, même si le mode opératoire diffère. Plus généralement, une grande variété de répétitions criminelles peut être détectée à l'aide de la comparaison et de l'interprétation des traces collectées sur différentes scènes de crime (par exemple, images, traces de souliers, profils ADN).

[17] « *The smoking gun* »

La mise en relation systématique d'événements criminels par la comparaison de traces contribue à l'analyse criminelle lorsqu'il s'agit de détecter les activités sérielles d'auteurs ou de groupes d'auteurs identiques (Ribaux, 1997; Ribaux, Taroni, & Margot, 1995). L'information sur les problèmes considérés porte sur différentes dimensions. Par exemple, des profils ADN correspondants peuvent indiquer la présence d'une seule personne sur différentes scènes de crime. Sur une autre dimension, des processus de fabrication similaires peuvent être inférés à partir de caractéristiques correspondantes extraites de faux documents d'identité détectés et saisis, par exemple lors de contrôles en rue ou à la frontière (Baechler, 2015; Baechler, Ribaux, & Margot, 2012).

Cette approche, appelée renseignement forensique, peut être exprimée comme un processus itératif de veille opérationnelle (Figure 8)(Baechler, Morelato, et al., 2015; Morelato et al., 2014; Ribaux, Genessay, & Margot, 2011).

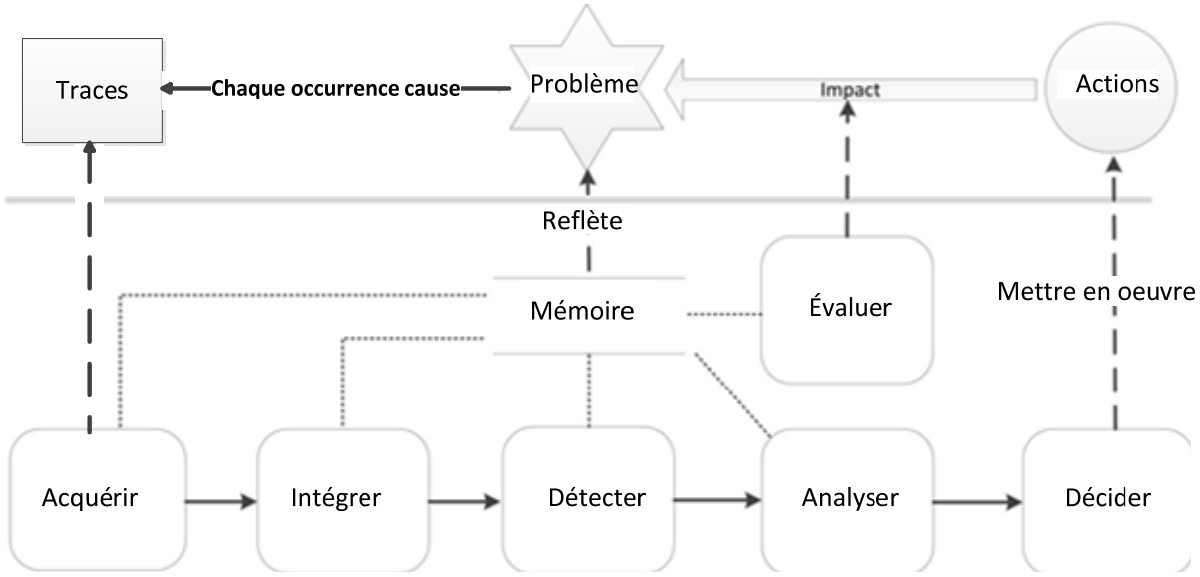


***Figure 8** : Le processus de veille opérationnelle (Ribaux, 2014) : le processus de veille est conçu comme étant la collection de traces, l'intégration de l'information, la détection de patterns, l'analyse et la diffusion du renseignement*

Ce genre de processus trouve une expression plus générale, et peut être considérés comme une « forme primaire » du renseignement (Hane, 2015). L'association française de Normalisation (AFNOR) définit un système de veille comme un « ensemble structuré réunissant les compétences répondant [à une] activité continue et en grande partie itérative visant à une surveillance active de l'environnement technologique, commercial, etc., pour en anticiper les évolutions » (AFNOR, 1998, p. 6).

La plupart des descriptions proposées s'accordent sur des étapes clés dans les processus de veille (Bloch, 1999; Hane, 2015; Rouach, 2010) :

- La collecte des données.
- L'intégration dans une mémoire.
- L'analyse des données.
- La prise de décision.
- L'évaluation.

Dans cette vision, l'étape de détection n'est pas distinguée de l'étape d'analyse. Notre système de veille en renseignement forensique se révèle donc un peu plus spécifique, en différenciant clairement ces deux étapes. En ce sens, il s'inspire des processus analogues qui relèvent de l'action de sécurité et de la résolution de problèmes. Par exemple, la méthode SARA (Clarke & Eck, 2005, étape 7), probablement la plus pratiquée en milieu policier, décompose quatre étapes :

1) La détection (*Scanning*) qui permet de découvrir des problèmes spécifiques.
2) L'analyse (*Analysis*) qui cherche à les définir, en exprimer les contours et à comprendre leurs causes.
3) La réponse (*Response*) qui se traduit par la recherche et le choix de solutions pour supprimer ces causes et/ou réduire de manière persistante les effets désagréables des problèmes identifiés.
4) L'évaluation (*Assessment*) des actions entreprises pour déterminer leurs impacts sur les problèmes.

Bien que ce processus soit exprimé à un niveau très général, il s'avère en pratique très utile. Les quatre étapes contrebalancent la tendance naturelle de chercher une réponse immédiate à une situation, en s'épargnant de définir et d'analyser le problème en profondeur, et en négligeant d'assurer un suivi et de mesurer l'efficacité des mesures choisies. Nous discuterons plus loin en profondeur ce lien entre la résolution de problèmes et le processus de veille en renseignement forensique, afin d'enrichir ce dernier.

En regard de notre problématique, il manque encore dans les systèmes de veille, qui connaissent un fort développement dans le domaine économique (Hane, 2015), une mention plus explicite au renseignement. Ainsi, au-delà de la détection et de l'analyse

des répétitions criminelles, le processus de veille opérationnelle en renseignement forensique, tel qu'illustré à la Figure 8, poursuit également les objectifs suivant :

- Cibler l'exploitation des ressources en fonction du renseignement accessible
- Assurer la qualité des produits de renseignement en vue de leur traduction dans des mesures concrètes.

Ce type de systèmes nécessite un fonctionnement simple, rapide et suffisamment flexible pour s'adapter à l'évolution de la criminalité tout en optimalisant l'exploitation des ressources.

Diverses expériences ont démontré le potentiel de l'approche, en se basant sur différents types de traces (p. ex. traces de souliers, profils ADN, profils chimiques de substances, projectiles, images), afin de suivre différents types d'activités criminelles répétitives (p. ex. crimes de haut volume, l'utilisation de faux documents, la production et la distribution de médicaments contrefaits) (Baechler, 2015; Dégardin, Roggo, & Margot, 2015; Lammers, 2013; Ribaux, Crispino, Delémont, & Roux, In press; Rossy, Ioset, Dessimoz, & Ribaux, 2013).

La mise en œuvre de ces processus forensiques rencontre beaucoup d'obstacles. Nourrir l'analyse des problèmes de nature criminelle avec des traces de diverses natures est une entreprise hardie. Conséquence de la spécialisation et du focus sur la cour de justice, les organisations relevant de la science forensique promulguent le traitement séparé de chaque type de traces au sein de structures spécialisées. L'information est ainsi fragmentée la plupart du temps et n'est pas directement accessible aux analystes criminels. Dans des structures judiciaires relativement petites et généralistes, p. ex. dans un pays fédéraliste comme la Suisse, ce problème n'est pas insoluble car les scientifiques forensiques sont fortement intégrés dans les investigations et demeurent proches de l'analyse criminelle. Dans d'autres types d'organisations, la fragmentation des processus rend ce genre d'approche beaucoup plus difficile (Ribaux, Crispino, & Roux, 2015).

L'intégration de différents types de traces dans les processus de comparaison n'est pas à négliger dans la production de renseignement criminel : certains cambrioleurs laissent des traces de soulier qui permet leur détection, alors que d'autres, à travers un certain degré de spécialisation, laisseront plutôt de manière répétée des traces d'oreilles sur les portes ou des traces biologiques à partir desquelles des profils ADN

comparables seront extraits. De plus, la diversité de l'information issue de la comparaison de traces (p. ex. une même personne présente lors de plusieurs événements, un processus de fabrication de faux documents similaire, ou des types d'objets similaires utilisés par l'auteur) requiert un certain niveau de contextualisation pour l'intégrer dans un processus d'analyse criminelle. La manière dont ce processus s'articule autour des traces est exprimée à travers la Figure 9.

Ainsi, la trace est produite par une activité, ou comportement, qui est généralement l'élément d'intérêt des processus d'analyse criminelle. La relation entre la trace et le comportement, ou mode opératoire, demande cependant à être approfondie afin d'exprimer le composant forensique au sein d'une architecture intégrée d'analyse criminelle.

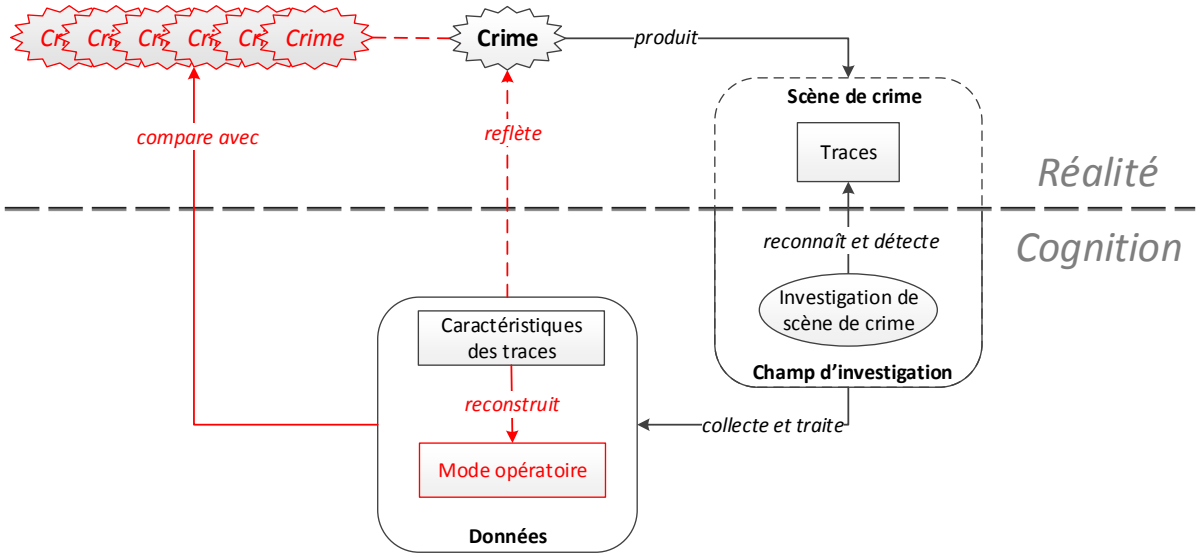


***Figure 9** : 2ème étape de formalisation : l'utilisation de la trace. Parmi l'ensemble des événements criminels, une activité spécifique produit des effets sous forme de traces, lesquelles sont reconnues et détectées par l'investigation de scène de crime. Une fois collectées et traitées, les caractéristiques des traces se retrouvent sous forme de données permettant de reconstruire le mode opératoire, reflétant ainsi l'activité, ou encore d'effectuer une comparaison avec d'autres événements afin de les lier à une même source.*

4.2. Entre traces et activités : l'intégration de la criminologie environnementale

L'analyse criminelle gère traditionnellement l'information liée aux situations (convergence de cibles adéquates pour des auteurs spécifiques, à des endroits particuliers et à certains moments). Afin d'y parvenir, l'analyse criminelle prend en considération les informations spatio-temporelles et les modes opératoires qui ont été reconstruits à partir des observations sur la scène de l'événement, c.-à-d. à partir des traces. Cette forme de reconstruction constitue un processus de base, mais l'intégration

de la science forensique en analyse criminelle peut cependant aller plus loin en termes d'inférences.

Lorsqu'il se rend sur une scène, l'examinateur de scènes de crime interprète sommairement une situation criminelle. Il imagine comment l'auteur a agi, quel était son mode opératoire. Il doit « penser comme un voleur »[18] afin d'interpréter la situation (Clarke & Eck, 2005, step 10) et guider l'examen de la scène. Cette forme d'inférence se trouve renforcée par la considération de l'environnement physique et social : quels étaient les obstacles/gardiens ? Que dire à propos de la configuration des lieux lorsque le problème a surgi ? Quelles étaient les cibles ? Peut-on imaginer le processus de décision de l'auteur dans une situation particulière ? Comment interpréter les situations afin de reconnaître et collecter des traces, ainsi que d'expliquer les traces observées ? En résumé, quelles sont les connexions qu'entretient l'investigation de scène de crime avec les théories issues de la criminologie environnementale ? Cette question est notamment investiguée par plusieurs chercheurs en science forensique (Ribaux, Baylon, Roux, et al., 2010; Ribaux, Crispino, Delémont, et al., 2015; Schuliar & Crispino, 2013).

Dans le but de comprendre la séquence d'événements qui constitue une activité litigieuse, Cornish (1994) propose le concept des « scripts ». Cette approche part du principe qu'une catégorie particulière de crimes nécessite un ensemble d'actions standards qui doivent être réalisées dans un ordre bien précis, comme dans le script d'une pièce de théâtre. Les scènes représentent les différentes étapes d'un crime, le casting est constitué des auteurs, des victimes et des témoins ; et les outils utilisés sont les accessoires. Par exemple, dans le cas d'un cambriolage de villa avec une chignole durant la nuit, le malfaiteur doit suivre une série d'étapes nécessaires à la réalisation de son méfait. Il doit se procurer un véhicule (souvent volé) pour se rendre sur les lieux ; il doit se procurer une chignole ; une fois sur place il doit trouver une voie d'accès, percer un trou à l'aide de la chignole dans le montant d'une fenêtre ou d'une porte-fenêtre, souffler dans le trou pour dégager la sciure produite, insérer une tige métallique pour ouvrir de l'intérieur, sans bruit (ce genre de mode opératoire est utilisé durant la nuit). Autant d'étapes nécessaires alors que le cambrioleur n'est même pas encore entré dans la villa. La décomposition en scripts permet d'orienter la recherche de traces (Ribaux, Baylon, Lock, et al., 2010). Dans notre exemple, le fait que le

[18] « Think like a thief »

malfaiteur doive souffler à l'intérieur du trou percé pour y insérer la tige métallique est susceptible de provoquer un transfert de traces biologiques dans le trou créé par la chignole. Un profil ADN peut éventuellement résulter de l'analyse d'un prélèvement. Devant ce type particulier de cambriolage, les investigateurs de scène de crime savent donc comment détecter ces « points de contacts » supposés (Barclay, 2009) désignant où aller rechercher des traces en priorité.

Dès lors que l'analyse criminelle se fonde sur les théories issues de la criminologie environnementale, concernées par le substrat physique sur lequel le crime survient (Boba, 2009; Felson & Clarke, 1998), un lien se cristallise entre les disciplines : les traces sont reconnues à l'aide des situations criminelles. À leur tour, elles aident à les interpréter et à détecter des liens entre les événements reflétant la nature sérielle et concentrée de ces formes de criminalité. Une troisième étape est alors possible dans notre essai de formalisation du processus en intégrant la notion de situations criminelles (Figure 10). Basées sur ces postulats, la collecte et l'intégration de données forensiques procurent une solide base d'information pour les processus d'analyse criminelle. Cependant, ce lien demande à être approfondi, notamment en partant d'un objet pivot en analyse criminelle. Si la trace en est la donnée la plus élémentaire, la découverte de « patterns » est son objectif le plus fondamental.

La notion de pattern tel que mentionnée ici soulève néanmoins toute une série d'interrogations. Qu'est-ce qu'un pattern exactement ? Quelle est sa définition ? Est-ce que ces patterns existent ? Et si oui, ont-ils un sens ? Quelle est leur nature ? Sont-ils détectables et comment ? Peut-on les utiliser afin d'élaborer et planifier une réponse à un problème ? Autant de questions à investiguer afin de construire un cadre de travail interdisciplinaire articulé autour des patterns criminels.

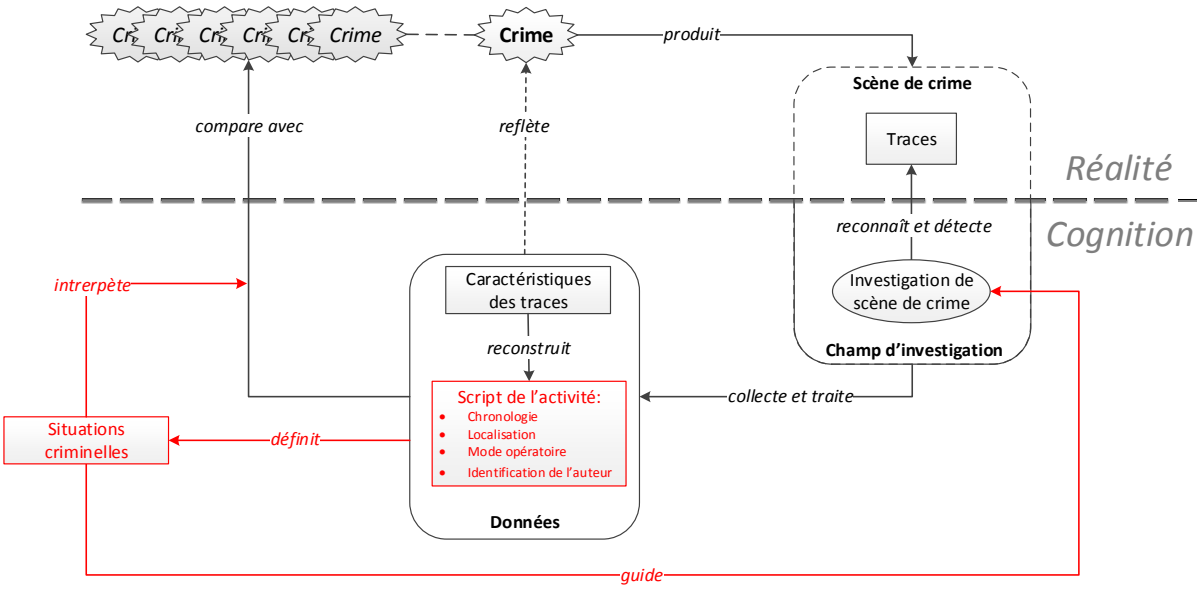


***Figure 10** : 3ème étape de formalisation : l'intégration des situations criminelles. L'analyse des traces collectées sur une scène de crime permet de reconstruire l'activité délictueuse, notamment sous forme de script. La détermination de la chronologie des événements, de leurs localisations, de leurs modes opératoires et de leurs auteurs potentiels peut mener à définir des situations criminelles spécifiques. Ces situations guident les futures investigations de scène de crime et aident à interpréter et détecter les liens entre les événements.*

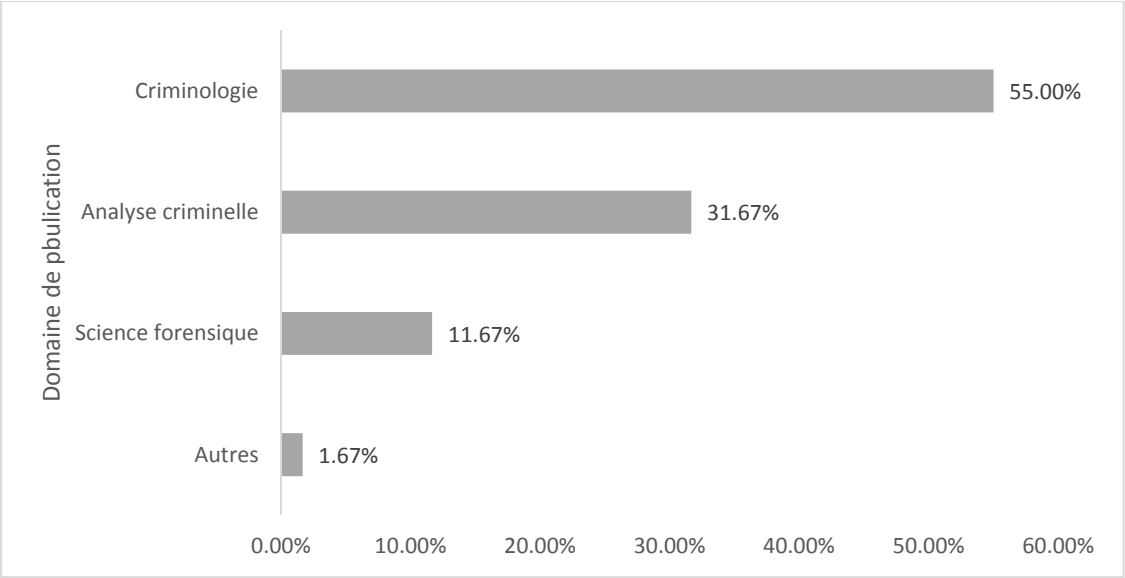
5. Le pattern : le définir, le détecter, l'utiliser

5.1. Définition

Définir ce qu'est un pattern n'est pas aussi évident qu'on pourrait l'imaginer. Il existe presque autant de définitions que de disciplines dans lesquelles cette notion est utilisée et sa traduction française est tout aussi problématique. Si l'on se réfère au Bureau de la traduction des services gouvernementaux du Canada, le mot « *pattern* » n'aurait pas moins de 52 traductions différentes selon le contexte d'utilisation[19]. Dans un souci de simplicité et de clarté, nous ne nous essaierons pas à l'exercice de traduction et emploierons le terme pattern tout au long de cette thèse. En revanche, il apparaît primordial de définir précisément ce que nous entendrons par « pattern » en relation avec notre problématique.

Dans le domaine des sciences criminelles, la notion de pattern est utilisée régulièrement, mais son sens peut varier fortement selon le contexte (Baud, 2015). En analyse criminelle, le pattern fait référence au caractère répétitif du comportement d'un auteur qui peut être visualisé dans les données collectées (Phillips & Lee, 2012; Rossy, 2011; Rossy & Ribaux, 2014). Il existe cependant d'autres situations où le pattern désigne une anomalie, un élément qui se distingue de la normalité (Kedma, Guri, Sela, & Elovici, 2013). C'est au regard de ces notions de répétition et d'anomalie que peut s'opérationnaliser la détection du pattern. Dans son côté plus forensique, le pattern peut également désigner les similitudes entre les traces collectées sur des événements distincts, permettant de formuler l'hypothèse d'une source commune ou d'un mode opératoire similaire. Cependant, l'étude exploratoire de Baud (2015) sur l'utilisation du pattern en sciences criminelles tend à montrer que les questions liées au pattern criminel sont moins appréhendées par la science forensique, où on s'intéresse plutôt aux cas particuliers, que par la criminologie qui recherche préférentiellement des schémas généraux (Graphique 5). Le pattern peut également avoir une connotation plus technique, par exemple les patterns de distribution de résidus de tirs ou de traces de sang et les patterns de configuration des fractures en médecine légale (Baud, 2015).

[19] Source : http://www.btb.termiumplus.gc.ca/tpv2guides/guides/chroniq/index-fra.html?lang=fra&lettr=indx_titls&page=9MwP2LfKJznE.html



***Graphique 5** : Fréquence des articles comprenant les termes « crime » et « patterns » en fonction du domaine de publication pour les 60 premiers résultats (n= 60) Source : Scopus et ScienceDirect.*

Ainsi, certaines notions spécifiques au concept semblent se dégager des différentes itérations possibles : la représentation schématique d'une réalité et la répétitivité. La définition trouvée dans le Larousse[20] se trouve être particulièrement adaptée au contexte des sciences humaines et sociales. Celle-ci stipule qu'un pattern est un :

« Modèle spécifique représentant d'une façon schématique la structure d'un comportement individuel ou collectif. »

Bien que la notion de représentativité de la réalité soit présente, il manque cependant la notion de répétitivité qui est pourtant cruciale dans l'analyse en fonction de la nature sérielle et concentrée d'une grande partie de la criminalité. De plus, l'analyse criminelle ne s'intéresse pas à n'importe quels patterns, mais précisément à ceux qui sont liés à des comportements problématiques. Ainsi, nous proposons une définition plus fine et adaptée à partir de la précédente au contexte de l'analyse criminelle.

Le pattern criminel peut donc être défini comme un modèle spécifique représentant d'une façon schématique les récurrences et les anomalies dans la structure d'un comportement litigieux individuel ou collectif.

Néanmoins, ces formes de patterns existent-elles ? Et si cela est le cas, comment le justifier d'un point de vue théorique ? Un élément de réponse se situe au sein des théories criminologiques.

[20] Source : http://www.larousse.fr/dictionnaires/francais/pattern/58738

5.2. Existence de patterns pertinents dans les données

L'analyse criminelle se fonde sur l'existence de patterns de données reflétant les régularités dans les comportements criminels et l'expérience tend à montrer leur validité quotidiennement. Cependant, ce postulat peut être nuancé et mieux délimité : la nature des données disponibles et les processus de reconstruction incertains semblent limiter drastiquement les attentes issues de la fouille des données.

Les théories des opportunités, ou approches situationnelles, en criminologie peuvent aider à situer la discussion autour de cette question, notamment grâce à leur expression pragmatique qui se traduit par le modèle du triangle du crime (Clarke & Eck, 2005).

Cet ensemble de théories s'articule autour des *activités routinières* de Cohen et Felson (1979). Pour qu'un crime se produise, une convergence dans le temps et l'espace est nécessaire entre un auteur motivé, une cible propice et une absence de gardien capable. La combinaison de ces composants caractérise un acte criminel comme une situation spécifique, ou « opportunité », qui distingue des schémas susceptibles de s'imprimer dans les données sous la forme de patterns criminels.

Contrairement aux activités routinières de Cohen et Felson où la motivation de l'auteur à commettre un crime est supposée constante, la théorie du *choix rationnel* (Cornish & Clarke, 1986; Felson & Clarke, 1998) défend quant à elle le postulat que l'auteur motivé est un être rationnel capable de s'adapter aux situations. Sa décision de commettre un crime dépend de l'environnement immédiat et de son analyse coût-bénéfice de la situation. Ainsi, les auteurs sont motivés par les bénéfices perçus à court terme et espèrent que ces derniers surpasseront les éventuels inconvénients susceptibles de découler de leurs actions (Cusson, 2005). La rationalité de l'auteur influence ainsi également la configuration spécifique des opportunités et prend part à la caractérisation des patterns criminels.

Finalement, la théorie des *patterns criminels* (P. J. Brantingham & Brantingham, 1978, 1981; P. L. Brantingham & Brantingham, 1993) explique comment les auteurs motivés rencontrent des cibles propices. Durant ses activités quotidiennes et licites, un auteur potentiel peut traverser des zones cibles qui offrent des opportunités criminelles. Ce sont surtout des espaces d'activités où les personnes peuvent vivre, travailler, s'amuser. Pour un auteur motivé, il est plus simple de commettre des crimes sur son trajet journalier, plutôt que d'effectuer un trajet spécial pour le faire.

Le modèle du triangle du crime intègre les divers éléments des théories mentionnées et appuie la compréhension et l'interprétation des patterns (Clarke & Eck, 2005). On retrouve les trois pans qui représentent l'auteur, la cible/victime et le lieu, et leur rencontre sous la forme de l'opportunité criminelle (Figure 11). Chacun de ces composants peut être influencé par l'un des éléments formant un triangle extérieur : l'éducateur pour l'auteur, le gardien pour la cible/victime et le responsable pour le lieu. Ces configurations sont connues pour être très spécifiques et distinctes les unes des autres car le moment et l'endroit pour réaliser l'activité sont restreints, de même que le mode opératoire potentiel utilisé par l'auteur dans le contexte immédiat de l'acte. Cette famille de théories criminologiques rassemblées dans le triangle converge ainsi vers certaines structures connues et délimitées pour certains types de crime. En conséquence, il peut être attendu que ces comportements fortement contraints impriment de manière répétée des caractéristiques comparables dans les données collectées, en particulier dans les traces. La troisième étape de formalisation du processus général résume l'apport des approches situationnelles en criminologie concernant la notion de patterns (Figure 12).

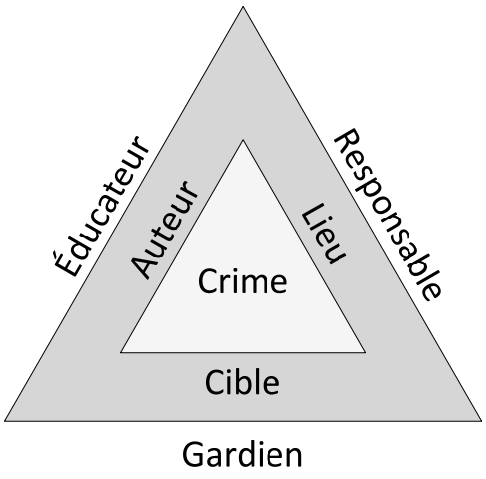


***Figure 11** : Le triangle du crime (Clarke & Eck, 2005)*

Il est possible de comprendre la manière dont se traduisent les patterns d'activités dans les situations et la manière dont ils sont ensuite imprimés dans les données à l'aide de deux exemples observés en Suisse romande au sein de l'unité d'analyse et de renseignement du CICOP.

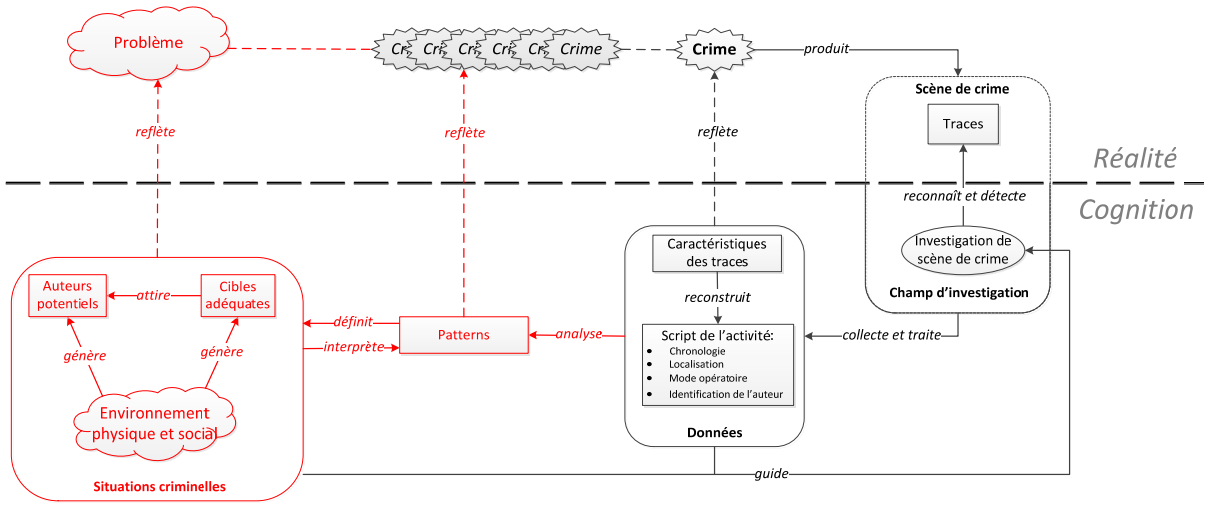
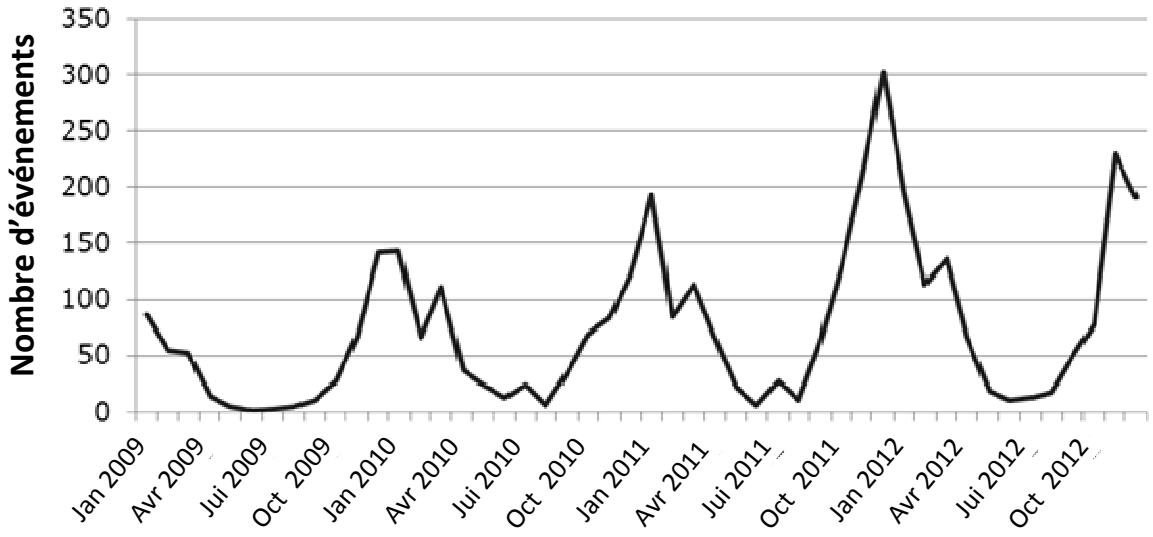


Figure 12 *: 4ème étape de formalisation : L'apport des approches situationnelles en criminologie. L'existence de patterns dans les données est justifiée par les approches situationnelles, notamment sous l'égide du triangle du crime. Ces théories criminologiques expliquent la présence d'opportunités criminelles susceptibles de se répéter et, par conséquent, d'imprimer des patterns particuliers dans les données collectées sur une scène de crime. Ces patterns reflétant l'activité criminelle.*

5.3. La distinction entre pattern d'activité et pattern de données

Il est important de distinguer ce que nous appelons « pattern d'activité » et « pattern de données ». Le pattern d'activité représente les répétitions dans les activités criminelles, par exemple le mode opératoire similaire d'un auteur, la victimisation répétée d'une cible ou la concentration de crime à un endroit particulier (*hotspots*). Le pattern de données est défini par la manière dont ces répétitions sont imprimées dans les données recueillies. À l'instar de la logique de l'investigation qui cherche à reconstruire les causes d'une activité à partir de ses effets (les traces), l'analyse criminelle cherche à représenter les activités répétitives (pattern d'activité) en se basant sur les patterns détectés dans les données collectées (pattern de données). Un exemple concret de pattern d'activité, issue de l'influence des trois pans du triangle du crime, peut être observé à travers une analyse des cambriolages d'habitation se déroulant le soir. Un pattern de données apparaît dans les données examinées par l'unité d'analyse criminelle couvrant la Suisse romande (Graphique 6). L'activité représentée par les données collectées est concentrée sur les périodes hivernales et semble être récurrente, même si l'intensité change selon l'année considérée. Des considérations situationnelles expliquent relativement facilement l'existence de ces patterns d'activité et de données (Birrer, 2010; Maguire, 1982). Durant l'hiver en Suisse, le soleil se couche tôt le soir (aux environs de 17 h). Les occupants des maisons ou appartements ont besoin de lumière pour continuer à vaquer à leurs occupations. Le cambrioleur cherchant une cible potentielle dans une zone résidentielle peut utiliser

l'absence de lumière comme un bon indicateur de sélection. Cela s'inscrit dans un schéma rationnel uniquement réalisable durant les soirées hivernales et seulement dans des zones résidentielles, où un accès discret par la fenêtre est rendu possible.



***Graphique 6** : Évolution des cambriolages d'habitation en soirée en Suisse romande (n= 3'549).*

Cela semble être un classique de l'analyse criminelle en l'occurrence. Cependant, l'analyse spécifique des cambriolages du soir n'est pas aussi facilement réalisable et dépend fortement des structures de base de données dont disposent les analystes : le déroulement exact des événements n'est pas toujours connu avec suffisamment de précision (il est reconstruit) et les bases de données sont généralement conçues sur la base d'autres systèmes de classification plus orientés vers la justice. En revanche, une classification situationnelle des cambriolages d'habitation permet le suivi de ce genre de phénomènes périodiques. Car même si les auteurs peuvent se diversifier dans leurs activités délictueuses, il existe des indices qu'ils restent actifs de manière répétée dans des situations spécifiques (Kempf, 1986). Détecter ces situations spécifiques en subdivisant les données à l'aide de typologies situationnelles semble être ainsi pertinent.

De plus, une telle approche encourage la détection de patterns plus subtils (Birrer, 2010; Sorensen, 2003). En effet, en observant la courbe, une autre régularité se démarque : en février de chaque année, le nombre de cas diminue et augmente après deux semaines. Ce nouveau pattern doit encore être interprété, par exemple en supposant une périodicité dans l'activité d'un groupe d'auteurs spécifiques.

L'existence de certains types de patterns d'activité et de données est ainsi justifiée par les théories des opportunités et il existe de sérieux espoirs de les détecter à l'aide d'une analyse centrée sur certains types d'information (p. ex. concentration de crimes, comparaison systématique de traces ou de modes opératoires). En retour, ces concentrations de crimes offrent du renseignement destiné à mettre en place des stratégies efficientes, tel que le ciblage des auteurs prolifiques (Ratcliffe, 2008), les opérations coup de poing sur des points chauds (Chilvers & Weatherburn, 2001) ou des réponses avec une orientation plus préventive (Clarke, 1997).

5.4. L'évolution temporelle comme pattern pertinent

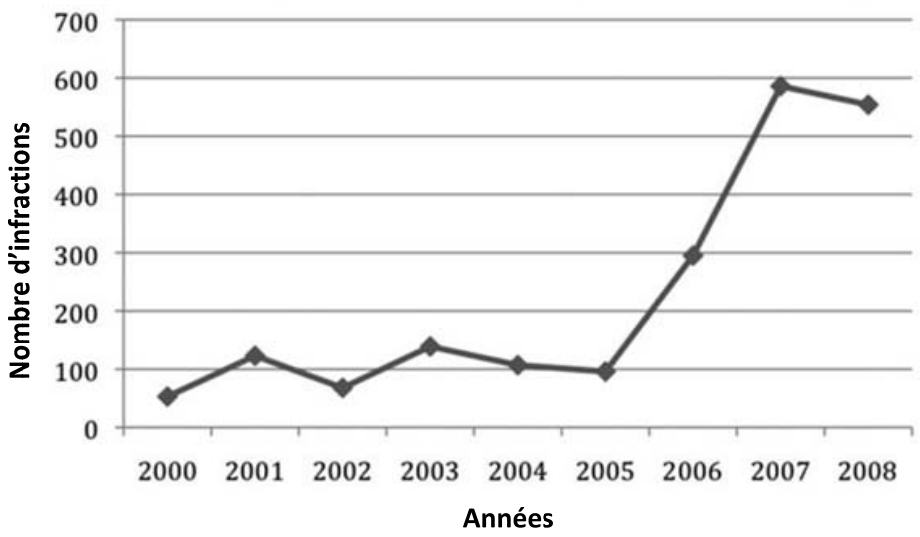
Boba (2009) intègre les concentrations de problèmes de manière simplifiée en utilisant le triangle du crime et le principe de Pareto, aussi appelé règle du 80/20. Ce principe issu des travaux du sociologue et économiste italien Vilfredo Pareto (1896/1965)[21] stipule que 20% des causes seraient responsables de 80% des effets[22]. Traduit en termes de criminalité, une minorité d'auteurs est responsable de la majorité des crimes (Heaton, 2000; Wolfgang, 1987), peu d'endroits génèrent de grandes proportions de crimes (Sherman, Gartin, & Buerger, 1989), et un petit nombre de victimes souffre d'une part importante d'infractions (Weisel, 2005).

Cela justifie l'existence d'une autre famille de patterns particulièrement pertinente en analyse criminelle. Lorsqu'un auteur prolifique, ou un groupe d'auteurs, initient une activité, changent radicalement leur comportement ou stoppent leurs activités, une influence significative est susceptible de se produire dans les données collectées, à cause de cette grande concentration des répétitions criminelles. Plus généralement, lorsque les opportunités (un complexe auteur/crime/espace-temps/cible) changent, la criminalité se modifie significativement, et par conséquent, les patterns dans les données collectées également. Ces changements, ou « nouveautés » doivent être cependant interprétés avec attention et contextualisés afin de tenter d'expliquer leurs causes et de savoir s'ils représentent un pattern d'activité pertinent.

Plus la concentration est élevée, plus les changements dans les opportunités et les activités seront potentiellement détectés et interprétables. L'arrivée des cambrioleurs géorgiens dans les pays d'Europe de l'Ouest et leur activité dans le canton de Vaud en

[21] Initialement, il avait observé que 20% de la population possédait 80% des richesses en Italie.
[22] Il s'agit d'une proportion symbolique ; pour Ratcliffe (2008), la proportion est de 60% des crimes commis par 6% des auteurs.

Suisse (Azzola, 2010) sont une excellente illustration. Ces auteurs prolifiques agissant à l'aide d'un mode opératoire plutôt stable ont soudainement causé une augmentation significative de certains types de cambriolages d'habitations. On le remarque dans les données depuis 2006 (Graphique 7).



***Graphique 7** : Évolution des cambriolages d'habitation par arrachage de cylindre dans le canton de Vaud en Suisse (n= 2'021).*

De manière générale, on s'attend à ce que l'apparition d'auteurs prolifiques agissant selon des situations spécifiques cause un impact significatif dans la répartition statistique du même type de crimes. En revanche, ces genres de changement ne sont pas forcément évidents à détecter parmi les grandes quantités de données agrégées qui portent sur la criminalité que les Anglo-saxons appellent criminalité de haut volume (*high volume crime*)[23]. En effet, dans l'exemple précédent, la rupture semble évidente une fois présentée *a posteriori* avec les dimensions adéquates. Elle n'a cependant pas été détectée immédiatement ou reconnue comme pertinente malgré le niveau important d'activités de ces auteurs. La perception au sein de l'unité de renseignement criminel est que ce pattern aurait dû être détecté, interprété et communiqué plus tôt. Elle ne se restreint pas à cet exemple. Les analystes et policiers sont souvent partagés lors de la détection d'un pattern dans les données, dont il est fait l'hypothèse qu'il s'agit d'une répétition criminelle. D'une part ils sont satisfaits de leur découverte, mais ils constatent aussi fréquemment que ces changements dans les données étaient détectables bien plus tôt. Cette incapacité à détecter ces répétitions ont des conséquences dramatiques lorsqu'elles portent par exemple sur les homicides (Kind,

[23] Par exemple, les cambriolages ou les vols à l'arraché.

1987). Cette incapacité a été même nommée par Egger (1984) « *linkage blindness* » et a conduit au développement d'une multitude d'approches, pas toujours convaincantes et souvent antérieures aux développements de la criminologie environnementale (Chopin, 2017), de comparaison systématique des données.

Considérer d'autres théories criminologiques dans la perspective de délimiter les patterns détectables est évidemment possible. Par exemple, l'étude des interactions sociales permet l'émergence d'autres structures de la criminalité (p. ex. les systèmes de délinquance) avec des actions, compositions, diffusions et durées spécifiques. Il est ainsi possible d'obtenir des indications complémentaires sur les genres de patterns à chercher dans les données collectées (Sutherland & Cressey, 1966; Tremblay, 2010). Il s'agit d'une extension possible afin de compléter une approche intégrée en analyse criminelle, mais qui ne sera pas traité dans le cadre de ce travail. Ce nonobstant, le triangle du crime forme un cadre de travail solide, prenant ses racines sur des théories criminologiques et guidant la détection de concentrations et de répétitions criminelles et l'expression des liens entre les disciplines impliquées en analyse criminelle.

Les méthodes de détection de patterns se répartissent à travers un large panel d'opérations comme l'adaptation de modèle de base de données, le tri, l'exploration et la visualisation des données. Une grande contribution des sciences de l'information est cependant attendue car certains patterns se révèlent évidents et leur détection semble facilement calculable. Parmi les outils les plus basiques, certaines tendances bien marquées peuvent être détectées en définissant des seuils et en comparant les taux de crimes avec les taux des semaines précédentes (Bruce, 2008). Dans d'autres situations, des modèles plus sophistiqués destinés à prédire les répliques sismiques sont adaptés pour suivre la distribution spatio-temporelle de la criminalité (Mohler et al., 2011). Évidemment, plus les dimensions des données sont explorées (informations spatio-temporelles, modes opératoires, données forensiques, etc.), plus les chances sont grandes de détecter un changement.

En connaissant les patterns, il est plus aisé de les détecter. Mais les modèles computationnels proposent idéalement un plus ambitieux projet : générer de nouvelles connaissances, c.-à-d., dans notre perspective, détecter des nouveaux patterns inattendus, sans se fonder sur des *a priori*.

5.5. L'utilité du pattern

Bien que l'existence et la possibilité de détecter des patterns soient démontrées, il reste à déterminer comment évaluer l'utilité de ces derniers. Les concepts présentés par Soergel (1994) permettent de nous guider dans cette tâche.

Soergel hiérarchise en trois niveaux les performances de la récupération d'information : l'indexation doit ainsi être congruente (*relevance*), pertinente (*pertinence*) et utile (*utility*). Selon Soergel, sur une problématique donnée, une entité est ***congruente*** si celle-ci est susceptible d'apporter des informations sur la problématique. Ensuite, une entité est ***pertinente*** si elle est congruente et qu'elle est appropriée pour l'utilisateur. Autrement dit, la personne utilisant une entité doit pouvoir la comprendre et appliquer l'information obtenue. Finalement, une entité est ***utile*** si elle est pertinente et qu'elle apporte une valeur ajoutée à la connaissance de l'utilisateur.

Ces concepts ont déjà été appliqués par analogie sur des problématiques forensiques, notamment au sujet de la pertinence en science forensique (Hazard, 2014) et de l'utilité de l'indice (Bitzer, Albertini, Lock, Ribaux, & Delémont, 2015; Bitzer et al., 2016). Les patterns présentant des caractéristiques similaires aux traces (vestige d'une activité pour les traces, reflet d'une activité pour les patterns), penser l'utilité du pattern selon ce prisme semble opportun.

Ainsi, le pattern doit avant tout être ***congruent***[24]**.** C'est-à-dire qu'il doit être lié à la problématique considérée et pouvoir apporter des informations sur celle-ci. Cette congruence peut être appréhendée sous la distinction entre l'activité dite normale ou habituelle et l'activité anormale ou inhabituelle. Des patterns sont susceptibles d'être détectés dans les structures des données et de refléter des structures d'activités, mais s'ils ne concernent pas la problématique considérée, alors le questionnement de leur utilité s'arrête là. En analyse criminelle, ce sont les patterns qui sortent de la norme qui retiennent particulièrement l'attention. Il peut s'agir de distinguer des patterns d'activités criminelles au sein de patterns d'activités tout à fait licites, mais également de détecter des patterns d'activités anormales au sein de patterns d'activités illicites

[24] Une autre traduction française est proposée pour les termes *relevance* et *pertinence* dans les travaux de Hazard (2014), il s'agit de *pertinence factuelle* et de *pertinence appropriée,* ce qui se justifie par l'intérêt centré sur la notion de pertinence dans sa recherche. Néanmoins, nous avons préféré opté pour *congruence* et *pertinence* qui semble plus approprié une fois que l'on applique ces concepts aux patterns étudiés.

habituels, par exemple l'apparition de nouveaux groupes d'auteurs ou un changement de mode opératoire pour un certain type de situations criminelles. L'étude de Reardon, Nance et McCombie (2012) sur la détection de l'utilisation de cartes bancaires contrefaites illustre bien cette distinction entre patterns d'utilisations habituelles (activités licites) et inhabituelles (activités illicites)(Figure 13).

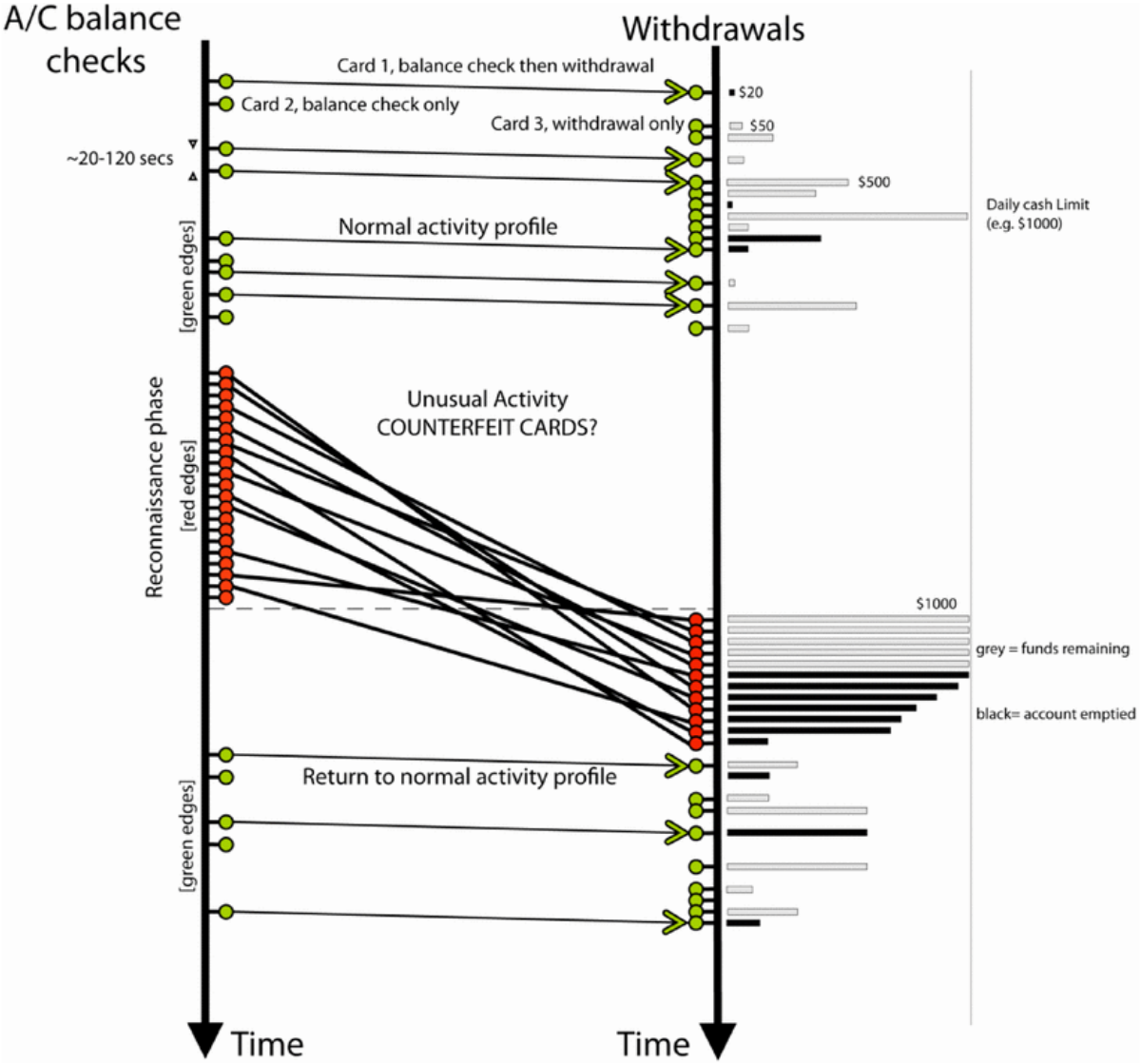


***Figure 13** : Exemple de pattern dans l'utilisation des cartes bancaires au DAB (Reardon et al., 2012) : La ligne temporelle de gauche montre les vérifications du solde d'un compte et celle de droite, les retraits effectués. Les correspondances entre une même carte pour la vérification du solde et le retrait sont indiquées par une flèche reliant les deux lignes temporelles. On distingue le pattern d'activité normale qui consiste en général à une vérification de solde suivi d'un retrait pour une même carte (indiqué en vert sur le schéma), et le pattern d'activité inhabituelle qui consiste en une séquence de vérification de solde sur plusieurs cartes sans retrait, puis une séquence de retrait avec une partie de ces mêmes cartes (indiqué en rouge dans le schéma).*

Pourtant, la congruence du pattern avec la problématique considérée ne suffit pas. Le pattern doit également être approprié pour l'utilisateur afin de devenir ***pertinent.*** Les compétences de la personne, sa formation, sa familiarité avec la problématique, ce sont autant d'éléments qui sont susceptibles d'influencer sa capacité d'appréhension du

pattern. Si l'on revient à l'exemple des patterns d'utilisation de cartes bancaires au DAB (Figure 13), le pattern est bel et bien présent dans les données et donc potentiellement détectable. Mais sa pertinence dépendra de l'utilisateur qui l'observe. On peut imaginer qu'un employé du service informatique de la banque remarque ce pattern dans les données, mais il ne sera pas forcément pertinent pour lui, car ses connaissances des phénomènes criminels sont généralement restreintes. Face à ce pattern, il se sentira comme un pingouin dans le désert pour reprendre l'expression consacrée. Au contraire, un analyste criminel pourra interpréter ce pattern à l'aide de ses connaissances sur les types de criminalités liés aux cartes bancaires. En effet, lorsqu'un auteur dérobe des cartes bancaires ou en copie les informations, à l'aide par exemple d'un dispositif de skimming, son objectif est de récupérer l'argent situé sur les comptes au plus vite. Dans un souci d'efficacité, l'auteur concentrera sa venue au DAB à des moments-clés. Afin de sélectionner les cartes bancaires disposant des plus grands capitaux, l'auteur vérifiera le solde de toutes les cartes avant de commencer à retirer l'argent des comptes les plus fournis. De plus, ces opérations se déroulent aux environs de minuit ce qui peut également être expliqué par une hypothèse situationnelle. La limite de retrait des cartes bancaires étant souvent remise à zéro à cette heure, cela permet au malfaiteur de retirer deux fois la somme maximale de retrait par cartes. Ces hypothèses sur l'activité inhabituelle sont reflétées dans la structure des données et sans connaissance des phénomènes criminels, il est très difficile d'interpréter ce type de pattern.

Finalement, pour qu'un pattern soit ***utile***, il doit non seulement être pertinent, mais il doit également apporter une valeur ajoutée, quelque chose que l'utilisateur ne connaissait pas sur la problématique considérée. En effet, un pattern peut tout à fait être pertinent, c'est-à-dire congruent et approprié, mais n'apporter aucune information nouvelle à l'utilisateur. Le gain d'information dépend bien évidemment des buts en rapport à la problématique. Dans l'exemple précédent des cambriolages du soir (Graphique 6), le pattern est congruent avec la problématique puisqu'il représente bien une fréquence inhabituelle des cambriolages en soirée durant l'hiver par rapport au reste de l'année. Il est également pertinent puisqu'un analyste criminel à la capacité de le détecter et de l'interpréter à l'aide d'une explication situationnelle. La lumière visible dans les habitations durant les soirées hivernales renseigne les potentiels cambrioleurs sur la présence d'occupants. Néanmoins, ce pattern n'est pas forcément utile car si l'objectif est de détecter des changements, alors le pattern est considéré

comme trivial puisque les analystes connaissent ce phénomène. Ils n'apprennent rien de nouveau et le pattern leur est inutile. En revanche, cet exemple trouve une autre utilité si l'on poursuit un objectif différent. Le fait de savoir que les cambriolages du soir augmentent chaque année à la même période permet de mettre en place des campagnes de prévention consistant par exemple à conseiller à la population de laisser un éclairage de veille lorsqu'ils ne sont pas chez eux[25]. De même, dans l'exemple du pattern d'utilisation des cartes bancaires au DAB, la connaissance du pattern permet de mettre en place des mesures visant à perturber l'activité criminelle, par exemple la mise en place d'un système de veille qui détecterait ce schéma d'activités et qui bloquerait les cartes à partir d'un certain seuil de soldes consultés à la suite sans retrait.

En résumé, l'utilité d'un pattern et sa capacité d'utilisation dépendent à la fois de la capacité de détection qui vise à remarquer les activités inhabituelles au sein des activités habituelles et de la capacité d'interprétation qui infère le lien entre les structures des données et les structures des activités. La détection et l'interprétation ne sont possibles que si le pattern est congruent avec la problématique considérée et qu'il est approprié pour l'analyste criminel (Figure 14).

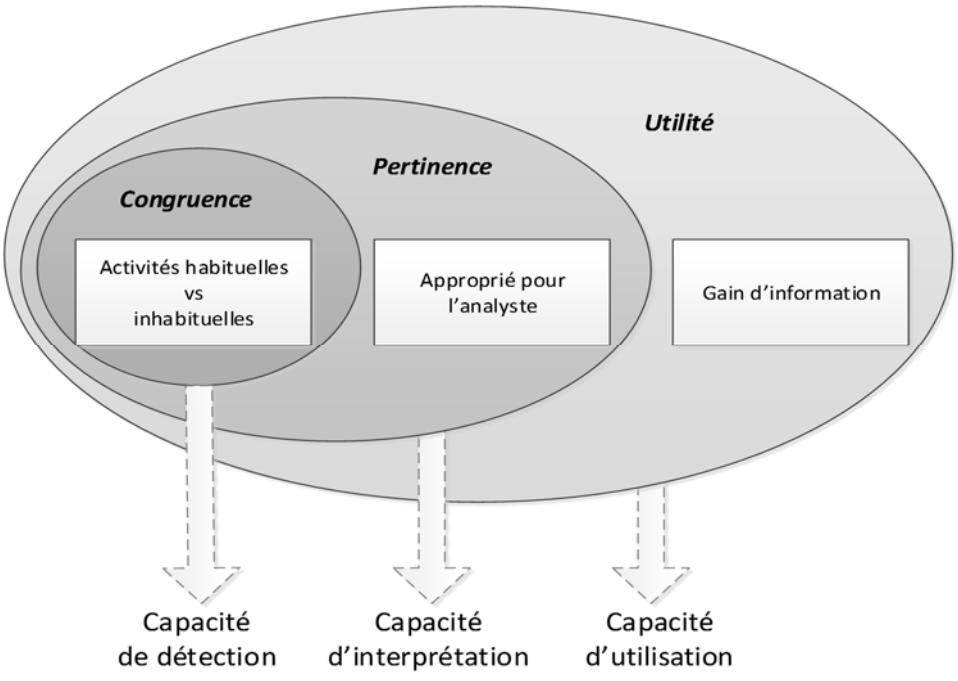


***Figure 14** : Utilité du pattern : la congruence du pattern avec la problématique considérée en cherchant l'inhabituel au sein de l'habituel fournit une capacité de détection. En étant appropriée pour l'analyste criminel, la pertinence du pattern permet alors une interprétation de ce dernier. Finalement, combinée à un gain d'information et dépendamment des objectifs fixés, l'utilité du pattern permet son utilisation concrète à des fins préventives, proactives ou réactives.*

[25] Voir Annexe B

6. Une approche crimino-forensique en renseignement criminel

6.1. Le pattern à la croisée des chemins

Le pattern de données joue un rôle central dans la compréhension des processus d'analyse et de renseignement criminel et la place qu'il occupe fait figure de carrefour entre les différents niveaux d'analyses (Figure 15). Ces niveaux d'analyse s'inscrivent dans la schématisation proposée par Birrer (2010) et inspirée des travaux de Ribaux et Margot (2007). Ce modèle permet de distinguer trois formes d'analyses : l'analyse criminelle opérationnelle (également connu sous le nom « analyse tactique » au Canada) qui soutient l'enquête et vient en appui du processus d'investigation traditionnel ; le renseignement criminel opérationnel cherche à identifier et à comprendre les phénomènes criminels pour assister les décideurs et coordonner les mesures à entreprendre ; finalement, le renseignement criminel stratégique tient compte des tendances générales de la criminalité afin d'en évaluer les menaces, ampleurs et impacts sur le long terme. Cette schématisation fait directement écho à la classification des niveaux d'abstraction de problèmes proposée par Cusson et Ribaux (2015) : la catégorie de problème, le niveau territorial ou du réseau, et le cas particulier et singulier. La catégorie de problème renvoie au niveau le plus général des problèmes de criminalités, par exemple les cambriolages dans les résidences, les braquages de bijouteries ou les fraudes par carte de crédit. Le niveau territorial représente la manière dont les catégories de problèmes se traduisent spécifiquement au sein d'une région, d'une ville ou d'un quartier, par exemple un groupe de cambrioleurs qui sillonne une région ou des violences dans une rue à la sortie des bars. Enfin, le cas particulier décrit l'événement litigieux porté à la connaissance des acteurs impliqués dans l'action de sécurité, par exemple un cambriolage ou une agression spécifique.

Tout en s'inscrivant dans la continuité de ces deux modèles, une précision supplémentaire est apportée dans notre proposition. La configuration environnementale, c'est-à-dire les composants de l'environnement physique et social, décrit le niveau général et stratégique dans lequel des catégories de problèmes sont susceptibles de survenir en fonction de la configuration. Au niveau opérationnel, le pattern d'activité regroupe un ensemble d'activités répétitives qui sont les effets directs du problème identifié en amont. Finalement à un niveau tactique, le cas particulier représente un événement singulier, le crime ou l'activité litigieuse, qui compose le pattern d'activité. Pour tenter de comprendre ces différents niveaux de réalité, les

processus en analyse et renseignement criminel cherchent à les expliquer à l'aide d'hypothèses formulées en se basant les données collectées. La situation criminelle reflète une configuration environnementale spécifique, le pattern de données reflète le pattern d'activité et les données reflètent le cas particulier. Ce processus n'est cependant pas linéaire, car la détection de pattern peut découler tant des données que des situations criminelles déjà connues de l'unité d'analyse. L'exploration sans *a priori* de données collectées peut amener à la détection de nouveaux patterns qui sont susceptibles de définir de nouvelles situations criminelles. D'un autre côté, la reconnaissance de situations criminelles, découlant par exemple d'expérimentations menées en criminologie, peut également guider la détection de patterns déjà connus, et par extension, la collecte de traces.

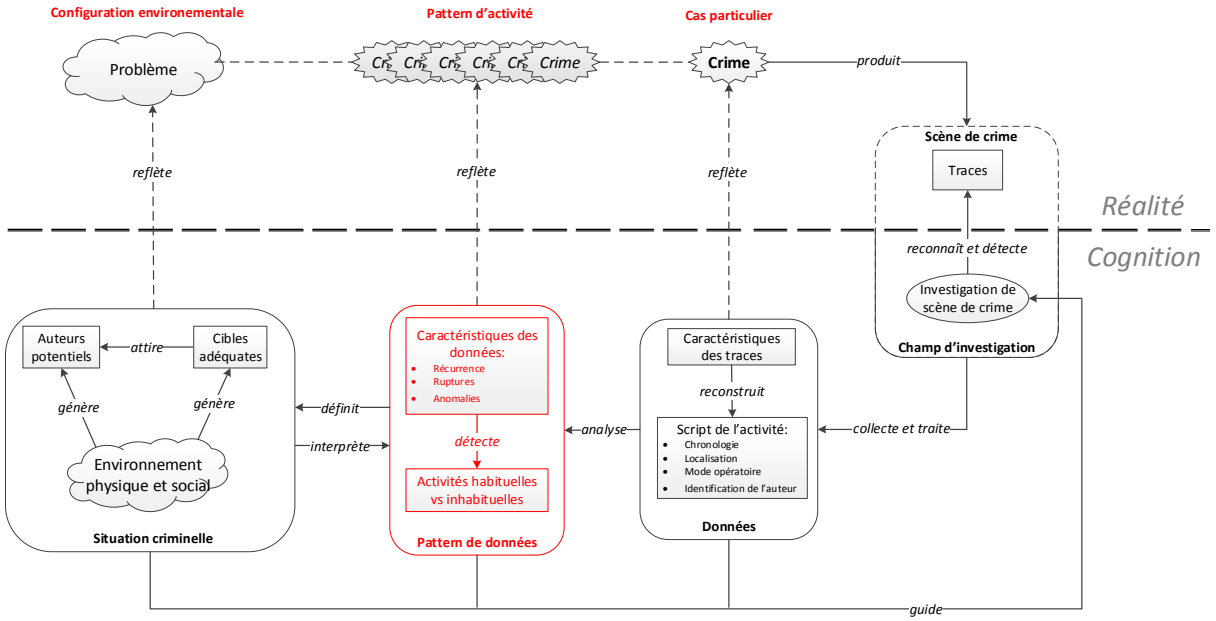


***Figure 15** : Dernière étape de formalisation : le pattern comme pivot. La présence de pattern congruent (c.-à-d. pouvant être détecté) et pertinent (c.-à-d. pouvant être interprété) dans les données permet de définir des situations criminelles. Le processus général est ainsi décomposé en trois niveaux : le problème, le pattern d'activité et le cas particulier. La trace fait figure de transition entre la réalité et la cognition, tandis que le pattern détient un rôle de pivot entre les différents niveaux d'analyses.*

La distinction en trois niveaux d'analyse est cruciale dans l'intégration des méthodes computationnelles, car les objectifs poursuivis, le timing, les contraintes et les ressources disponibles varient entre les niveaux. Ainsi, la performance des techniques utilisées est fortement dépendante de l'étape à laquelle ces dernières sont appliquées. Une technique d'analyse de séries temporelles peut se révéler efficace si elle est utilisée à long terme afin d'analyser des tendances globales sur plusieurs mois, mais inutile à niveau opérationnel qui nécessite une capacité de proaction rapide. L'application de

méthodes computationnelles nécessite de se poser ces questions en amont et de connaître le contexte dans lequel elles s'inscrivent afin de sélectionner les bonnes pratiques à mettre en œuvre. Cela peut s'apparenter à un véritable jeu d'équilibriste où il est nécessaire d'avancer pas après pas sans se précipiter, sous peine de chuter.

6.2. À la recherche de l'équilibre

Notre unité d'analyse criminelle fait donc face à des choix critiques lorsqu'elle franchit une étape de développement : intégrer des modèles computationnels de manière cohérente dans sa méthodologie basée sur la science forensique et la criminologie environnementale.

Ainsi se pose la question de la délimitation des objectifs concrets qu'il s'agit d'atteindre. Ces améliorations peuvent prendre différentes formes durant le processus du renseignement, par exemple :

- Assurer la validité, la fiabilité et l'intégrité des données lors de l'encodage et de la structuration de l'information décrite avec des métadonnées
- Fournir une capacité de détection de patterns plus complète, plus rapide et plus fine grâce à une meilleure formalisation des connaissances
- Permettre une gestion multitâche afin de réduire le temps d'analyse de l'information, par exemple en opérant un suivi de différents types de cambriolage en simultané ou en observant les données sous différents niveaux d'agrégation spatio-temporelle.

C'est alors que l'enjeu est de gérer l'équilibre fragile entre l'intégration de différents types de connaissances dans le processus et la capacité réaliste d'extraire de la connaissance à l'aide de méthodes computationnelles. C'est-à-dire pour nous, de découvrir des patterns encore inconnus plutôt que de détecter de nouvelles occurrences à partir de patterns déjà connus (par exemple, les cambriolages du soir). Cet équilibre peut être envisagé à travers les processus de découverte de connaissance dans les bases de données (Knowledge Discovery Databases) qui visent à produire de la connaissance à partir des données (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). Cela initie les questions sur quand et où les connaissances sur les comportements criminels doivent être intégrées afin de soutenir la découverte de patterns. Certains d'entre eux restent indétectables par des algorithmes totalement autonomes, mais trop de

connaissances injectées dans le processus peuvent créer un effet tunnel qui limite la détection des nouveautés (Kahneman, 2013).

Le data mining guidé par le domaine (D3M) (Cao, 2008) est un mouvement en modélisation computationnelle qui semble adéquat pour guider la recherche de l'équilibre subtil à atteindre dans un environnement et une méthodologie préexistants. L'intégration adéquate des connaissances du domaine durant tout le processus y est exprimée comme une nécessité. Fondamentalement, ce mouvement reconnaît que posséder une compréhension globale des données et des règles inhérentes au domaine améliore l'efficacité des méthodes de fouille de données et le sens de ses résultats. Le data mining guidé par le domaine peut aussi être présenté comme une réponse « pour combler le gouffre entre le domaine académique et opérationnel »[26] (Cao, Yu, Zhang, & Zhao, 2010, p. vii) en fournissant des solutions pratiques et en maîtrisant les connaissances du domaine et de leurs contraintes. C'est dans un esprit similaire à cette approche qu'il est possible d'envisager l'intégration des méthodes computationnelles en renseignement criminel à l'aide de la démarche méthodologique décrite dans cette partie.

6.3. Synthèse

En prenant la méthodologie de l'unité d'analyse du CICOP comme fondement, une approche intégrative et itérative dans l'application des méthodes computationnelles en renseignement criminel est exprimée. L'approche proposée est basée sur un postulat fondamental en analyse criminelle : les activités litigieuses suivent des patterns susceptibles d'être détectés et analysés à l'aide des données disponibles. Le raisonnement est basé sur la plus élémentaire de ces données : la trace, vestige physique (et numérique) de l'activité litigieuse, qui a été reconnue et collectée sur les scènes de crime. En combinant cet aspect forensique avec les approches situationnelles en criminologie, il est possible de décomposer l'approche en trois niveaux d'analyse (Figure 16).

[26] Citation originale: "to bridge the gap between academia and business".

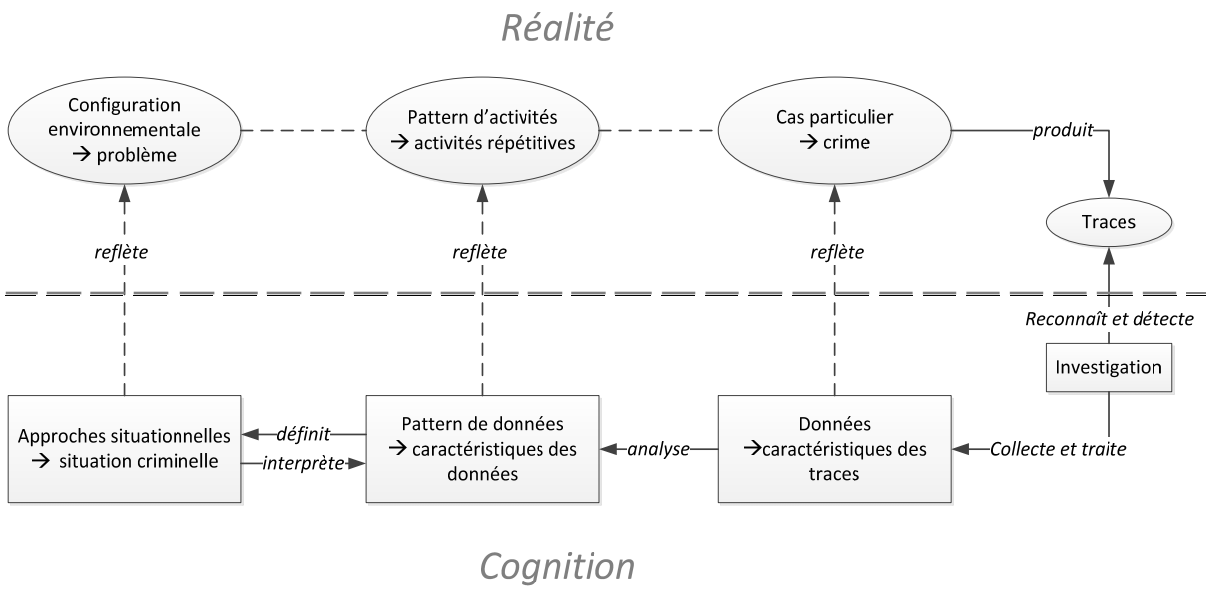


***Figure 16** : Formalisation des inférences en analyse et renseignement criminel : l'analyse et le renseignement criminel se décompose en trois niveaux d'analyse qui correspondent à différents niveaux de réalités, c.-à-d. la configuration environnementale qui est interprétée par les approches situationnelles, le pattern d'activité qui est reflété par le pattern de données et le cas particulier qui est représenté par les données.*

PARTIE III : VERS UNE APPLICATION CONCRÈTE ET PERTINENTE

La partie précédente ayant permis d'exprimer la démarche méthodologique susceptible de favoriser l'intégration des méthodes computationnelles en analyse criminelle, cette partie vise à intégrer quelques méthodes computationnelles simples de classification et de détection de ruptures dans les processus du CICOP, en exploitant le cadre développé plus haut. Il ne s'agit pas d'évaluer la performance des méthodes computationnelles utilisées, mais plutôt de montrer comment le cadre de travail proposé est susceptible d'aider à identifier leur contribution potentielle à des fins de renseignement criminel dans un environnement et une méthodologie qui intègrent des éléments de criminologie environnementale et de renseignement forensique préexistants.

Ainsi, nous procéderons en quatre temps :

1. L'identification et l'expression de quelques processus d'analyse criminelle exploités dans le cadre du CICOP, encore tacites, permettront d'isoler des champs d'applications prometteurs pour des méthodes computationnelles ; pour cela, un langage appelé « Business Process Model and Notation (BPMN) » sera utilisé (chapitre 7).
2. Parmi ces champs, la classification automatique des nouveaux cas dans le système de classification fondé sur les situations criminelles, exploité par le CICOP, apparaîtra comme particulièrement pertinente; des méthodes computationnelles classiques seront expérimentées pour tester cette possibilité sur le jeu de données rassemblé (chapitre 8).
3. Une autre application consistera en la détection de ruptures dans les tendances susceptible d'attirer l'attention vers un changement dans la structure des crimes répétitifs; une méthode moderne qui relève du « *Change Point Analysis* » sera testée sur les jeux de données à disposition (chapitre 9).
4. Finalement, le potentiel de ces méthodes et leur implémentation dans le processus global sont discuté dans une synthèse, de même que la réponse aux hypothèses de recherche (chapitre 10).

7. Démarche d'application axée sur les processus

L'unité d'analyse du CICOP opère en suivant le cadre de la veille opérationnelle décrite au chapitre 4.1. L'un des objectifs du CICOP étant le suivi de la criminalité sérielle et itinérante, le suivi des cambriolages représente une partie importante de leurs activités. Les cambriolages, et particulièrement les cambriolages d'habitations, représentent une proportion importante du volume de la criminalité en Europe (Brown, Ross, & Attewell, 2014; Lammers, Bernasco, & Elffers, 2012; Ritter, 2008) et de ce fait, également en Suisse (Baechler, Cartier, Schucany, & Guéniat, 2015). Si l'on observe l'ensemble des infractions enregistrées par la Police cantonale vaudoise en 2008 (N= 38'638)[27], 24.02% (9'281) des infractions concernent des cambriolages[28], commerces et habitations confondus. De même, les données de PICAR en 2008 nous montrent que sur les 18'922 événements enregistrés, 22.04% (4'171) sont des cambriolages d'habitation.

Les analystes du CICOP utilisent un système de classification fondé sur ce qu'ils appellent des « codes phénomènes ». Ces codes délimitent des situations criminelles, au sens de la criminologie environnementale, présentant des patterns de données spécifiques (voir chapitre 3.2.1). Les systèmes d'analyse des cambriolages utilisent traditionnellement plutôt des systèmes de classification fondés sur le mode opératoire, ou la technique utilisée par le malfaiteur pour opérer (Birkett, 1989; Vollmer, 1919). La codification du CICOP est originale, car elle intègre ces éléments dans une conception situationnelle englobant le moment de la journée ou le type de cible. Néanmoins, son efficacité dépend de la qualité de l'intégration des nouveaux cas dans PICAR.

Le processus de veille opérationnelle constitue le point de départ le plus général de ce travail de modélisation. Comme illustré par le processus SARA décrit au chapitre 4.1, l'étape d'analyse est clairement distinguée de l'étape de détection. La détection apparaît comme une des étapes clés au même titre que l'intégration (Figure 17). Cette modélisation se révèle cependant insuffisante si l'on cherche à localiser les tâches susceptibles d'être améliorées par un composant computationnel. Une démarche

[27] Base de données Zéphyr qui enregistre tous les cas enregistrés par la police et ayant fait l'objet d'une plainte sur le canton de Vaud.
[28] Les cambriolages dans Zéphyr sont définis par les vols par effraction et les vols par introduction clandestine.

d'application axée sur les processus est ainsi nécessaire pour identifier et exprimer certains processus encore tacites d'intégration et de détection exploités par le CICOP.

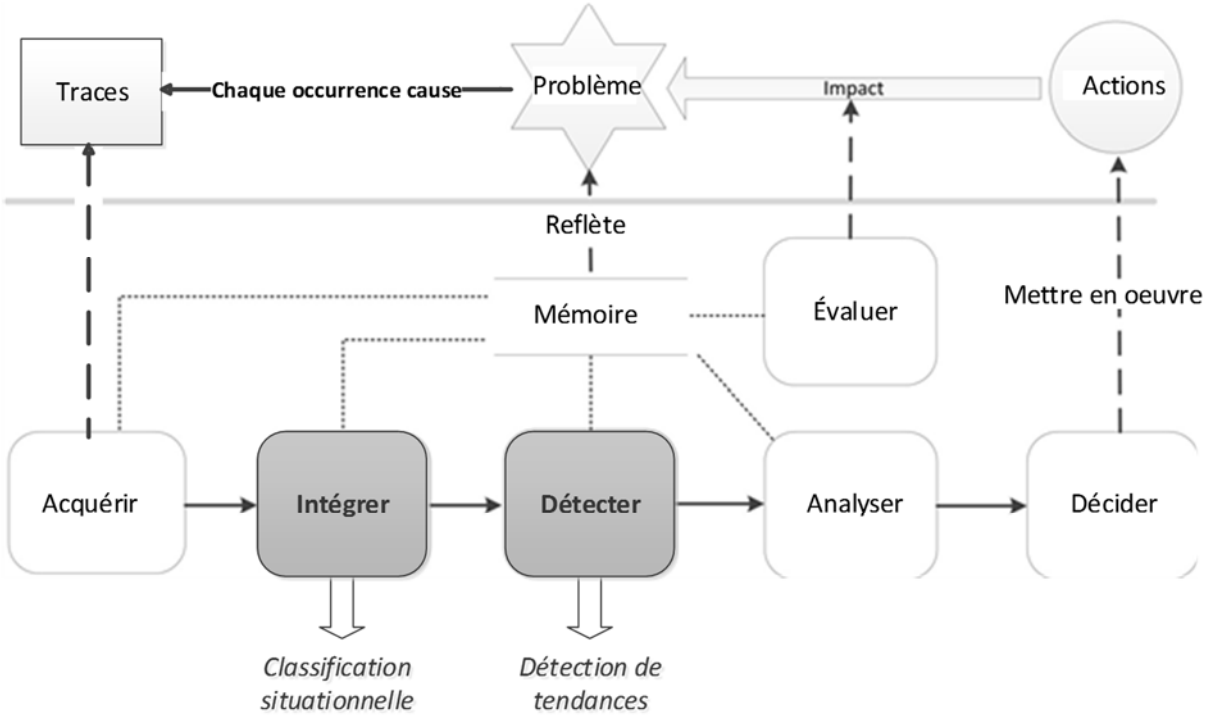


***Figure 17** : Sélection des étapes de la veille opérationnelle : les étapes d'intégration et de détection sont illustrées à travers le processus de classification situationnelle des cambriolages d'habitation et du processus de détection de tendances des activités criminelles.*

Concrètement, la démarche d'application des méthodes computationnelles se décompose en trois étapes (Figure 18). Il convient d'exprimer et modéliser les différentes structures d'inférence identifiées en renseignement criminel à l'aide d'une notation standardisée appelée Business Process Model and Notation (BPMN) (1), afin d'identifier les contributions computationnelles potentielles en fonction des besoins définis (2) et, enfin, d'interpréter les résultats à l'aide d'une analyse forensique et criminologique (3).

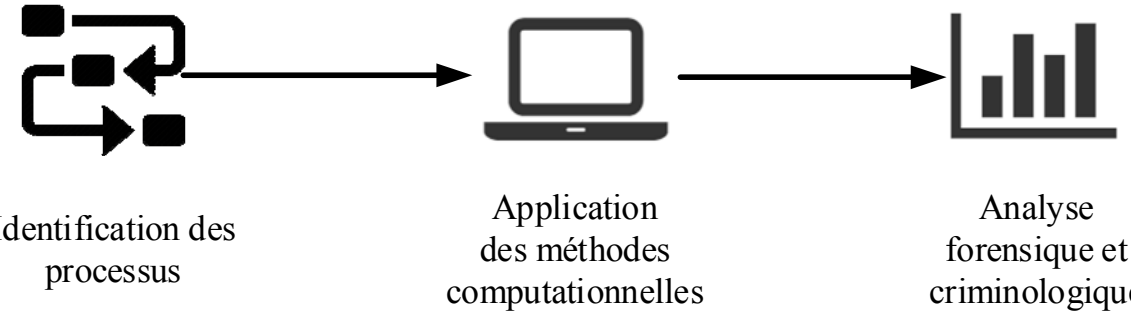


***Figure 18** : La démarche d'application des méthodes computationnelles basée sur les processus : Cette démarche voit se succéder l'identification et la formalisation des processus en renseignement criminel, l'application des méthodes computationnelles et enfin, l'analyse forensique et criminologique des résultats.*

7.1. Modélisation des processus

La première étape consiste donc à exprimer des processus propres au renseignement criminel, et d'autre part à les formaliser à l'aide d'une notation standardisée. Cette formalisation poursuit trois objectifs :

- Exprimer les raisonnements implicites tenus par les analystes.
- Identifier les processus où l'application de méthodes computationnelles peut se révéler pertinente.
- Définir les raisons pour lesquelles l'application d'une méthode computationnelle est réaliste et souhaitable.

Cette étape permet une meilleure compréhension des contraintes et spécificités inhérentes à l'analyse criminelle dans un environnement particulier, ce qui est crucial si l'on veut piloter de manière adéquate l'intégration de méthodes computationnelles. En effet, une partie de ces techniques nécessite, non seulement une préparation adéquate des données pour l'analyse, mais aussi une interprétation adaptée des résultats. Cela demande une solide connaissance des processus qui produisent ces données et de leur utilisation. Nous utilisons deux sources principales pour déterminer la nature des processus propres au renseignement criminel. La première vient puiser dans la littérature scientifique afin d'en dégager les processus qui pourraient s'appliquer au contexte étudié. Des structures d'inférences permettant d'intégrer des données forensiques dans un système basé sur le renseignement ont notamment été formalisées par Ribaux et Margot (1999). La seconde source puise dans le savoir et l'expérience des analystes criminels, car ces derniers utilisent quotidiennement des structures de raisonnement de manière implicite. Pour atteindre cet objectif, une observation en immersion dans la conduite des activités au sein de la division

Coordination et renseignement judiciaire (DCRJ)[29] de la Police cantonale vaudoise a été réalisée durant une semaine. L'observation constitue un outil adéquat pour identifier des éléments qui ne sont pas verbalisés en entretien, en lien notamment avec des postures, attitudes et (inter)actions (Chapoulie, 2000 ; Diaz, 2005). Cette observation nous a permis d'initier la collaboration avec l'unité d'analyse et d'effectuer une première approche de leur cadre de travail. Elle fut suivie de rencontres régulières avec les collaborateurs de cette unité tout au long de la réalisation de ce travail. Ces contacts nous ont ainsi permis de saisir très concrètement la diversité des missions assignées à une unité de renseignement criminel et la variété des cas auxquels elle peut être confrontée. Ils colorent de l'intérieur notre appréhension du milieu et de la sorte ont considérablement enrichi notre question de recherche. L'observation a été couplée à une analyse de documents jugés pertinents pour exprimer les processus de travail des analystes. Ce travail a consisté en un dépouillement de la littérature grise et de l'ensemble des rapports et synthèses de travail à disposition, puis en une analyse des tâches, de la durée et de l'organisation des processus identifiées. Il est ainsi possible à l'aide du croisement de ces différentes sources de données d'exprimer ces processus implicites et de les modéliser.

Après l'identification des processus, il convient alors de les modéliser à l'aide du formalisme adéquat. Afin d'obtenir un résultat satisfaisant, le formalisme sélectionné doit pouvoir être relativement simple pour pouvoir être compris par des personnes non expertes en formalisation de processus, tout en étant suffisamment exhaustif afin de représenter de manière précise des processus potentiellement complexes. Parmi les nombreuses manières de modéliser des processus, nous avons dû effectuer un choix qui remplisse les critères exposés. Cette décision a également été guidée par l'expertise de l'équipe de recherche en sciences de l'information avec laquelle nous collaborons (voir chapitre 3.1). Ainsi, le Business Process Model and Notation (BPMN) s'est présenté comme un choix adéquat de par sa simplicité et sa flexibilité et sa proximité avec une approche orientée sur les processus.

7.1.1. Business Process Model and Notation (BPMN)

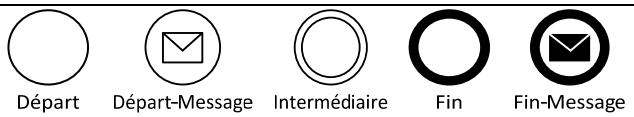
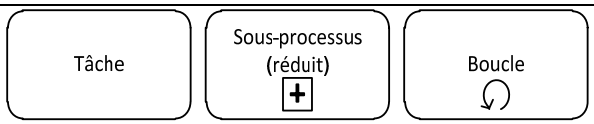
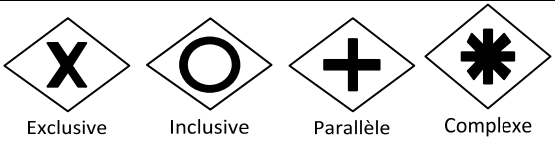
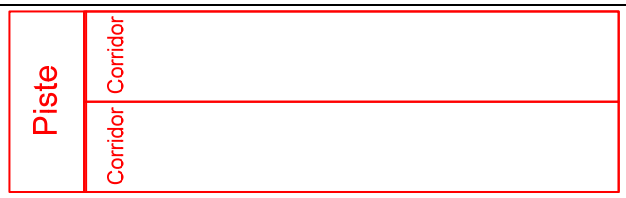
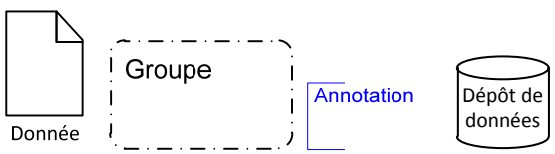
Il s'agit d'une notation graphique standardisée qui décrit les étapes d'un processus métier et qui a été spécialement conçue afin de coordonner le séquençage des processus (OMG, n.d.). À l'aide de cette notation, il est possible de coordonner des acteurs

[29] Appelée division Coordination judiciaire à l'époque.

multiples en formalisant des processus métier pouvant être complexes. Il s'agit d'une approche orientée vers les processus afin de modéliser des systèmes. Les principaux symboles de notation sont présentés au Tableau 1.

Étant donné la présence de structures d'inférence dans le cadre du renseignement criminel, il est possible dès lors de les formaliser à l'aide de la notation BPMN. La même formalisation peut également être réalisée pour les différents processus qui permettent aux analystes des services de police de classifier et analyser les différents événements portés à leur connaissance. Afin de vérifier l'adéquation de la notation BPMN, les structures d'inférences identifiées par Ribaux et Margot (1999) ont été exprimées sous la forme de processus BPMN[30].

***Tableau 1** : Notation BPMN adapté de Cotofrei et Stoffel (2011)*

Catégories	***Éléments***	***Quelques exemples (notations graphiques)***
Objets de flux	Événements	
	Activités	
	Passerelles	
Connecteurs	Flux de séquence	
	Flux de message	
	Associations	
Couloirs	Piste	
	Corridor	
Artefacts	Objet de donnée	
	Groupe	
	Annotation	
	Dépôt de données	

[30] Les processus formalisés sont disponibles à l'Annexe C.

7.1.2. Processus d'enregistrement des événements dans PICAR

Le processus utilisé par la coordination judiciaire vaudoise, qui intègre le système PICAR, peut s'interpréter dans le cadre de la veille opérationnelle (suivi permanent des cambriolages d'habitations). Un niveau de modélisation plus fin est toutefois exigé pour mieux situer les possibilités d'intégrer des méthodes computationnelles, tant au niveau du choix des techniques que de leurs localisations dans le processus de renseignement criminel. C'est grâce à l'observation en immersion et l'analyse de documents décrites dans le chapitre 7.1 que le cheminement d'un événement de son annonce aux services de police jusqu'à son inscription dans PICAR a pu être formalisé à l'aide de la notation BPMN. Comme constaté à la Figure 19, tout événement traité par la police passe à travers deux filtres. Le premier est automatisé et transmet à la coordination judiciaire à partir du journal d'événements police (JEP) une liste d'événements grossièrement triée selon le type d'infraction. L'objectif de ce premier filtre est d'éliminer la majorité des événements non pertinents à des fins de renseignement criminel. Par exemple, les événements impliquant la loi sur la circulation ou la loi sur les stupéfiants sont automatiquement écartés, pour peu qu'ils ne soient pas caractérisés par une catégorie cible[31]. Une fois la liste d'événements épurée, c'est au tour du second filtre d'agir. Celui-ci est entièrement manuel, puisque c'est l'analyste qui va prendre un par un les événements présélectionnés dans le power JEP (les événements pré-triés du JEP) afin de décider s'ils doivent être traité et intégré dans PICAR. Cette étape qui permet de classifier les événements à l'aide des codes CICOP est une étape indispensable dont dépendent les processus de détection et d'analyse ultérieurs. Cette dernière peut cependant prendre du temps et la fiabilité de la classification reste tributaire de la variabilité entre les différents analystes traitant les événements. Les avantages d'une technique computationnelle apparaissent alors plus clairement dès lors qu'ils sont susceptibles d'améliorer le temps de traitement des événements et la fiabilité de la classification.

[31] Voir **Annexe 16**

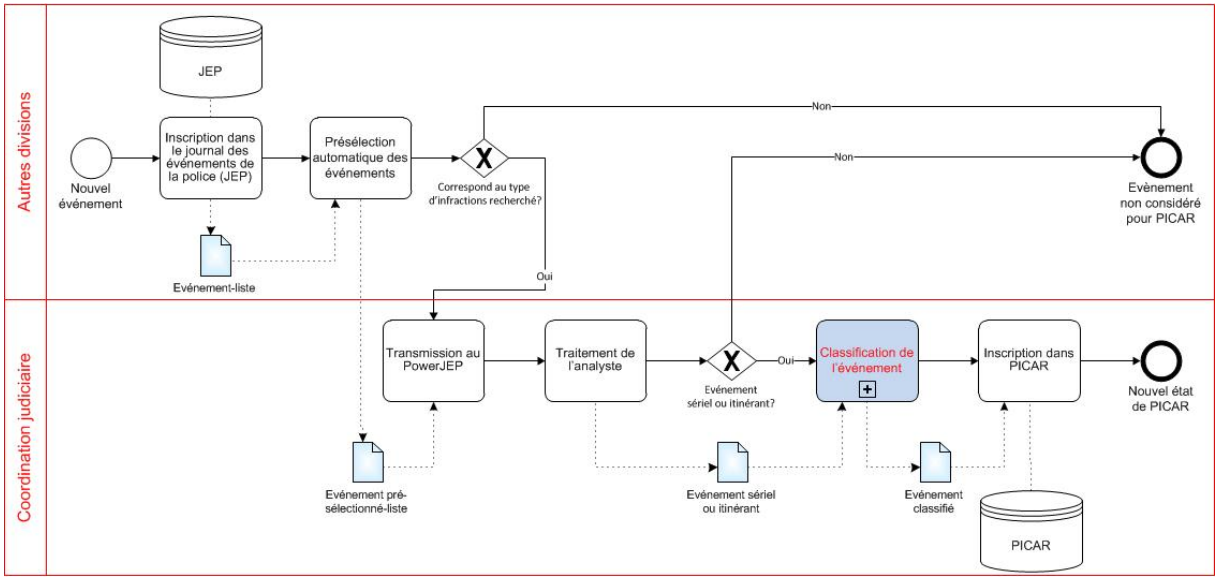


Figure 19 *: Processus BPMN de l'enregistrement d'un événement dans la base de données PICAR*

7.2. Application des méthodes computationnelles

Une fois les processus formalisés à l'aide de la notation BPMN, des méthodes computationnelles sont sélectionnées pour répondre aux besoins identifiés et sont appliquées. Il est important non seulement de savoir à quel moment utiliser les techniques computationnelles, mais également quel type de techniques considérer. Cela va dépendre de différents facteurs : but de la tâche à effectuer, nature et quantité des données disponibles, utilisation des résultats à court, moyen ou long terme, personnes à qui sont destinés les résultats. Tous ces facteurs vont conditionner le choix et l'utilisation des méthodes computationnelles. De ces facteurs spécifiques découlent quatre critères principaux auxquels la sélection des techniques doit à notre avis idéalement répondre :

- La simplicité d'utilisation.
- La flexibilité.
- La rapidité de traitement.
- L'insertion dans les processus de travail existant.

On peut distinguer 2 grandes familles de techniques computationnelles[32] (Berry & Linoff, 2004) : l'apprentissage supervisé et l'apprentissage non supervisé. Le but de l'apprentissage supervisé est de trouver la valeur d'une variable cible particulière. À

[32] Cette distinction est définie originalement pour les techniques de *data mining*, mais nous l'appliquons par analogie aux méthodes computationnelles en général.

l'inverse, l'apprentissage non supervisé cherche à révéler des structures dans les données sans tenir compte d'une variable cible particulière.

Dans l'étude des techniques de *data mining* dans la détection des fraudes financières, Ngai et ses collègues (2011) distinguent six types de techniques :

- La **classification**, qui permet de construire et utiliser un modèle pour prédire la catégorie à laquelle appartient un objet inconnu afin de distinguer des objets en différentes classes, les catégories étant prédéfinies.
- Le ***clustering***, qui divise les objets en groupes conceptuels significatifs sans avoir de catégories prédéfinies.
- La **prédiction**, qui estime des valeurs numériques en se basant sur les patterns d'un jeu de données.
- La **détection de valeurs aberrantes**, qui permet de détecter des objets qui significativement différent ou inconsistant par rapport au jeu de données.
- La **régression**, qui utilise une méthodologie statistique afin de révéler la corrélation entre des variables indépendantes et dépendantes.
- La **visualisation**, qui se réfère à la méthodologie entourant la présentation de données en vue de rendre clairs et exploitables des patterns complexes pour les utilisateurs.

Le clustering apparaît comme le type de techniques le plus souvent utilisé en science forensique (Sghaier, 2015), notamment de par la volonté de lier des cas entre eux. Cela s'inscrit dans les processus de détection et de découverte de patterns figurant parmi les enjeux principaux en renseignement criminel. Les techniques permettant cette détection requièrent trois étapes dans leur application : le traitement des données, l'application des méthodes de détection de pattern et l'interprétation des patterns détectés (Terrettaz-Zufferey et al., 2006).

7.3. Interprétation forensique et criminologique

Finalement, il reste à interpréter les résultats à l'aune d'une analyse forensique et criminologique. En effet, la plupart des techniques computationnelles vont d'une certaine manière produire des résultats bruts qui nécessitent une interprétation humaine afin de devenir du renseignement exploitable comme nous l'avons vu à travers le processus KDD (voir chapitre 6.2). Dans notre cas, il paraît alors approprié de parler d'analyse forensique et criminologique. Il s'agit concrètement de vérifier la

cohérence des résultats avec les principales théories situationnelles en criminologie vues au chapitre 5.2, c.-à-d. les activités routinières (Cohen & Felson, 1979), le choix rationnel (Felson & Clarke, 1998), le triangle du crime (Clarke & Eck, 2005) et les patterns criminels (P. J. Brantingham & Brantingham, 1990). Cette étape permet non seulement de vérifier la validité des résultats, mais également leur pertinence. Les techniques computationnelles sont susceptibles de produire des résultats triviaux, déjà connus des analystes, mais également des résultats incohérents et ininterprétables. Il s'agit dès lors de pouvoir déterminer si les résultats produits apportent réellement une valeur ajoutée par rapport aux connaissances préalables du phénomène, c.-à-d. d'apprécier l'utilité du pattern détecté (voir chapitre 5.5). Lorsque des traces sont exploités pour filtrer des groupes de cas au sein desquels des patterns sont recherchés, des connaissances forensiques sont évidemment nécessaires afin d'interpréter les résultats (p. ex. le pouvoir discriminatoire des traces). Finalement, les résultats qui auront passé ce filtre pourront être considérés comme des nouvelles connaissances produites et qui pourront servir de support à la prise de décision, que ce soit au niveau opérationnel ou stratégique.

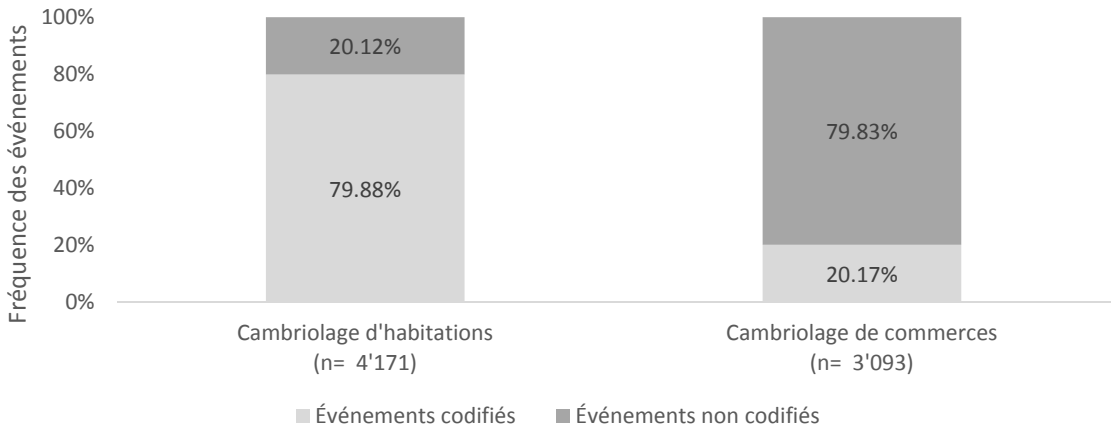
Dans cette optique, les méthodes computationnelles appliquées sont évaluées à travers la confrontation avec les pratiques de l'unité d'analyse de la police, ainsi qu'avec les connaissances forensiques et criminologiques.

8. Classifier : codification des cambriolages d'habitation

Comme première étape d'application de l'approche développée, nous avons choisi la classification d'événements. Concrètement, chaque nouvel événement, par exemple un cas de cambriolage, sur lequel des données ont été collectées est codifié et classifié dans un système prévu pour aider à la détection de répétitions criminelles (et donc de patterns). Une codification/classification automatisée des événements constituerait un avantage déterminant : elle réduirait les temps de saisie dans le système informatisé, et, potentiellement augmenterait la fiabilité de cette saisie (ou diminuerait la dépendance envers celui qui enregistre les données). En l'espèce, il a été choisi de considérer les cambriolages d'habitations comme domaine d'application.

8.1. Les cambriolages d'habitation

La décision de se concentrer sur les cambriolages d'habitation pour cette première étape tient principalement par la longue expérience de l'unité d'analyse du CICOP qui suit ces événements. On constate que sur les données provenant de PICAR, pour les 4'171 cambriolages d'habitations recensés en 2008 sur le canton de Vaud, 79.88% (3'332) sont codifiés à l'aide des codes CICOP (Graphique 8). Le reste n'a pas pu être classifié dans une des catégories pour différentes raisons, par exemple, le manque d'information disponible sur les cambriolages enregistrés ou des erreurs de saisies lors de l'enregistrement des événements.



***Graphique 8** : Codification des cas de cambriolages dans le canton de Vaud en 2008 (source : PICAR).*

On observe cependant une tendance inverse pour les cambriolages de commerce, avec 20.17% (624) des événements codifiés sur les 3'093 cambriolages de commerces recensés. Cela veut dire que ces catégories, ou codes, ne couvrent qu'une partie de ces

formes de délinquances : les analystes policiers semblent démontrer une maîtrise plus complète des cambriolages d'habitations par rapport à ceux dans les commerces.

Nous nous concentrerons sur les cambriolages d'habitation en fonction de ces arguments. Il s'agit de plus d'un phénomène bien présent dans la criminalité suisse avec une forte caractérisation sérielle et itinérante (Birrer, 2010; Ribaux & Birrer, 2008). La question des cambriolages de commerces ne serait toutefois pas sans intérêt d'un point de vue computationnel car des analyses non supervisées pourraient aboutir à la découverte de situations criminelles encore non identifiées par les analystes, qui pourraient elles-mêmes faire l'objet de nouveaux codes permettant de couvrir/classifier une proportion plus grande de cas.

8.2. Processus de classification

En examinant le processus général, il apparaît que certaines tâches peuvent être détaillées et éventuellement permettre l'application de techniques computationnelles. Si l'on considère le sous-processus de classification mis en avant dans la Figure 19, il est possible de descendre d'un niveau de détail et de le représenter également sous la forme d'un processus BPMN (Figure 20). On distingue à ce niveau le raisonnement implicite des analystes qui permet de classifier un événement en un code phénomène (dans notre cas, les codes liés aux cambriolages d'habitation). Cette décomposition a plusieurs avantages :

Tout d'abord, il est possible d'identifier les variables liées à un événement et qui sont prises en considération par les analystes comme critère de classification. Ainsi, pour classifier un événement avec un code phénomène appartenant aux cambriolages d'habitation, les variables suivantes sont utilisées : le *type d'infraction*, le *type de lieu général*, le *moment de la journée*, le *mode opératoire*, et la *voie d'introduction*.

Certaines de ces variables ne sont pas présentes telles quelles dans la mémoire, mais découlent de connaissance forensique et criminologique des analystes. Par exemple, le moment de la journée est déterminé implicitement par les analystes sans qu'un quelconque champ apparaisse dans PICAR. Cela permet d'illustrer l'apport de connaissances implicites d'experts humains auquel ne peut se substituer un algorithme de data mining. Le code phénomène lui-même est sissu de l'expérience pratique et théorique des enquêteurs.

Enfin, il est possible d'identifier les tâches pouvant être automatisées par des méthodes computationnelles et également de guider leurs applications par le choix et la recodification des variables. Il existe ainsi un premier filtre automatique qui effectue une classification générale en « cambriolage d'habitation » et un second filtre qui classifie les cambriolages d'habitation à l'aide des codes phénomènes. C'est ce second filtre qui codifie les cambriolages d'habitation que nous allons tenter d'automatiser à l'aide d'une technique computationnelle. L'un des avantages de cette (semi-) automatisation est de soutenir les analystes sur les nouveaux cas quotidiens à codifier, en leur procurant ainsi un gain de temps appréciable.

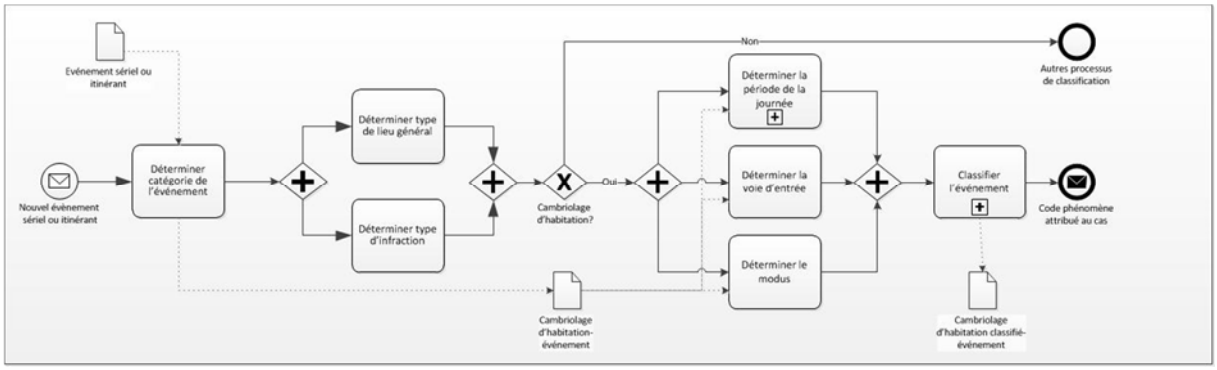


***Figure 20** : Processus BPMN de classification des cambriolages d'habitation*

8.3. Méthode

8.3.1. Échantillon

L'échantillon considéré provient de la base de données PICAR qui répertorie tous les événements reportés à la police et qui sont relatifs à la délinquance sérielle et itinérante (voir chapitre 3.2.1 pour la description détaillée). Il ne s'agit pas uniquement d'infractions clairement définies, mais il peut également s'agir de signalement, de découverte de véhicule volé ou encore de tentatives. Nous considérons tous les cambriolages d'habitations enregistrés dans PICAR en 2008 sur le canton de Vaud, ce qui représente 4'171 événements.

8.3.2. Variable dépendante

Dans les cas d'apprentissage supervisé où une variable cible est nécessaire, nous avons pris en considération le code CICOP qui est déterminé par les analystes en vue de classifier l'événement dans PICAR. La liste et le descriptif de ces codes sont disponibles à l'Annexe 17.

8.3.3. Variables indépendantes

Ces variables ont pu être identifiées en étudiant les processus formalisés de raisonnement utilisés par les analystes. Cet examen indique que la variable cible, le code CICOP, peut être déterminée par les valeurs de ces variables.

- *Type de lieu*, *voie d'introduction*, et *mode opératoire* :

Nous avons repris la classification opérée dans PICAR pour ces trois variables[33]

- *Jour de la semaine* :

Pour tous les événements dont l'intervalle de temps est de 24 h ou moins, nous avons déterminé le jour de la semaine à partir de la date de l'heure médiane.

- *Jours fériés* :

Les jours fériés sont définis par les samedis et dimanches, ainsi que tous les jours fériés officiels du canton de Vaud en 2008.

- *Saison* :

Nous avons déterminé la saison à partir de la date de début de l'événement.

- *Moment de la journée* :

Afin de déterminer le moment de la journée d'un événement, nous n'avons considéré que les cas où l'intervalle d'heure ne dépassait pas 24 h, la mesure devenant impossible au-delà et la variable non pertinente. Nous avons déterminé l'heure médiane des événements, pour ensuite la placer dans des bornes fixes délimitant trois moments de la journée. Ces bornes diffèrent selon les périodes de l'année où l'heure d'été et l'heure d'hiver sont appliquées (Tableau 2).

***Tableau 2** : Détermination du moment de la journée*

Moment de la journée	***Période de l'année***	
	Heure d'hiver	*Heure d'été*
Jour	7-18h	6-21h
Soir	18-23h	21-23h
Nuit	23-7h	23-6h

[33] Voir Annexes 18,19 et 20

Lorsqu'un événement est classifié à l'aide des codes phénomènes du CICOP, les modes opératoires et les voies d'introduction ne sont pas spécifiés dans PICAR, car ils font partie intégrante des critères de classification. Il en découle que nous avons dû manuellement rajouter les modes opératoires et les voies d'introduction manquants à partir des codes phénomènes, ceci afin de faciliter les analyses ultérieures. Cela introduit bien évidemment un biais qu'il faudra garder à l'esprit lors de l'interprétation des résultats. Une des solutions pour contourner cette limite serait de recourir à des techniques de *text mining* directement sur le texte libre lié aux événements (p. ex. description de l'événement). Cependant, nous n'avions pas accès à ces données dans le cadre de cette recherche pour des raisons de confidentialité.

8.3.4. Stratégie de classification automatique : le réseau neuronal « perceptron multicouche »

La technique choisie ici relève des réseaux neuronaux. Nous avons fait ce choix car cette méthode est particulièrement pertinente dans notre contexte. En effet, elle relève des méthodes computationnelles d'apprentissage supervisé, c'est-à-dire qu'elle cherche à trouver la valeur d'une variable cible particulière, en l'espèce le code CICOP. Elle est susceptible d'apparaître comme complexe et peu intuitive, notamment à cause de l'aspect « boîte noire » du processus (Figure 21). Elle demeure néanmoins incontournable dans le domaine du data mining de par sa capacité à gérer des problèmes complexes et sa puissance de modélisation.

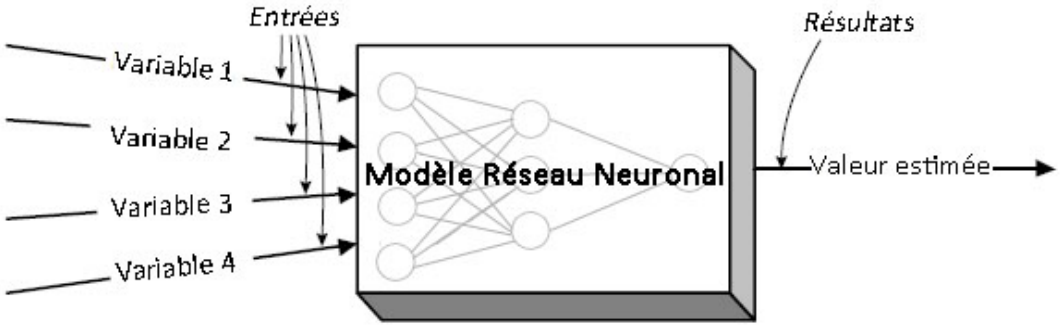


***Figure 21** : Principe d'un réseau neuronal (traduit de Berry & Linoff, 2004)*

Le principe de fonctionnement de cette technique est, de manière simplifiée, semblable au comportement des neurones biologiques présents chez les êtres vivants. Fondamentalement, chaque réseau est constitué d'unités qui, une fois activées par

certaines informations en entrée, produisent un résultat. Selon la complexité du réseau, il peut y avoir plusieurs couches de neurones et plusieurs résultats possibles.

Afin de tester les capacités de classification automatique sur notre jeu de données, nous avons utilisé un algorithme classique de réseau neuronal appelé « perceptron multicouche ». Concrètement, l'algorithme fonctionne en deux temps avec une procédure d'entraînement et de test. Il apprend à classifier les événements en code CICOP en fonction des variables indépendantes sur les 2/3 des données (n= 2'894), puis il teste le modèle obtenu sur les données restantes (n= 1'277). Le taux de précision est calculé en comparant avec la classification réalisée par les analystes du CICOP.

8.4. Résultats

Le Tableau 3 nous indique le taux de précision de l'algorithme par rapport à la classification manuelle des événements, tandis que le Tableau 4 représente la matrice de confusion. Il est remarquable de constater que ce taux est relativement élevé pour la majorité des catégories. Environ 85% des classifications effectuées à partir des règles apprises par l'algorithme correspondent à la classification du CICOP. Il est également intéressant de noter que 67.7% des erreurs de l'algorithme sont liés à la catégorie des événements non codifiés par l'algorithme, ou par les analystes du CICOP (Tableau 4). Les classifications les moins précises (« Giorno » et « Notte ») correspondent à des catégories générales, caractérisées uniquement par la période de la journée durant laquelle s'est déroulé le cas, ce qui peut expliquer la faible performance de l'algorithme. D'autant plus que presque tous les événements codifiés « Giorno » ou « Notte » par les analystes et qui ont été mal classés par l'algorithme se retrouvent dans la catégorie des événements non classifiés.

***Tableau 3** : Taux de précision d'un réseau neuronal dans la classification de cambriolages d'habitations (n = 1'277)*

Code CICOP	***Réseau neuronal*** [34]
HALL *(cambriolage de jour par introduction clandestine et fouille du hall d'entrée, n=68)*	100.00%
NOTTE CHIGNOLE *(cambriolage de nuit par la fenêtre avec une chignole, n=12)*	100.00%
GIORNO CILINDRO *(cambriolage de jour par la porte avec arrachage du cylindre, n=232)*	98.90%
GIORNO PIATTO *(cambriolage de jour par la porte avec outil plat, n=147)*	98.20%
GIORNO EPAULEE *(cambriolage de jour par la porte à coups d'épaule, de pieds, ou de vive force, n=29)*	92.00%
GIORNO CHIAVE *(cambriolage de jour avec introduction clandestine avec clé trouvée, n=6)*	90.90%
GIORNO FINESTRA *(cambriolage de jour par la fenêtre, n=258)*	88.90%
Non classifié *(n=287)*	82.60%
NOTTE FINESTRA *(cambriolage de nuit par la fenêtre, n=41)*	67.60%
SERA *(cambriolage du soir sans détail, n=99)*	65.70%
SERA BLOKO *(cambriolage du soir en bloquant la porte d'entrée, n=4)*	60.00%
NOTTE CILINDRO *(cambriolage de nuit par la porte avec arrachage du cylindre, n=12)*	57.10%
NOTTE *(cambriolage de nuit sans détail, n=41)*	18.90%
GIORNO *(cambriolage de jour sans détail, n=41)*	0.00%
Total	***84.5%***

[34] Calculé avec le logiciel SPSS

***Tableau 4** : Matrice de confusion du réseau neuronal perceptron multicouches (n =1'277). Les classifications correctes sont indiquées en vert et les classifications erronées en rouge.*

	Prévisions														
	GIORNO	GIORNO CHIAVE	GIORNO CILINDRO	GIORNO EPAULEE	GIORNO FINESTRA	GIORNO PIATTO	HALL	NOTTE	NOTTE CHIGNOLE	NOTTE CILINDRO	NOTTE FINESTRA	SERA	SERA BLOKO	Non classifié	Observations
	4	0	1	0	0	0	0	2	0	0	0	3	0	31	**GIORNO**
	0	6	0	0	0	0	0	0	0	0	0	0	0	0	**GIORNO CHIAVE**
	0	0	230	0	0	0	0	0	0	2	0	0	0	0	**GIORNO CILINDRO**
	0	0	1	27	0	0	0	0	0	0	0	0	0	1	**GIORNO EPAULEE**
	0	0	0	0	243	0	0	0	0	0	6	8	0	1	**GIORNO FINESTRA**
	1	0	0	0	0	144	0	0	0	0	0	0	0	2	**GIORNO PIATTO**
	0	0	0	0	0	0	68	0	0	0	0	0	0	0	**HALL**
	0	0	0	0	0	1	0	16	0	0	0	2	0	22	**NOTTE**
	0	0	0	0	0	0	0	0	12	0	0	0	0	0	**NOTTE CHIGNOLE**
	0	0	6	0	0	0	0	0	0	6	0	0	0	0	**NOTTE CILINDRO**
	0	0	0	0	9	0	0	0	0	0	25	6	0	1	**NOTTE FINESTRA**
	0	0	0	0	10	0	0	1	0	0	3	62	0	23	**SERA**
	0	1	0	0	0	0	0	0	0	0	0	1	1	1	**SERA BLOKO**
	4	2	0	1	9	1	0	14	0	0	2	19	0	235	**Non classifié**

8.5. Discussion

Les résultats observés au niveau de la classification des cambriolages d'habitations se révèlent intéressants pour plusieurs raisons. En premier lieu, on constate que le taux de précision est relativement élevé avec la méthode de classification automatique testée. Cela nous amène à penser que l'algorithme est capable de reconstituer de manière précise et complète la classification opérée par les analystes du CICOP. Néanmoins, il n'est pas possible à ce stade de se prononcer sur la cohérence de cette classification par les analystes, puisqu'une partie des variables a été rajoutée manuellement à partir des codes CICOP (qui sont la cible de cette classification). L'autre point d'intérêt relève de l'analyse des erreurs. Comme suggéré par les hypothèses, les codes CICOP caractérisés par un nombre moins élevé de variables sont moins facilement classifiés par l'algorithme. Il y a deux raisons possibles à cela : la première est que certains codes CICOP sont des codes par défaut, qui sont appliqués

lorsqu'aucun autre code ne peut l'être. Citons par exemple le code GIORNO qui est appliqué sur les cambriolages d'habitation commis le jour, mais qui ne correspond à aucun autre code CICOP. Il est donc possible que l'algorithme de data mining éprouve des difficultés à comprendre la logique de classification derrière ce raisonnement implicite de l'analyste. Il faut clairement considérer le code GIORNO comme une catégorie plus générale qui englobe les autres codes CICOP relevant des cambriolages de jour. La seconde raison concerne le faible nombre de cas de certains phénomènes, notamment dans les SERA et les NOTTE, qui sont relativement peu nombreux comparé aux cas GIORNO. La seule exception est le phénomène NOTTE CHIGNOLE qui peut s'expliquer par le fait que le mode opératoire « chignole » est spécifique à ce type de phénomène et ne se retrouve dans aucun autre. Une autre explication possible est la nature approximative de la détermination du moment de la journée. En effet, comme nous l'avons déjà évoqué, les intervalles horaires ne permettent pas toujours de cibler avec précision le moment de la journée où s'est déroulé le cas. Ce problème est particulièrement visible avec les SERA et les NOTTE dont la période de transition est relativement rapprochée.

Cette expérimentation reste néanmoins exploratoire en regard des objectifs plus généraux du projet sur le potentiel du data mining dans de tels processus. Elle indique néanmoins la possibilité de soutenir, par un moyen automatisé, l'encodage des données, et de procurer ainsi un gain de temps non négligeable dans la mise en œuvre du processus, un processus plus complet, ainsi que davantage de fiabilité. Ces résultats ouvrent aussi des perspectives dans l'utilisation d'approches qui ne nécessitent pas une classification déterministe des cas, mais qui mémorise plutôt des degrés d'appartenance à une ou plusieurs catégories. Ces approches offrent évidemment davantage de souplesse pour l'analyse, car elles laissent la possibilité d'interpréter les cas sous différentes hypothèses. Ce développement avait été partiellement testé par Ribaux (1997) qui avait considéré, au moyen d'ensembles flous, simultanément plusieurs catégorisations possibles d'un cas en fonction du moment de la journée.

Une fois les événements codifiés selon une approche situationnelle, il est maintenant possible de les monitorer afin de détecter des tendances dans les données de la criminalité.

9. Détecter : tendances des activités criminelles et patterns de rupture

La seconde étape d'application de l'approche développée illustre l'analyse de tendances dans les activités criminelles et se focalise sur la détection des patterns de données (voir chapitre 5). La détection s'impose comme l'une des étapes clés du processus de veille opérationnelle, car c'est cette opération qui fait office de déclencheur pour les étapes ultérieures d'analyse et de prise de décision. Souvent, les analystes se rendent compte *a posteriori* qu'ils auraient dû détecter plus rapidement des changements dans les répétitions criminelles, par exemple une nouvelle série, un passage d'un groupe d'auteurs dans une région ou le désistement de groupes criminels.

9.1. Détection computationnelle des patterns de ruptures dans les tendances

Le but du composant computationnel est de concevoir une méthode calculatoire capable de détecter les changements dans les données de la criminalité de la manière la plus automatique possible en se basant sur la classification situationnelle utilisée par l'unité d'analyse.

Par hypothèse, une méthode automatique présente les avantages suivants : (1) elle peut détecter les changements dans les tendances plus rapidement qu'une détection humaine et à partir d'un plus grand jeu de données, au-delà de toutes capacités cognitives humaines ; (2) elle est capable de détecter des patterns inattendus ; (3) elle peut s'appliquer selon une grande variété de dimensions, telles que les fréquences d'événements, l'information spatio-temporelle, les caractéristiques du mode opératoire ou les données forensiques ; et (4) elle peut réaliser différentes analyses simultanément en variant les niveaux d'agrégation (par ex. temporel, spatial, granularité des modes opératoires).

L'expérience qui va suivre tend à investiguer les enjeux soulevés par les tendances des activités criminelles. Il s'agira notamment de déterminer si l'application de modèles computationnels simples en complément de la méthodologie existante peut potentiellement améliorer la détection automatique de changements de tout type de données. De tels patterns peuvent être définis comme étant des « patterns de rupture ».

9.2. Processus de détection

De la même manière qu'au chapitre 8.2, c'est à travers une observation en immersion réalisée avec les analystes de la division Coordination judiciaire de la Police cantonale vaudoise et une analyse documentaire que le processus permettant de comprendre la détection de tendances a pu être formalisé à l'aide de la notation BPMN. Comme constaté à la Figure 22, on distingue trois séquences qui permettent d'aboutir à la détection d'une tendance dans les données de la criminalité. La première est la réaction face à une information transmise par des partenaires externes à la division. Il peut s'agir par exemple d'un enquêteur d'une autre division ou d'une autre brigade, d'une communication d'une autre police cantonale, ou encore d'un message provenant d'entités externes à la police (p. ex. Corps des gardes-frontière). La seconde séquence intervient lors de l'intégration des nouveaux événements dans PICAR (voir chapitre 8.2). Durant cette étape, les analystes réalisent une activité de monitoring au fur et à mesure que les événements sont traités. Lorsque des événements insolites ou qu'une succession inhabituelle d'événements similaires se produit, le monitoring des analystes est susceptible de mener à une détection de tendances. Finalement, la dernière séquence identifiée intervient lorsque les analystes effectuent une extraction de PICAR en vue de réaliser leur rapport hebdomadaire et qu'ils analysent les données sans *a priori*. Comme pour l'activité de monitoring, la présence d'événements insolites ou d'une succession inhabituelle d'événements similaires est susceptible de mener également à une détection de tendances. Dans tous les cas, une fois la tendance détectée, les analystes vont évaluer la pertinence de celle-ci en recherchant des cas similaires dans la base de données et posant des hypothèses susceptibles de l'interpréter. Le cas échéant, la tendance peut ainsi se confirmer et aboutir à une diffusion de l'information aux entités pouvant entreprendre des actions adaptées. Elle peut également se révéler non pertinente pour différentes raisons. La tendance peut être considérée comme triviale si elle est déjà connue de l'unité d'analyse. Il peut aussi s'agir d'une tendance incohérente qui ne peut être expliquée à l'aune d'une analyse forensique et criminologique et qui relèverait plutôt d'un phénomène statistique. Il est donc crucial de bien distinguer l'étape de détection de l'étape d'analyse. Au moment de la détection, il n'est pas encore possible de savoir s'il s'agit effectivement d'une tendance criminelle.

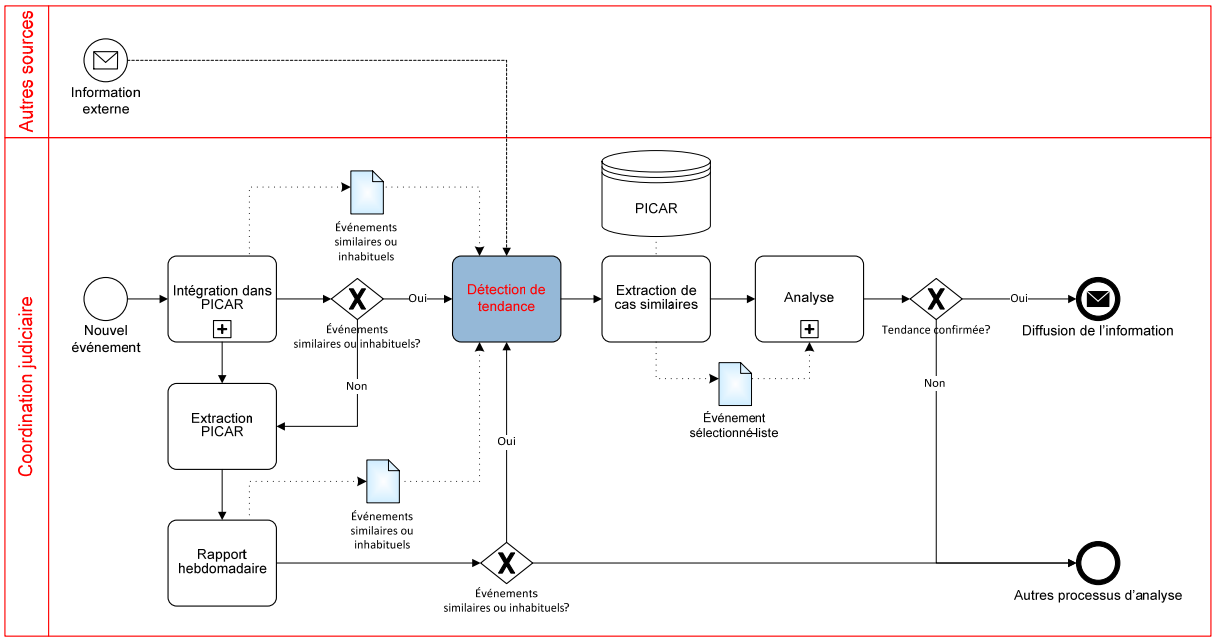


Figure 22 *: Processus BPMN du déclenchement de la détection de tendances dans les données de la criminalité au sein de la coordination judiciaire.*

Les patterns de rupture susceptible d'être détectés par l'unité d'analyse peuvent se traduire par cinq types de patterns (Figure 23).

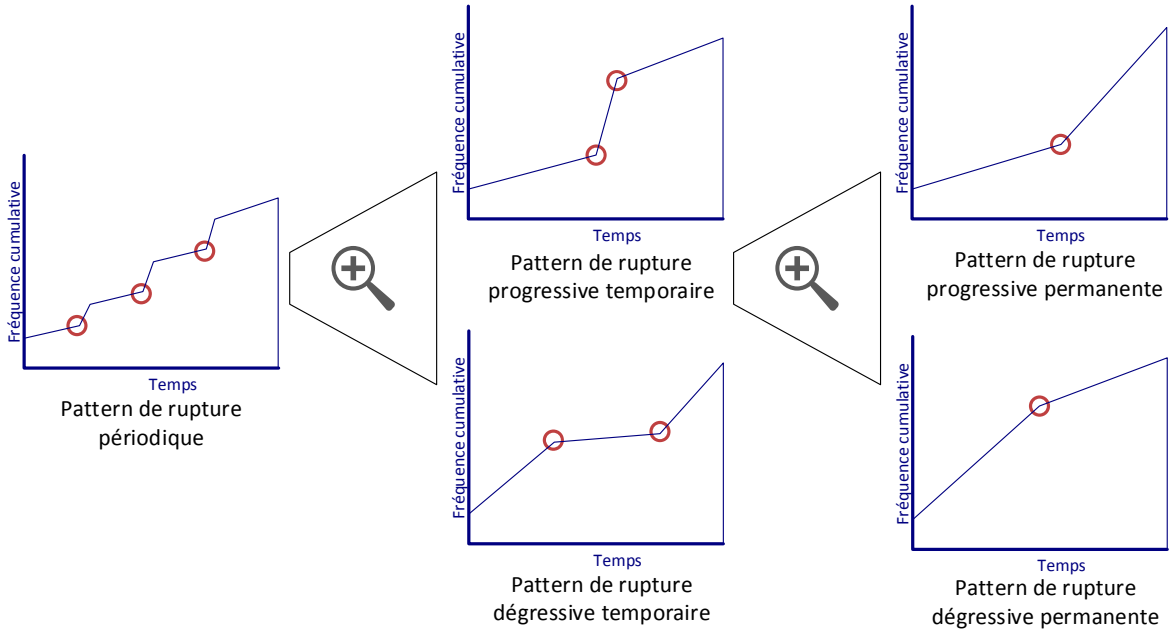


Figure 23 *: Type de patterns de rupture : À l'aide d'une visualisation des données sous forme de fréquences cumulatives, il est possible de catégoriser les patterns de rupture en cinq types. Le pattern de rupture périodique correspond à des augmentations ou diminutions régulières d'un phénomène criminel. Le pattern de rupture temporaire représente une augmentation (rupture progressive) ou une diminution (rupture dégressive) temporaire. Le pattern de rupture permanente représente une augmentation (rupture progressive) ou une diminution (rupture dégressive) qui se maintient dans le temps formant ainsi une nouvelle tendance stable. Dépendamment du niveau d'agrégation, une rupture permanente peut s'intégrer dans une rupture temporaire, qui elle-même peut faire partie d'une rupture périodique.*

La rupture périodique indique des fluctuations régulières dans la fréquence d'un phénomène criminel. Les cambriolages du soir présenté au chapitre 5.3 en sont un cas d'école. Pour ce type de patterns, l'enjeu de la détection prend deux formes différentes. La première forme est la détection de la périodicité du phénomène. Le phénomène est susceptible d'être connu par les analystes, mais sa régularité peut ne pas avoir été remarquée ou alors le phénomène a évolué et la périodicité apparaît alors comme nouvel attribut. La seconde forme est la détection d'anormalité au sein de la périodicité. Dans ce cas, la régularité dans la fréquence du phénomène est connue, mais une fréquence anormale est observée. Par exemple, dans le cas des cambriolages du soir (Figure 24), l'augmentation observée durant l'hiver 2011 apparaît comme inhabituelle par rapport aux autres hivers.

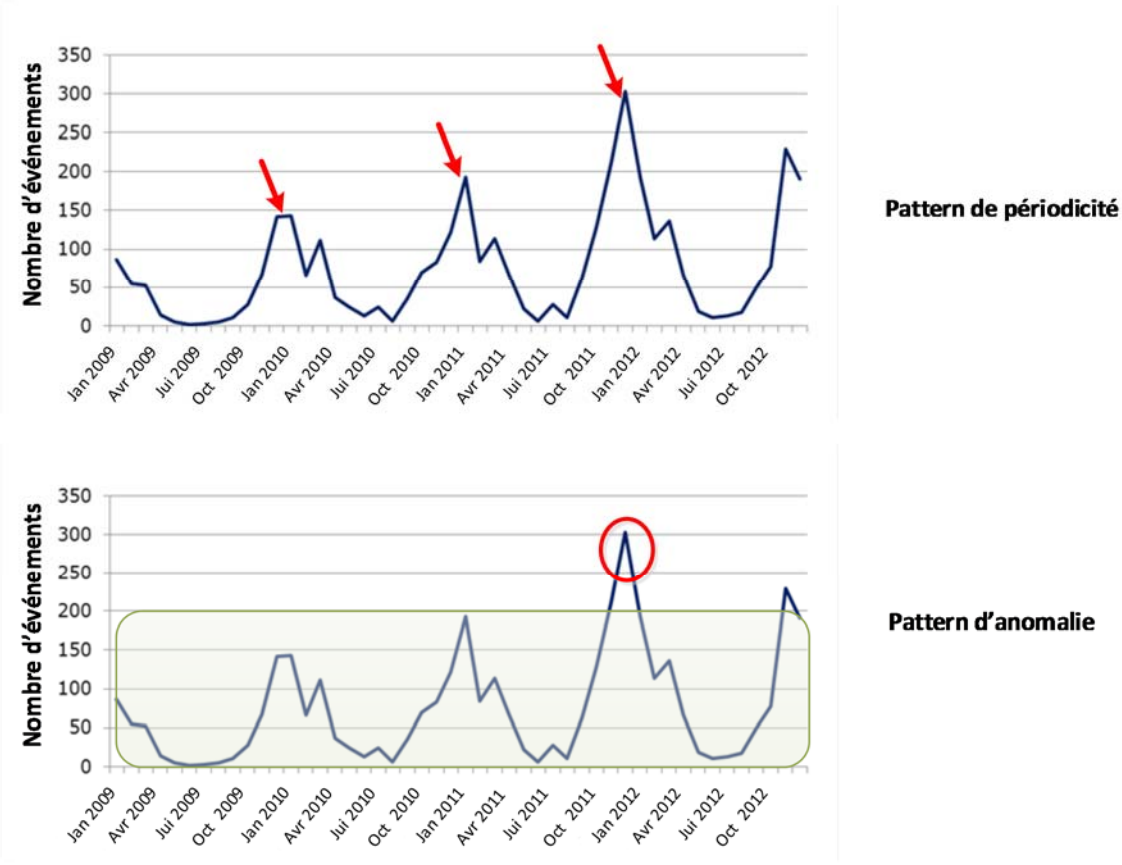


***Figure 24** : Différents types de pattern dans l'évolution des cambriolages du soir : un premier type de pattern est détecté avec l'augmentation périodique durant l'hiver de ce phénomène (en haut). Cette périodicité définit ensuite un nouvel état normal de la tendance. Un second pattern peut finalement être détecté avec une augmentation anormale des cambriolages par rapport à la normalité précédemment établie (en bas).*

Les ruptures temporaires progressive et dégressive définissent des augmentations ou des diminutions temporaires dans la fréquence d'un phénomène criminel. Cela peut résulter de l'activité d'un auteur ou d'un groupe d'auteurs prolifique qui agit sur une période temporelle relativement circonscrite. Une fois les auteurs partis ou neutralisés, la tendance retourne à sa fréquence initiale.

Les ruptures permanentes progressive et dégressive indiquent des augmentations et des diminutions dans la fréquence d'un phénomène criminel qui demeurent stables après leur apparition. L'exemple de l'arrivée de cambrioleurs géorgiens en Suisse romande présenté au chapitre 5.4 illustre bien ce type de pattern de rupture. La nouvelle tendance apparaît alors comme la nouvelle normalité du phénomène dès lors qu'elle perdure dans le temps.

La frontière entre ces catégories peut être floue si l'on considère le niveau d'agrégation sous lequel les données sont observées. Un pattern de rupture permanente peut se révéler être une partie d'un pattern de rupture temporaire à un niveau d'agrégation supérieur. De même, ce dernier peut être intégré dans un pattern de rupture périodique à niveau encore supérieur. Ces patterns sont susceptibles d'être générés par des changements résultants des individus (l'activité des auteurs) et/ou des changements dans l'environnement (modification des opportunités criminelles ou de l'activité policière). Une analyse forensique et criminologique peut alors fournir des hypothèses d'explication afin de caractériser les causes d'un pattern de rupture et ainsi guider le choix des niveaux d'agrégation.

En résumé, les enjeux liés à la détection des tendances des activités criminelles peuvent être diligenté par les questions suivantes : (a) définir la normalité de la tendance du phénomène criminel considéré, (b) détecter les ruptures dans la tendance normalisée, et (c) analyser les patterns anormaux détectés.

9.3. Méthode

9.3.1. Échantillon

L'échantillon considéré provient de la base de données PICAR qui répertorie tous les événements reportés à la police et qui sont relatifs à la délinquance sérielle et itinérante (voir chapitre 3.2.1 pour la description détaillée). Il ne s'agit pas uniquement d'infractions clairement définies, mais il peut également s'agir de signalements, de découvertes de véhicules volés ou encore de tentatives. Nous considérons tous les

événements enregistrés dans PICAR entre le 6 janvier 2014 et le 30 juin 2015 sur le territoire du CICOP (les cantons de Vaud, Genève, Fribourg, Valais, Neuchâtel, et Jura), ce qui représente 72'727 événements. Parallèlement, nous avons collecté toutes les tendances détectées par l'unité d'analyse entre Mai 2014 et Juin 2015, ce qui représente 27 détections. Pour tester la capacité d'adaptation d'un algorithme de détection automatique, d'autres sources d'information complémentaire ont également été exploitées, à savoir les saisies de faux documents d'identités et les traces de souliers.

9.3.2. Stratégie de détection humaine : procédure de collecte

Afin de représenter le plus fidèlement possible l'activité de détection de tendances au sein d'une unité de renseignement criminel opérationnel, une procédure de collecte a été mise au point en collaboration avec la Division Coordination judiciaire de la Police cantonale vaudoise. Un module « Suivi des tendances » développé sous FileMaker fut adjoint sur l'interface utilisé par l'unité d'analyse. Une capture d'écran du module permet d'avoir un aperçu de cette interface (Figure 25). Ce module se caractérise par deux parties distinctes : la description de la tendance et la détection de la tendance.

Concernant la description de la tendance, les variables suivantes sont enregistrées :

- *Numéro de tendance*
- *Titre*
- *Texte libre*
- *Phénomène/type d'auteurs*
- *Mode opératoire*
- *Type de lieux général*
- *Type de lieux détail*

En plus de l'identifiant unique de la tendance et de son titre, le texte libre permet une description plus détaillée notamment sur le type d'événements, le type de changement et la quantité de cas concernés. Le phénomène, le mode opératoire et le type de lieux général/détail reprennent les mêmes modalités que celles utilisées par le CICOP[35].

[35] Voir Annexes D pour le détail.

Concernant la détection de la tendance, les variables suivantes sont enregistrées :

- *Date de détection*

Date à laquelle la tendance a été détectée par le coordinateur

- *Coordinateur*

Le nom du coordinateur qui a détecté la tendance. Il est ainsi possible de se renseigner au cas où des informations ou des précisions supplémentaires seraient nécessaires.

- *Mode de détection*

Il s'agit de la manière dont la tendance a été détectée. Conjointement avec l'unité d'analyse, trois modalités ont été définies (Tableau 5). Celles-ci vont d'un mode de détection plus proactif (*Analyse coordi*) à un mode plus réactif (*Info extérieure*) en passant par un mode intermédiaire de veille (*Monitoring coordi*).

- *Originalité de la tendance*

L'originalité correspond au fait de savoir si une tendance a déjà été détectée par le passé (*connue*) ou s'il s'agit d'une tendance inédite au sein de la coordination judiciaire (*inconnue*). De plus, si la tendance est connue, il reste à déterminer si elle est connue au niveau cantonal (*VD*), au niveau romand (*CICOP*) ou au niveau suisse (*CH*).

- *Type de changement*

Concernant la manière dont la tendance évolue, trois types de changement peuvent être enregistrés : *l'augmentation ou l'apparition* d'un phénomène, *la diminution ou la disparition* d'un phénomène et *les autres changements* qui ne rentreraient pas dans les deux premières catégories.

- *Variables de détection*

Il s'agit de la ou des variable(s) qui ont permis la détection de la tendance. Il y a six types de variables possibles : *mode opératoire, phénomène, spatial, trace, type d'auteurs* et *type d'événements*.

***Tableau 5** : Le mode de détection de la tendance et ses différentes modalités détaillées*

Mode de détection		
Analyse coordi	***Monitoring coordi***	***Info extérieure***
Mise en évidence d'une tendance suite à une analyse. Pas de sentiment préalable avant la découverte de la tendance. *Exemple : Extraction de PICAR sur une période, pour la tendance hebdomadaire, et mise en évidence d'un phénomène particulier non repéré avant*	Impression générale ou discussion au rapport. *Exemple : En enregistrant une série de cas, un coordinateur soupçonne une activité inhabituelle et décide d'effectuer une analyse plus poussée.*	Réception d'une information externe à la coordination judiciaire du canton de Vaud. *Exemple : Un responsable de région qui transmet une information à la coordination judiciaire sur un événement particulier qui touche sa région.*

Concrètement, dès qu'un coordinateur détecte une tendance :

1. il avise le coordinateur responsable du module de suivi de tendance
2. il remplit une fiche basée sur le modèle de l'interface FileMaker[36] à l'aide du coordinateur responsable
3. la tendance est introduite dans le module « suivi de tendance » par le coordinateur responsable et un analyste du CICOP

Ce protocole permet d'assurer autant que possible une fiabilité dans l'enregistrement des tendances dans le module. La centralisation de la saisie sur un coordinateur responsable évite une partie des variations liées aux erreurs de saisie. En remplissant la fiche de description et de détection de tendance en binôme avec le coordinateur ayant détecté la tendance, le coordinateur responsable s'assure ainsi une certaine standardisation. De plus, la présence d'un analyste du CICOP au moment de remplir la base de données renforce cette standardisation et limite ainsi la subjectivité inhérente à ce type de collecte de données.

[36] Voir Annexe E pour le modèle de la fiche

Figure 25 *: Capture d'écran du module FileMaker « suivi des tendances » : les tendances détectées par les analystes sont enregistrées par l'intermédiaire de cette interface.*

9.3.3. Stratégie de détection automatique : l'analyse de changement de points

Afin de pouvoir éprouver l'approche méthodologique destinée à guider l'intégration de méthodes computationnelles en analyse criminelle, il faut à nouveau opérer des choix quant au type de techniques à utiliser. L'objectif est de pouvoir détecter des changements dans une série temporelle en supposant une certaine régularité dans son évolution. Les techniques d'analyse de séries temporelles semblent alors adéquates pour répondre à ce type d'opération. Mais une fois encore, les praticiens et académiciens sont susceptibles de se perdre dans la diversité des méthodes destinées à détecter des changements abrupts dans des données de séries temporelles (Gustafsson, 1996; Hido, Idé, Kashima, Kubo, & Matsuzawa, 2008; Kawahara & Sugiyama, 2009; Liu, Yamada, Collier, & Sugiyama, 2013; Yamada, Kimura, Naya, & Sawada, 2013, entre autres).

L'analyse de séries temporelles est souvent considérée au niveau de l'analyse stratégique et ses applications sont bien établies dans les études longitudinales en criminologie (Bennett, 1991; Cantor & Land, 1985; Greenberg, 2001) et plus généralement en sciences sociales (Box-Steffensmeier, Freeman, Hitt, & Pevehouse, 2014). Cependant, l'utilisation de l'analyse de séries temporelles dans le contexte de l'analyse et du renseignement criminel est sous-estimée et il n'existe pratiquement aucune application destinée à soutenir opérationnellement des unités d'analyse criminelle en intégrant des composants forensiques et dans une optique de résolution de problèmes. Dans ce contexte, il ne s'agit pas de considérer des tendances stratégiques de la criminalité sur plusieurs années, mais plutôt des tendances opérationnelles dans les activités criminelles sur plusieurs mois ou semaines (Tableau 6). Les frontières entre ces deux catégories de tendances sont cependant poreuses, dès lors que l'analyse de tendances de la criminalité peut amener à détecter des problèmes au niveau opérationnel et que l'analyse des tendances des activités criminelles est susceptible d'alimenter l'étude de l'évolution de la criminalité. Néanmoins, à l'inverse de son application sur les données historiques en criminologie, l'analyse de séries temporelles en renseignement criminel nécessite un suivi en temps réel pour être capable de déployer des réponses au temps opportun.

***Tableau 6** : Distinction entre les types de tendances dans les données de la criminalité.*

Type d'analyse	**Tendances de la criminalité**	**Tendances des activités criminelles**
Niveau d'analyse	Stratégique	Opérationnel
Objet	Évolution de la criminalité	Activités répétitives
Objectif	Détection de changement à long terme	Détection de changement à court et moyen terme
Approche	Criminologie	Résolution de problème
Temporalité	Années	Mois/semaines
Spatialité	International et national	Local et régional
Méthode	Rétrospective	Veille
Exemple	L'analyse de l'évolution des homicides par armes à feu durant les 50 dernières années en Suisse	L'analyse et le suivi des cambriolages d'habitation par arrachage de cylindre par semaine en Suisse romande

Parmi les différentes méthodes d'analyse de séries temporelles existantes, nous avons sélectionné une technique bien établie en informatique (Fuchs, Gruber, Nitschke, & Sick, 2010). Cette approche est basée sur la segmentation en temps réel des séries temporelles à l'aide d'approximations polynomiales des moindres-carrés. Appliquée concrètement à notre problème de criminalité, cette méthode définit des points de segmentation qui reflètent potentiellement un changement situationnel dans le phénomène criminel considéré. Ces points sont déterminés par des critères et des seuils de segmentation, en utilisant une combinaison linéaire de polynôme pour approximer les séries.

Nous avons fait ce choix car cette technique présente deux avantages principaux qui sont susceptibles de répondre aux exigences d'une application orientée vers la résolution de problèmes en analyse criminelle.

Le premier est représenté par deux propriétés importantes de la technique sélectionnée, à savoir le fait d'être adaptative et connectée (*online*). Ces propriétés permettent ainsi une utilisation dans un environnement connecté avec des adaptations rapides. Elles se réfèrent essentiellement au même objectif : le calcul efficient de l'approximation et sa mise à jour lorsque de nouveaux points sont disponibles. Cet objectif peut être atteint en implémentant un processus spécifique appelé *updating/downdating* (Elhay, Golub, & Kautsky, 1991). Ce processus utilise uniquement les nouvelles données pour mettre à jour les coefficients de la base polynomiale lorsque la fenêtre temporelle contenant les données à analyser est en train de changer. Actuellement, la plupart des méthodes ne sont pas adaptatives, c'est-à-dire qu'elles évaluent toutes les données de la nouvelle fenêtre temporelle sans considérer les précédents résultats calculés. Cela est difficilement applicable dans le contexte de l'analyse criminelle où les phénomènes criminels sont en constante évolution et la prise de décision requiert un traitement rapide des données. Ainsi, le côté adaptatif de cette méthode permet de tenir compte des précédentes données constituant la tendance des activités criminelles, tandis que son côté *online* permet d'analyser les données en temps réel sans avoir à attendre d'avoir l'ensemble des données à disposition. Ces propriétés s'avèrent donc cruciales dans une application opérationnelle orientée vers la résolution de problèmes en analyse criminelle.

Le second avantage repose sur le fait que les coefficients des fonctions de base sont porteurs de sens et facilement interprétables, non seulement dans leur pur sens mathématique, mais également en respect du domaine d'application. Cela est rendu possible en utilisant une base orthogonale spécifique construite à l'aide des polynômes discrets de Chebyshev. L'expansion orthogonale du polynôme approximé est définie par :

$$p(x) = \sum_{k=0}^{K} \frac{a_k}{\|p_k\|^2} p_k(x)$$

Les coefficients $\frac{a_0}{\|p_0\|^2}$, $\frac{a_1}{\|p_1\|^2}$ et $\frac{a_2}{\|p_2\|^2}$ (correspondant aux poids w_0 , w_1 et w_2, etc. pour w_3) de l'expansion peuvent être considérés comme des estimateurs optimaux (en termes de moindres-carrés) de la moyenne, de la pente et de la courbure, respectivement, des séries temporelles représentées par la variable x. Cette propriété rend l'interprétation de cette approximation relativement simple.

Ces deux avantages ont été cruciaux dans le choix de cette méthode et il est important de noter que les points de segmentation définis par l'algorithme dépendent de critères et de seuils qui sont dépendants uniquement des valeurs des séries temporelles, et non du domaine d'application.

Cependant, ce type de méthodes de segmentation présente intrinsèquement deux inconvénients. Premièrement, une difficulté repose sur la recherche du « bon » seuil pour le critère de segmentation, alors qu'il n'y a aucune méthode spécifique pour résoudre cette question, si ce n'est de se baser sur le rythme opérationnel d'un service d'analyse. Ensuite, ces méthodes présentent une certaine lacune dans leur flexibilité lorsqu'il s'agit de considérer les types d'entrées et de résultats de la segmentation. En effet, les données d'entrée de ces méthodes sont basées sur une seule valeur (y_i) par unité de temps (x_i), et les résultats prennent la forme de points de segmentation. Ces contraintes laissent peu de possibilités lorsque d'autres types d'entrées doivent être envisagés (p. ex. des données catégoriques ou floues au lieu de données numériques et nettes).

Une méthode de détection automatique a été élaborée par l'équipe de l'institut du management de l'information de l'Université de Neuchâtel sous la forme d'un programme en Python appelé *Fuzzy Change Points Detection* (FCPD). Ce programme permet de faire varier trois seuils afin de paramétrer la détection des tendances :

- ***Déviation de la valeur prédite (DPU).*** Ce premier critère correspond à la différence entre la valeur attendue de la courbe prédite et sa valeur réelle. Si le seuil est atteint, un point de segmentation est défini.

- ***Nombre de changement de signe de la pente (SSS).*** Le second critère correspond au nombre de fois que la pente de la série temporelle change de signe, c'est-à-dire si elle diminue ou augmente. De la même manière que le précèdent critère, lorsqu'un seuil est atteint, un point de segmentation est défini.

- ***Nombre de seuils à atteindre (SEG).*** Pour le dernier critère paramétrable, il faut décider du nombre de seuils à atteindre pour qu'un point de segmentation soit défini. Dans notre cas, il peut donc varier entre un et deux.

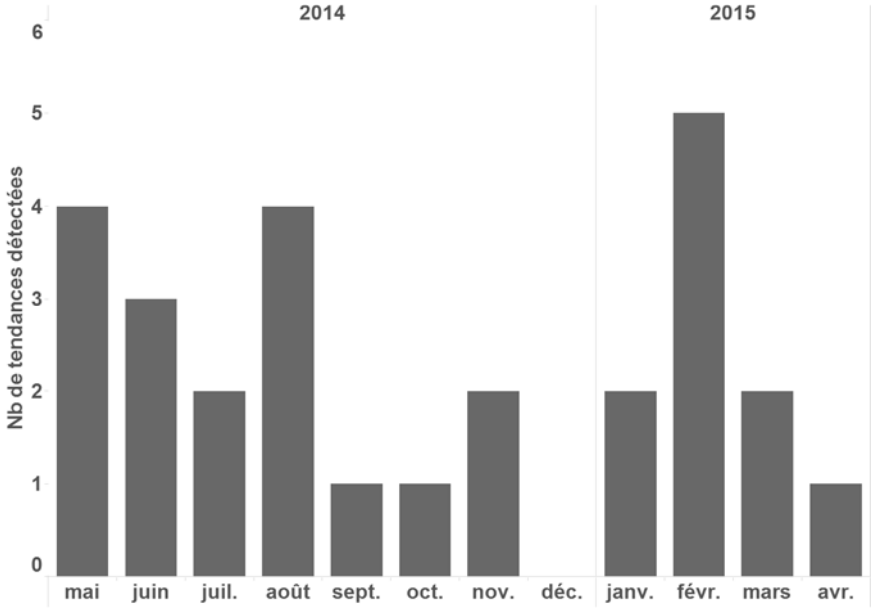
Ces paramètres sont susceptibles d'influencer la sensibilité de la détection et doivent ainsi être pris en compte dans l'application de l'algorithme.

9.4. Résultats

Les résultats sont présentés en quatre temps. Tout d'abord, une description des tendances détectées par l'unité d'analyse est présentée, puis la seconde partie illustre la capacité d'adaptation de l'algorithme en l'appliquant sur d'autres types de données. La troisième partie est consacrée à la comparaison entre la détection humaine et automatique. Et finalement, le dernier segment revient brièvement sur l'évaluation technique de la méthode de détection automatique.

9.4.1. Description des tendances détectées

Durant la période à l'étude, 27 tendances ont été détectées par l'unité d'analyse[37]. Le Graphique 9 indique la répartition des détections humaines sur la période à l'étude. Même s'il est difficile d'observer une orientation claire, on remarque un nombre plus faible de détection réalisée à la fin de l'année 2014. Une hypothèse serait que la charge de travail des analystes est orientée vers d'autres types d'activités en fin d'année, par exemple la préparation de bilans annuels. Moins de temps serait dévolu au monitoring et par conséquent moins de tendances seraient détectées. Une comparaison sur plusieurs années serait néanmoins souhaitable pour confirmer cette hypothèse.

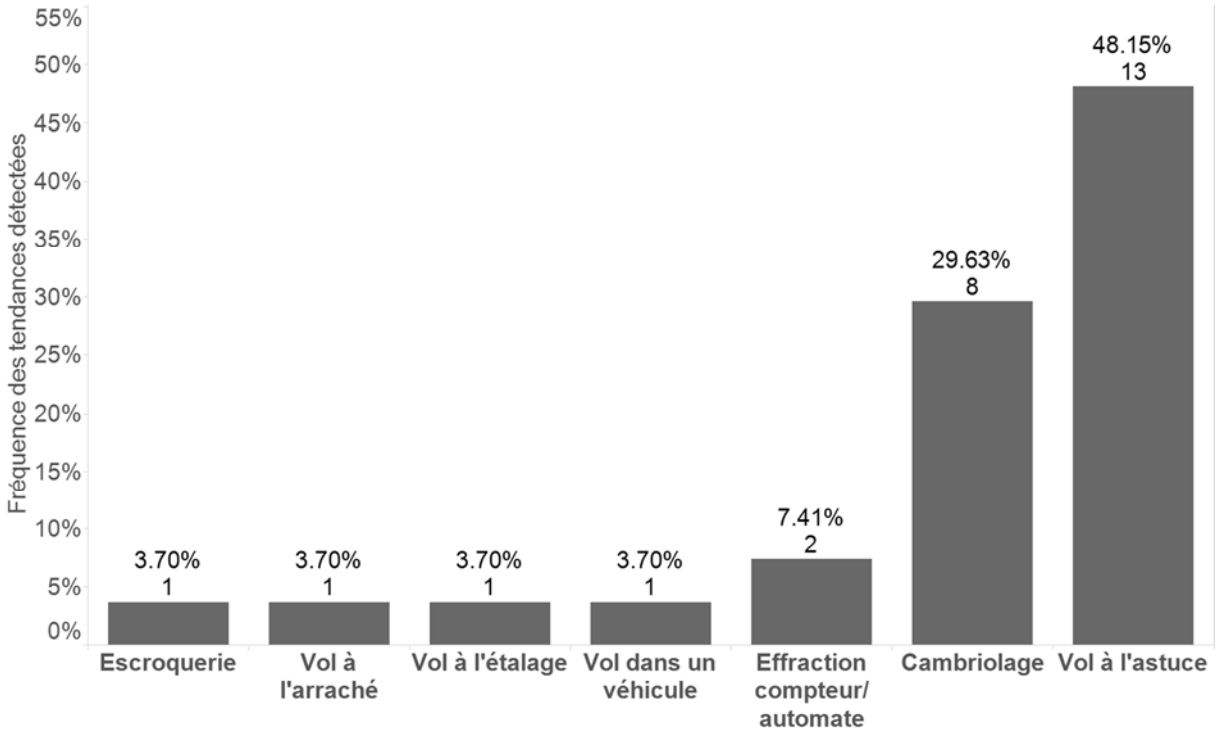


***Graphique 9** : Incidence des détections de tendances de l'unité d'analyse par mois (n =27).*

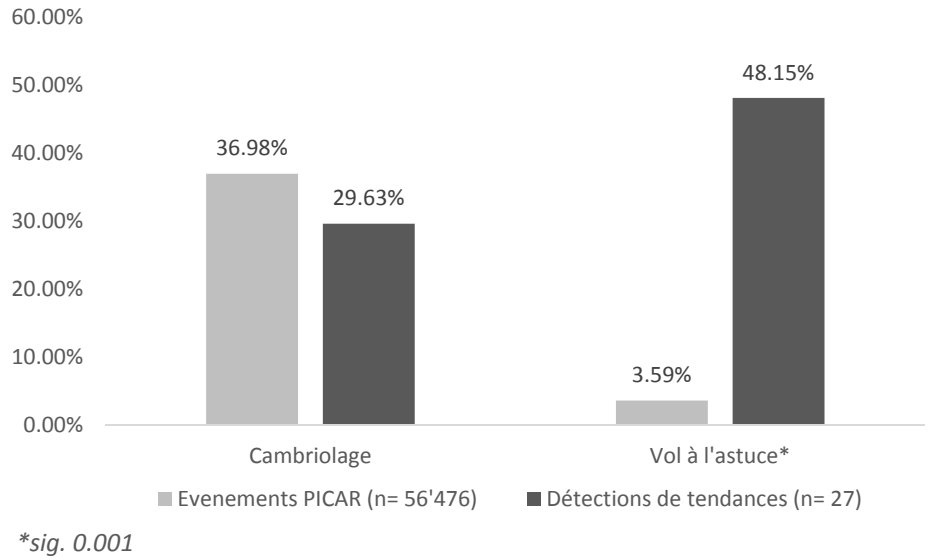
La majorité des tendances détectées concerne des vols à l'astuce (48.15%) et des cambriolages (29.63%) (Graphique 10). Ce n'est pas vraiment une surprise puisque ce sont précisément les types d'événements principaux qui sont enregistrés dans PICAR.

[37] Voir l'Annexe 22 pour les descriptions des tendances.

En revanche, en comparant avec la proportion d'événements enregistrés durant la même période (Graphique 11), on constate une différence significative pour les vols à l'astuce. Alors qu'ils représentent 3.59% des événements enregistrés dans PICAR, ils constituent 48.15% des détections de tendances.



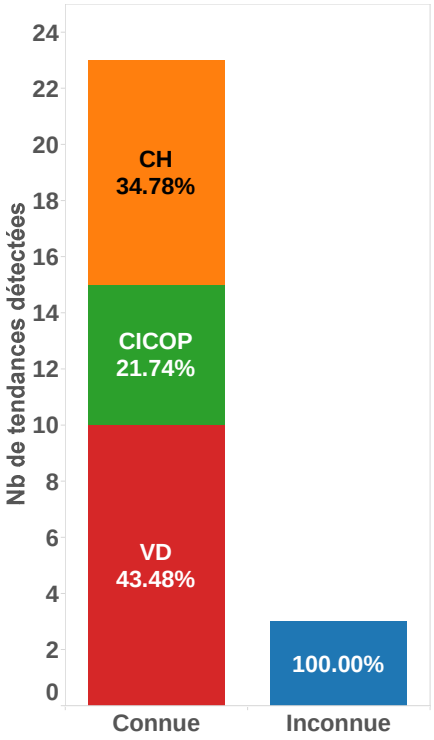
***Graphique 10** : Fréquence des détections de tendances de l'unité d'analyse en fonction du type d'événement (n= 27).*



**sig. 0.001*

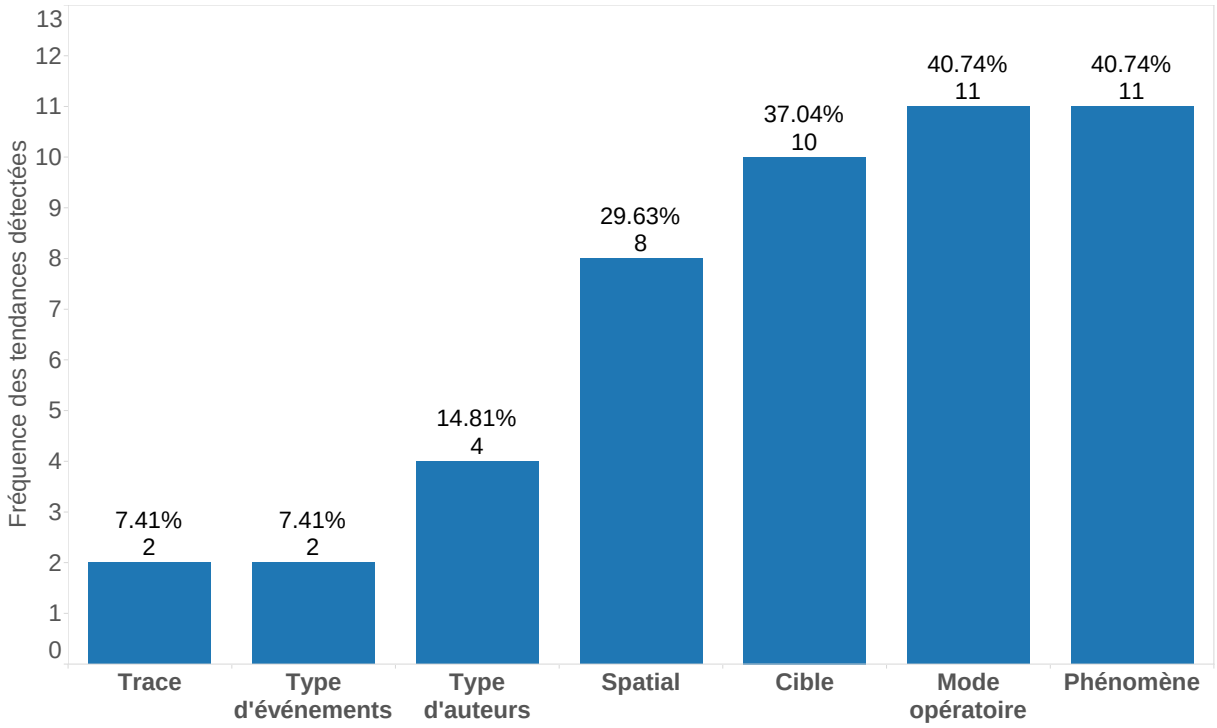
***Graphique 11** : Fréquence d'événements et de détections de tendances pour les cambriolages et les vols à l'astuce entre mai 2014 et juin 2015. Seules les deux catégories d'événements les plus représentées au Graphique 10 ont été considéré dans cette analyse.*

Concernant la connaissance des tendances détectées (Graphique 12), la grande majorité des tendances (~85%) ont déjà été au moins une fois détectée par le passé et la plupart au niveau du territoire vaudois.

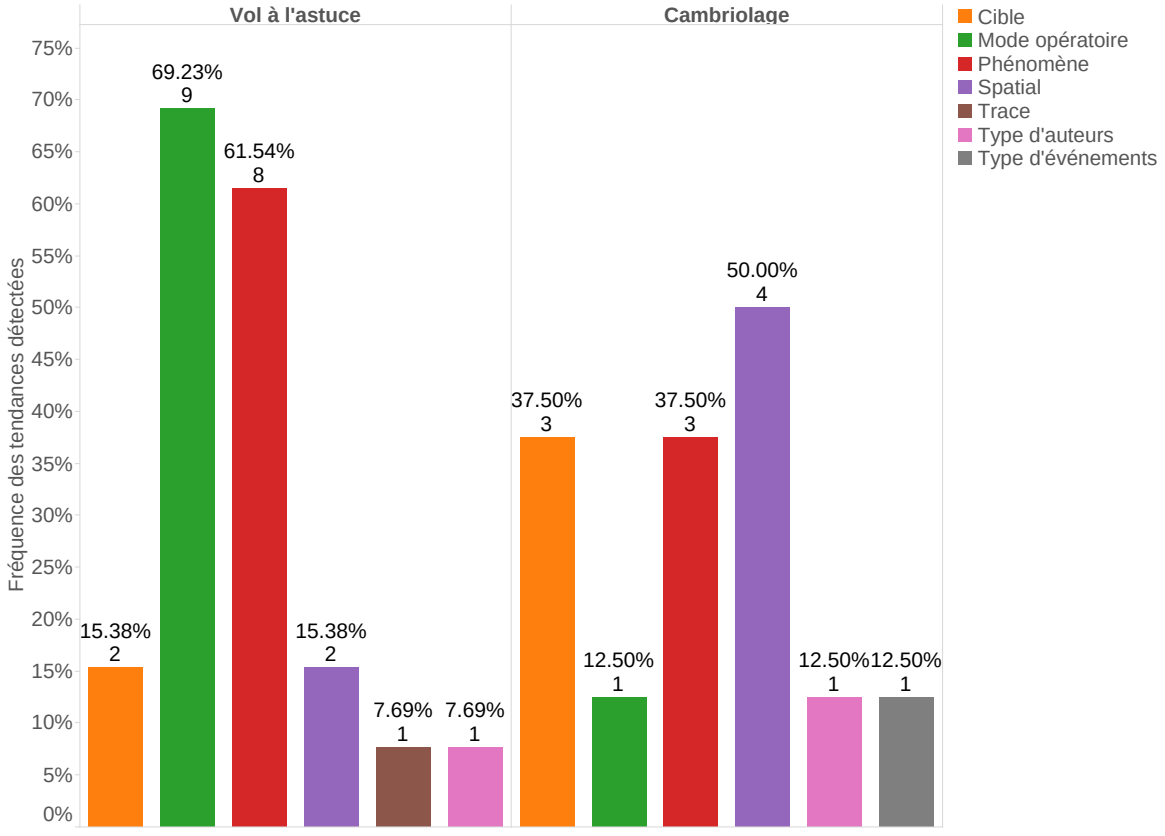


***Graphique 12** : Degré de connaissance des tendances détectées et fréquence des détections connues en fonction de l'étendue géographique (n= 27). En bleu figurent les tendances détectées qui étaient inconnues des analystes à tous les échelons géographiques considérés (national, régional et cantonal).*

Par rapport aux variables qui ont permis la détection, on constate que le code phénomène, le mode opératoire et la cible figurent parmi les plus utilisées (Graphique 13). En regardant cette même distribution, mais en fonction des deux principaux types d'événements, on note une répartition différente des variables (Graphique 14). Pour le vol à l'astuce, c'est principalement le code phénomène et le mode opératoire qui a permis les détections de tendances, alors que pour les cambriolages, c'est surtout la dimension spatiale, le code phénomène et le type de cible qui a rempli ce rôle.



Graphique 13 *: Fréquence des détections de tendances en fonction de la variable de détection (n= 27). Une détection peut résulter de plusieurs types de variables à la fois.*



Graphique 14 *: Fréquence des détections de tendances en fonction de la variable de détection pour les vols à l'astuce et les cambriolages (n= 27). Une détection peut résulter de plusieurs types de variables à la fois.*

En résumé, l'analyse descriptive des tendances détectées par l'unité d'analyse montre que les enjeux de détection se situent surtout au niveau des vols à l'astuce et des cambriolages. Cependant, le fait qu'il y ait peu de détection de tendances concernant les autres types d'événements est susceptible de soulever des questions liées à la découverte de phénomène. Il y a peu de détections de tendances qui n'ont jamais été vues auparavant et le code phénomène apparaît comme la variable principale permettant la détection. De manière plus spécifique, le mode opératoire permet surtout de détecter les tendances de vols à l'astuce, tandis que la localisation, ainsi que le type de cible permettent de détecter les tendances de cambriolages.

En se basant sur ces résultats, il a été décidé de retenir que les tendances qui présente un lien avec au moins un code phénomène ou un mode opératoire. Les tendances uniquement identifiées par un type d'auteurs sont retirées de l'analyse pour ne pas biaiser les résultats de la comparaison avec la détection automatique. C'est donc un total de 24 tendances qui seront comparées avec l'algorithme de détection automatique.

9.4.2. Capacité d'adaptation

Avant la comparaison effectuée avec la détection humaine, l'algorithme de détection a également été appliqué à d'autres variétés d'événements criminels et en utilisant plusieurs dimensions pour illustrer sa capacité potentielle d'adaptation en analyse criminelle.

La Figure 26 montre les fréquences cumulatives de différents actes illicites. Chaque trait rouge représente une détection effectuée par l'algorithme. Comme mentionné précédemment, l'algorithme est adaptatif et online. Cela signifie que chaque détection est réalisée sans prendre en considération les données subséquentes. De plus, les paramètres restent inchangés, peu importe le type de données utilisées. L'algorithme semble plus performant sur les ruptures marquées, notamment la rupture progressive permanente des vols à l'astuce au bancomat, la rupture progressive temporaire des *skimming* et la rupture périodique des cambriolages d'habitation en soirée. Les essais sur les saisies de faux documents semblent également prometteurs avec une rupture progressive permanente sur les saisies de cartes d'identité grecques contrefaites et une forme de rupture périodique sur les saisies de permis de conduire congolais contrefaits. Et, même si elle moins apparente, la rupture progressive permanente des cambriolages

d'habitation avec outils plats permet également des détections automatiques relativement précises.

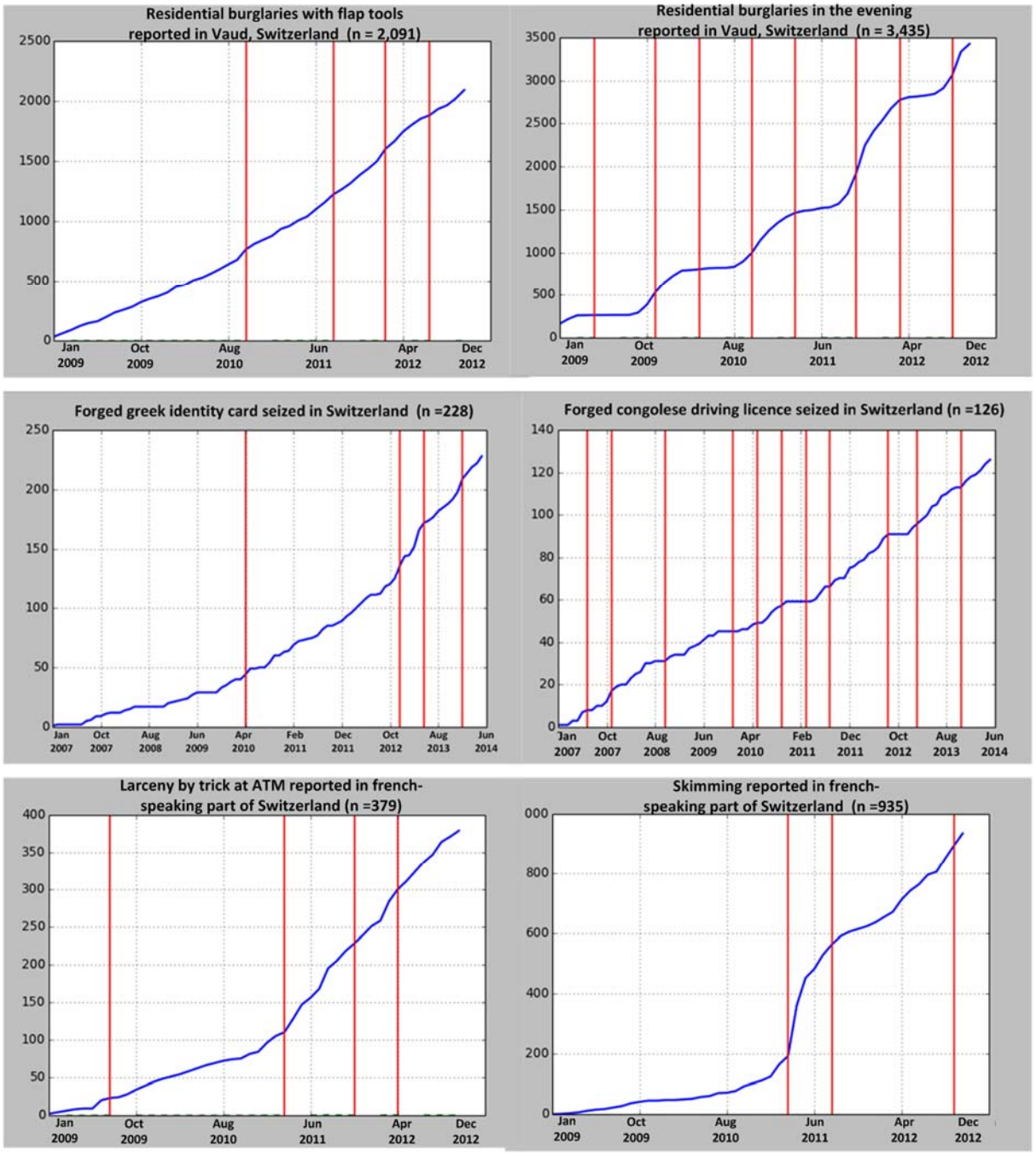


***Figure 26** : Détection automatique de pattern de rupture sur les fréquences cumulatives de 6 différents types d'actes illicites: les cambriolages d'habitation avec un outil plat (en haut à gauche), les cambriolages d'habitation en soirée (en haut à droite), les saisies de cartes d'identité grecque contrefaites (au milieu à gauche), les saisies de permis de conduire congolais contrefaits (au milieu à droite), les vols à l'astuce au bancomat (en bas à gauche) et les cas de skimming (en bas à droite).*

De plus, l'algorithme a également été appliqué sur des données de nature forensique, et plus spécifiquement sur des traces de souliers collectés sur différentes scènes de crime, à différents moments (Figure 27). En se basant sur une analyse des patterns

spatio-temporels de traces de soulier (Rodrigues, 2012), la Figure 27 montre l'évolution de motifs de traces de soulier similaires collectés sur des cambriolages d'habitation durant l'été 2010 dans le canton du Valais en Suisse. Même avec beaucoup moins de données que les exemples précédents et une fenêtre temporelle plus courte, l'algorithme de détection a identifié les ruptures, qui se sont révélés correspondre à l'activité croissante d'un seul auteur sériel.

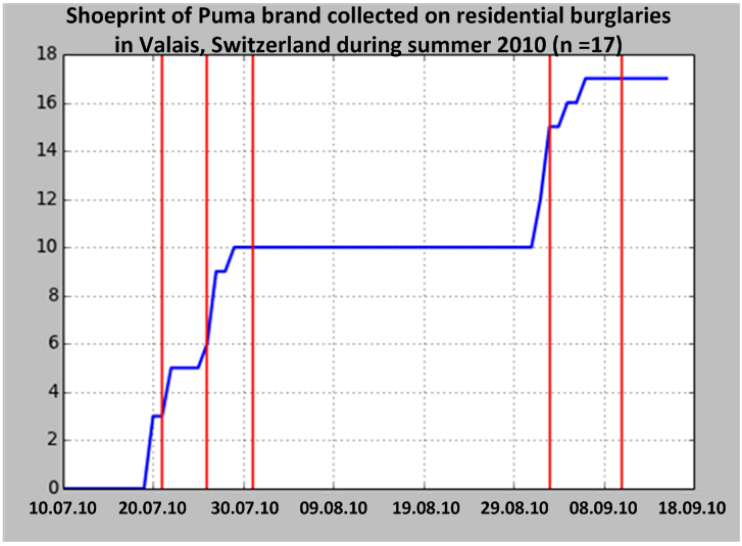


***Figure 27** : Détection automatique de pattern de rupture sur la fréquence cumulative de traces de soulier.*

Ces résultats sont prometteurs et forment des propositions concrètes en vue du développement de la méthodologie intégrative en analyse criminelle et de ses outils nécessaires. La méthode de détection montre un potentiel d'adaptation à toutes les sortes de données évoluant dans le temps et reflétant des caractéristiques sérielles. Il peut s'agir d'événements criminels, de saisies de faux documents ou de traces collectées sur les scènes de crime. Le point commun réside dans l'objet recherché : les changements dans les opportunités et l'activité d'auteurs sériels à l'aide de l'analyse des données. Ces résultats montrent également que l'enjeu de la méthode est la capacité à pouvoir subdiviser le jeu de données global en sous-groupes au sein desquels des changements de tendances sont recherchés (p. ex. codes phénomène, traces, types de cibles, etc.). Une application de l'algorithme de détection sur un jeu de données global (p. ex. tous les cambriolages confondus) ne serait pas pertinente. C'est la combinaison de cette étape de sélection des données avec l'analyse de tendances qui permet d'aboutir à la détection de problèmes spécifiques. Le type d'information exploité pour diviser le jeu de données initiales est susceptible alors de conditionner le

type d'inférence porté sur les changements détectés (p. ex. activité d'un auteur sériel, passage d'un groupe d'auteurs, vulnérabilité de cibles, etc.)

Cependant, ce premier constat initie un projet plus ambitieux consistant en l'implémentation de la méthode. Pour ce faire, il est encore nécessaire de définir les paramètres optimaux de l'algorithme afin de réduire les faux positifs (il n'y a pas de réels changements, mais l'algorithme en détecte un) et les faux négatifs (un changement réel n'a pas été détecté). L'utilité de l'approche est testée empiriquement dans le prochain chapitre. En effet, une capacité d'adaptation ne suffit pas comme seul argument pour intégrer la méthode de détection. Est-ce que les analystes équipés d'outils de détection plus simple détectent les changements après ou avant l'algorithme de détection automatique ? Est-ce que les heuristiques utilisées par les analystes peuvent être intégrées dans le modèle computationnel ?

9.4.3. Comparaison entre détection humaine et automatique

Les précédents exemples d'application de l'algorithme sur différentes sources de données illustrent bien l'importance de subdiviser le jeu de données global en sous-groupes avant de réaliser la détection de tendances. Ainsi, pour effectuer la comparaison entre la détection humaine et automatique, l'algorithme de détection a été appliqué sur différents jeux de données filtrés à partir de l'échantillon des événements PICAR. Le choix de variable filtre pour constituer les jeux de données a été réalisé de manière qualitative en fonction de caractéristiques des tendances détectées. L'objectif de cette sélection qualitative est de se limiter à un nombre minimum de filtres pour éviter de biaiser la comparaison en orientant l'algorithme. Concrètement, le choix des filtres s'est basé sur les variables qui ont permis la détection humaine en privilégiant le code phénomène. Si ce dernier n'est pas suffisant, il est complété par le type d'événement, l'étendue géographique, le mode opératoire et le type de lieux. Les choix effectués pour constituer les jeux de données correspondant à chaque tendance sont présentés à l'Annexe 23. Le nombre d'événements a été agrégé par semaines afin de se calquer sur une temporalité d'analyse exploitable en renseignement criminel. Sur chacun de ces jeux de données, l'algorithme d'analyse de changement de points a été appliqué sans en modifier les paramètres. Les résultats bruts de l'analyse sont disponibles à l'Annexe 24.

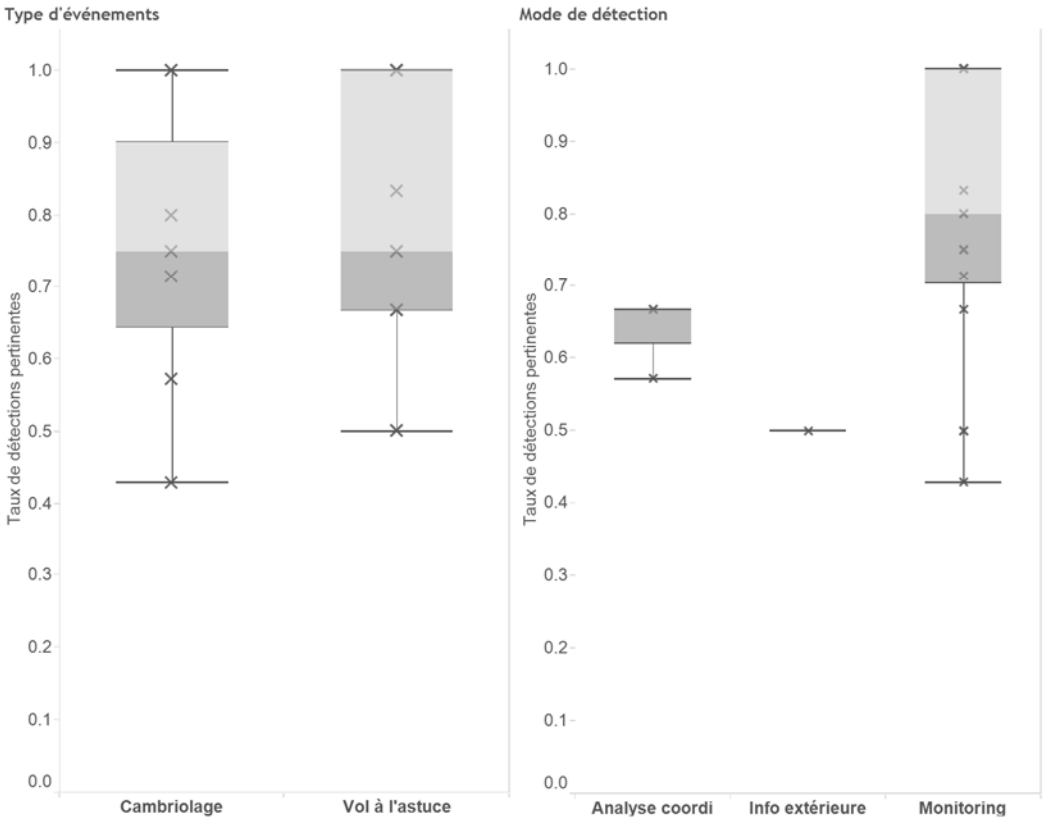
9.4.3.1. Taux de pertinence

Le taux de pertinence représente le nombre de détections pertinentes par rapport au nombre de détections totales de l'algorithme. Une détection est jugée pertinente sur la base d'une observation qualitative de la série temporelle. Si la détection correspond à un pattern de rupture tel que décrit au chapitre 9.2, elle est alors définie comme pertinente.

En moyenne, on note un taux total de pertinence de 0.71 (Tableau 7). Ce taux est légèrement plus élevé pour les vols à l'astuce (0.73) que pour les cambriolages (0.70)(Graphique 15 à gauche). On note une plus grande différence concernant le mode de détection (Graphique 15 à droite). Les tendances détectées grâce au monitoring présentent un taux de pertinence moyen de 0.74, alors que les détections réalisées grâce à une analyse proactive et une information extérieure obtiennent respectivement un taux moyen de 0.63 et 0.5. Toutefois ces résultats sont sujets à caution car il y a peu de tendances détectées grâce à une analyse proactive ou une information extérieure.

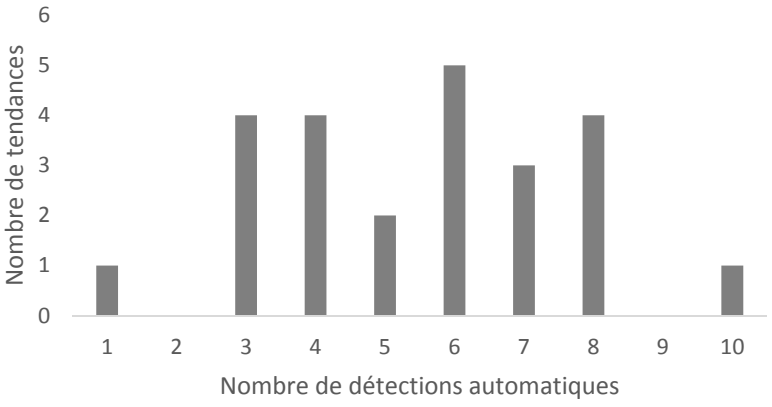
***Tableau 7** : Taux de pertinence des détections automatiques par tendance : pour chaque tendance détectée, le taux de pertinence est le rapport entre le nombre de détections jugées pertinentes sur le nombre de détection total réalisé par l'algorithme sur la tendance considérée. Un taux de 1 signifie que toutes les détections automatiques de l'algorithme sur une tendance sont pertinentes.*

ID tendance	Taux de pertinence
8	1.00
11	1.00
16	1.00
17	1.00
25	1.00
26	1.00
28	1.00
34	1.00
35	0.83
7	0.80
36	0.80
22	0.75
27	0.75
32	0.75
30	0.71
9	0.67
10	0.67
21	0.67
24	0.57
12	0.50
15	0.50
19	0.50
23	0.50
31	0.43
Total	**0.71**

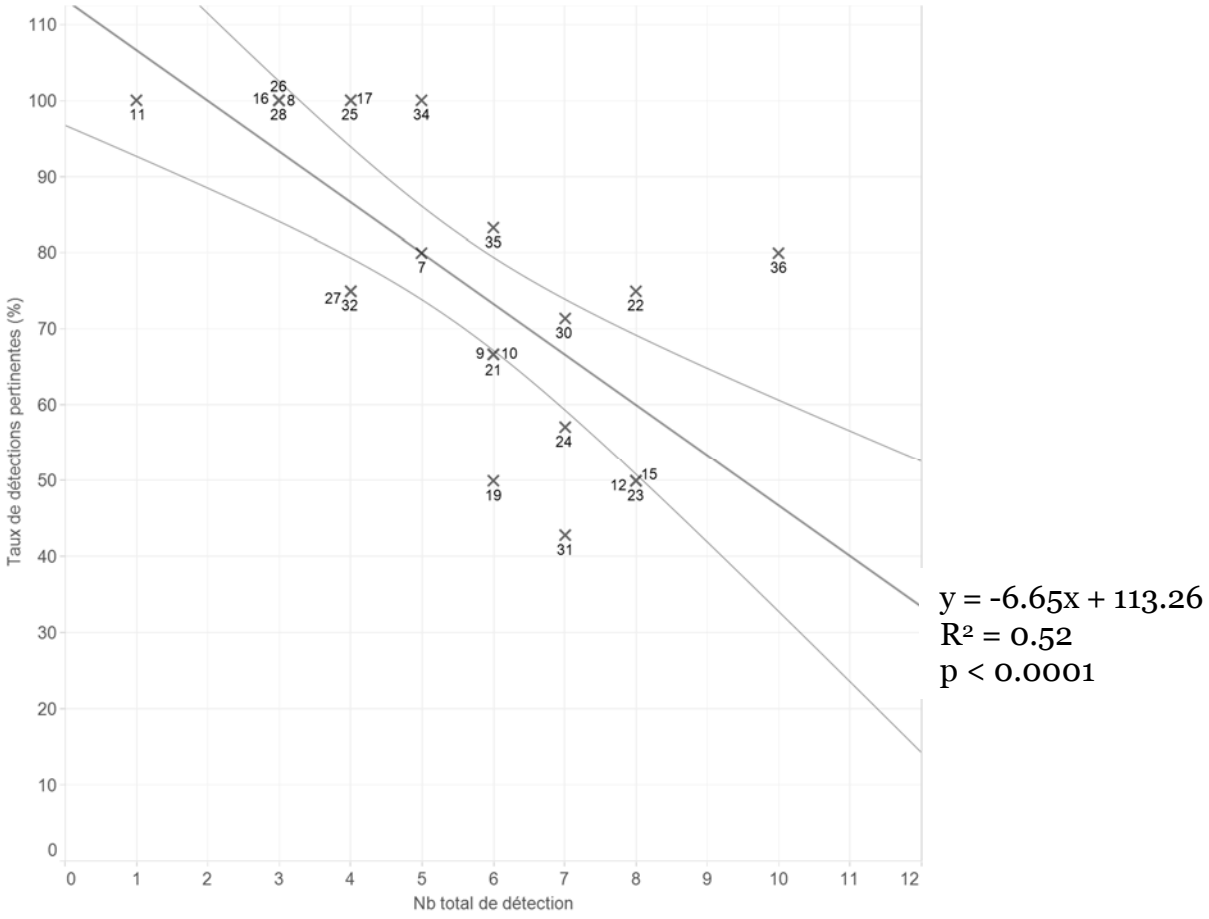


Graphique 15 *: Boxplot du taux de détections pertinentes pour les cambriolages et les vols à l'astuce et en fonction du mode de détection (n= 24).*

L'algorithme a réalisé en moyenne 5.5 détections par tendances (Graphique 16) et la majorité des tendances se situent entre 3 et 8 détections. Le rapport entre le taux de pertinence et le nombre de détections nous indique que plus le nombre de détections augmente, plus le taux de pertinence diminue (Graphique 17).



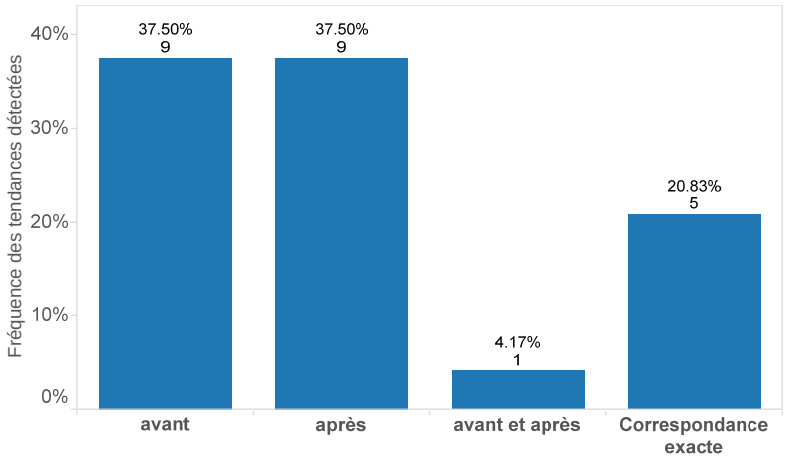
Graphique 16 *: Distribution des tendances par nombre de détection (n= 24).*



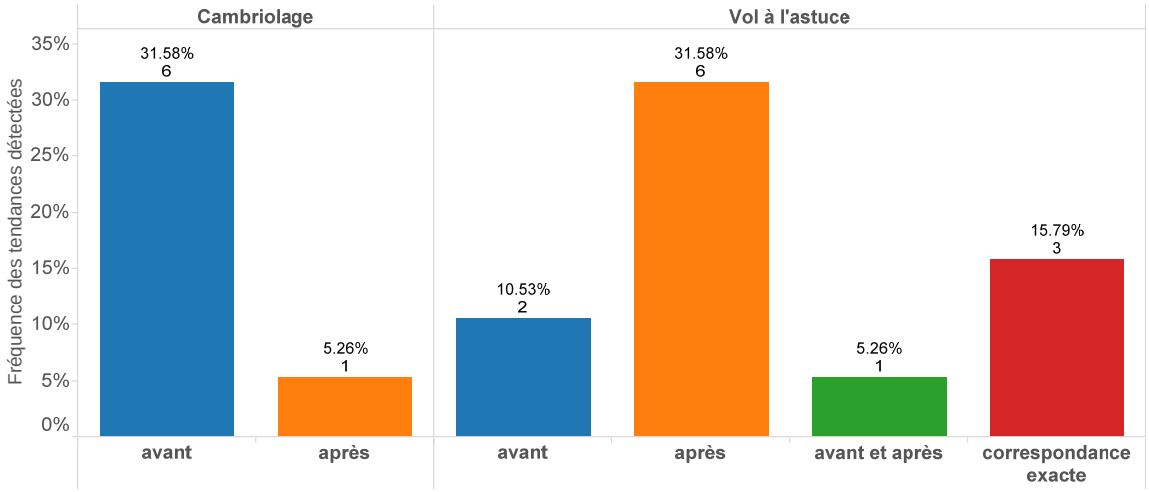
***Graphique 17** : Taux de détections pertinentes en fonction du nombre de détections (n= 24). Les étiquettes sur le graphique correspondent au numéro des tendances. L'équation de la droite de régression est indiquée, ainsi que le coefficient de détermination (R^2).*

9.4.3.2. Précision de la détection

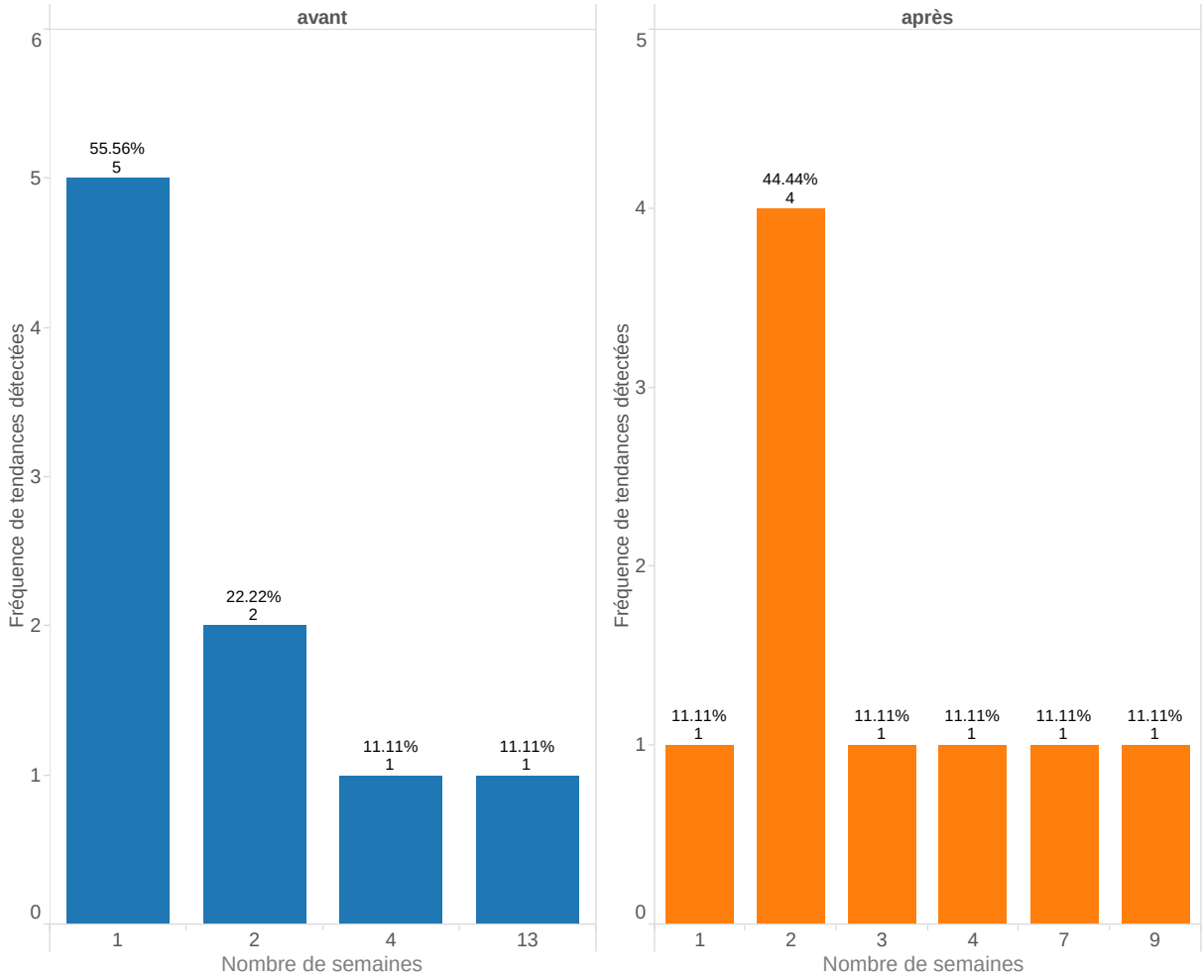
Afin d'estimer la précision de l'algorithme par rapport à la détection humaine, nous avons observé la distance temporelle entre la plus proche détection automatique et la détection humaine. Selon le Graphique 18, 20.83% des tendances présentent une correspondance exacte entre la détection humaine et algorithmique. Il n'y a pas de différence entre le nombre de tendances où l'algorithme détecte avant la détection humaine et le nombre de tendances où il détecte après. En revanche, si on l'on regarde par type d'événements (Graphique 19), on constate que pour les cambriolages la détection automatique la plus proche à le plus souvent lieu avant la détection humaine, alors que pour les vols à l'astuce, elle a lieu après. Concernant le nombre de semaines séparant la détection automatique et humaine, le Graphique 20 indique que la majorité des détections automatiques sont réalisées 1 semaine avant la détection humaine dans les cas où la détection la plus proche se trouve avant (55.56%) et 2 semaines après, lorsque la détection la plus proche se trouve après la détection humaine (44.44%).



Graphique 18 *: Fréquence des tendances détectées par rapport à la distance temporelle de la détection humaine (n= 24).*



Graphique 19 *: Fréquence des tendances détectées par rapport à la distance temporelle de la détection humaine pour les cambriolages et les vols à l'astuce (n= 24).*



Graphique 20 *: Fréquence des tendances détectées en fonction du nombre de semaines les séparant de la plus proche détection pour les détections ayant eu lieu avant la détection humaine la plus proche (à gauche) (n= 9) et les détections ayant eu lieu après (à droite) (n= 9).*

En résumé, l'algorithme de détection automatique présente de bons résultats surtout sur les tendances qui sont détectées par l'activité de monitoring de l'unité d'analyse. Ce type de tendance étant le plus souvent détecté, la valeur ajoutée d'une méthode automatique prend ici tout son sens. L'enjeu des faux positifs est cependant soulevé, car plus le nombre de détections augmente, plus le taux de pertinence des détections diminue. Concernant sa précision, la détection automatique se montre plus rapide que la détection humaine sur les tendances de cambriolages, tandis que pour les tendances de vols à l'astuce, c'est la détection humaine qui se montre en général plus rapide. Une hypothèse sur ces résultats serait qu'il y a plus d'événements de cambriolages qui sont enregistrés dans PICAR. Les analystes feraient alors face à beaucoup plus de données sur ce type d'événements par rapport aux vols à l'astuce. La détection d'un changement dans les tendances pour des cambriolages prendrait alors plus temps. Néanmoins, la plupart des détections automatiques ont lieu entre 1 ou 2 semaines avant ou après la

détection humaine. Nous rappelons à nouveau que ces résultats sont à titre indicatif et ne représentent pas une évaluation formelle de l'algorithme considéré.

9.4.4. Paramétrisation

Afin de déterminer si la paramétrisation de la méthode de détection automatique a une influence sur la détection de rupture de tendances, nous avons procédé à une estimation des paramètres optimaux selon le type de tendances. Cet aspect a principalement été traité dans le cadre d'un travail de maîtrise à l'École des sciences criminelles (Vivas Ramos, 2016). La première étape consiste à classifier les tendances pour tous les codes phénomènes enregistrés dans PICAR entre 2009 et 2013 en observant les séries temporelles pour les comparer à la catégorisation présentée au chapitre 9.2. Cette classification qualitative a permis de confirmer la classification théorique en subdivisant les ruptures dans les tendances en 5 catégories principales[38] :

- Pattern de rupture périodique
- Pattern de rupture progressive temporaire
- Pattern de rupture dégressive temporaire
- Pattern de rupture progressive périodique
- Pattern de rupture dégressive périodique

À cela s'ajoute également une catégorie regroupant les phénomènes constitués de peu de cas observés. Ces phénomènes n'affichent pas des tendances similaires aux phénomènes plus fréquents et lorsqu'un cas survient, il est rapidement détecté de par sa rareté. La comparaison avec des phénomènes plus fréquents s'avère donc difficile.

La seconde étape consiste à faire varier les trois paramètres, DPU, SSS et SEG, de l'algorithme afin d'identifier les paramètres optimaux pour chaque catégorie de rupture de tendances. Ce travail fut réalisé de manière qualitative en effectuant une première segmentation manuelle des tendances, puis en appliquant l'algorithme automatique avec différentes variations de paramètres. L'analyse a été réalisée en agrégeant les données par mois et par semaine. Les paramètres optimaux sont résumés au Tableau 8.

[38] Voir Annexes G

***Tableau 8** : Paramètres optimaux pour le programme FCPD en fonction du type de rupture à détecter dans les tendances et du niveau d'agrégation temporelle.*

Type de rupture	Paramètres					
	Agrégation par mois (2009 à 2013)			Agrégation par semaine (mai 2012 à juin 2013)		
	DPU	*SSS*	*SEG*	*DPU*	*SSS*	*SEG*
Périodique	0.007	2	1	0.04	3	1
Progressive temporaire	0.03	3	1	0.06	3	1
Régressive temporaire	0.04	3	1	0.065	3	1
Progressive permanente	0.02	3	1	-	-	-
Dégressive permanente	0.035	3	1	0.03	3	1
Peu de cas	0.006	1	1	0.006	1	1

Si les résultats montrent une certaine différence entre les tendances présentant des ruptures périodiques et les autres lorsque l'on considère une agrégation par mois, en agrégeant par semaine, cette distinction semble s'estomper. En revanche, les tendances constituées de peu de cas semblent se distinguer dans les deux cas. Le paramètre variant le plus est le DPU (la déviation par rapport à la valeur attendue). Le SSS (le nombre de changements de signe de la pente) reste majoritairement à 3, mais tombe à 1 pour les tendances présentant peu de cas, ce qui semble logique puisque ces phénomènes étant rares, une seule augmentation ou diminution est susceptible de déclencher une rupture à détecter. Finalement le SEG (nombre de seuils à atteindre) est à 1 dans tous les cas.

9.4.5. Évaluation de l'algorithme

La validité de la méthode au niveau technique n'est pas présentée dans le détail ici dès lors que cette thèse se consacre à la dimension intégrative de la problématique et moins à la dimension technique. Néanmoins, cette question est traitée en détail dans deux publications (Albertetti, 2016; Albertetti, Grossrieder, Ribaux, & Stoffel, 2016) dont les principaux résultats sont brièvement résumés ici.

Afin de démontrer les avantages de l'approche Fuzzy Change Points Detection (FCPD), l'algorithme a été comparé à un algorithme de détection de changement de points similaires, BFAST, sur des données policières (PICAR) et des données financières (Swiss Exchange Market). Les résultats en termes de précision sont similaires avec l'algorithme de détection BFAST, mais en termes de complexité, la méthode FCPD se

révèle plus efficiente que ce dernier, notamment car la méthode BFAST est *offline* et de ce fait n'est pas adaptée pour l'analyse des tendances dans les activités criminelles contrairement à l'approche FCPD. La combinaison d'une représentation significative, d'une segmentation dynamique et d'un système d'inférence floue permet aux analystes non familiers[39] aux questions computationnelles de trouver intuitivement des ruptures dans les tendances en décrivant les propriétés géométriques en termes linguistiques, qui sont notre façon naturelle de décrire une courbe.

9.5. Discussion

Peu importe le modèle computationnel, le processus de détection reste une opération semi-automatique. Cela renforce la considération de l'humain comme un composant essentiel dans le processus complet. Afin de pouvoir appliquer l'algorithme de détection, le problème considéré doit avoir déjà été structuré au préalable à l'aide de systèmes de codification adéquats, d'où l'importance de la classification automatique présentée au chapitre 8. La catégorisation situationnelle utilisée par le CICOP est un excellent exemple, tout comme l'utilisation de certaines caractéristiques issues des traces. Ces formes de catégorisations peuvent être considérées dans l'idée d'adapter certains paramètres (p. ex. le niveau d'agrégation de la fenêtre temporelle ou de l'aire géographique) en fonction de la situation criminelle. De la même manière, une fois que le pattern de rupture est détecté, il doit encore être analysé. La relation subtile entre le grand potentiel des méthodes computationnelles et le pilotage par des connaissances humaines est ainsi au cœur de l'analyse computationnelle de tendances des activités criminelles.

Les résultats montrent que la détection automatique de tendances est susceptible de soutenir le travail des analystes. Mais une application aveugle de l'algorithme sur les données risque de péjorer ce travail. Il semble qu'il existe une distinction entre les tendances relevant des cambriolages et des vols à l'astuce. Les tendances de vols à l'astuce sont détectées lorsque des modes opératoires particuliers sont repérés, tandis que la dimension spatiale et le type de cible jouent un rôle plus important dans le cas des cambriolages. Dans tous les cas, le type de phénomènes représenté par les codes CICOP permet un grand nombre de détections pertinentes.

[39] Mais également aux analystes familiers des questions computationnelles, car la détermination des seuils dépend des données.

La tentative de paramétrisation ne permet pas une application opérationnelle en l'état. Cependant, cela a permis de montrer un potentiel d'optimisation en tenant compte du type de rupture à détecter. De même le type de phénomène est susceptible également d'influencer cette optimisation. Les seuils à définir étant différents entre les phénomènes présentant peu de cas et ceux plus fréquents. Des recherches ultérieures sont nécessaires en vue d'approfondir cet aspect, notamment pour déterminer sur quel type de phénomènes une détection automatique peut être utile.

Ces résultats doivent cependant être pris avec précaution. Il est parfois difficile de déterminer à quelle tendance dans les données fait référence une détection humaine. La même observation peut être tirée des détections automatiques. Cette difficulté est inhérente à l'opération de détection de tendance, dès lors que l'étape de détection est distincte de l'étape d'analyse. Au niveau de la détection, la question de la nature de la tendance n'entre pas encore en considération. Il s'agit dans un premier temps d'observer une anomalie dans les données. Ce n'est que dans un deuxième temps, l'analyse, qu'une signification sera attribuée à cette anomalie. La difficulté de faire correspondre les détections humaines et automatiques aux données découle directement de cette observation. Une détection peut sembler incohérente a posteriori lorsque l'on dispose d'une série temporelle complète, mais au moment T de la détection, il est presque impossible de le dire. Néanmoins, ces résultats ouvrent la voie à de futures expérimentations destinées à approfondir ces questions

10. Synthèse : vers un processus semi-automatique de détection de tendances

Afin d'illustrer l'approche méthodologique développée, une démarche d'application des méthodes computationnelles en trois temps est appliquée à deux étapes-clés de la veille opérationnelle, l'intégration et la détection. Les exemples de la classification de cambriolages d'habitation et de la détection de tendances des activités criminelles ont été choisis pour expérimenter la démarche. Dans un premier temps, l'expression et la modélisation des processus ont permis de comprendre comment les analystes opèrent pour classifier les événements dans PICAR, puis comment les tendances sont détectées. Fort de ces acquis, il a ensuite été possible de tester une méthode de réseau neuronal sur les cambriolages d'habitations pour codifier automatiquement les événements en code phénomène, puis d'appliquer une analyse de changement de points sur des séries temporelles afin de détecter des patterns de ruptures dans les tendances des activités criminelles. L'application des algorithmes a été pilotée dans le but de répondre aux besoins et contraintes de la tâche à effectuer. Finalement, les résultats sont discutés au regard des savoirs forensiques et criminologiques exposés à la partie II. Le potentiel avéré de ces méthodes permet d'envisager leur implémentation dans le processus global (Figure 28). La couche computationnelle intervient alors en tant que soutien des analystes et s'intègre dans les processus déjà existants sous la forme d'une brique complémentaire.

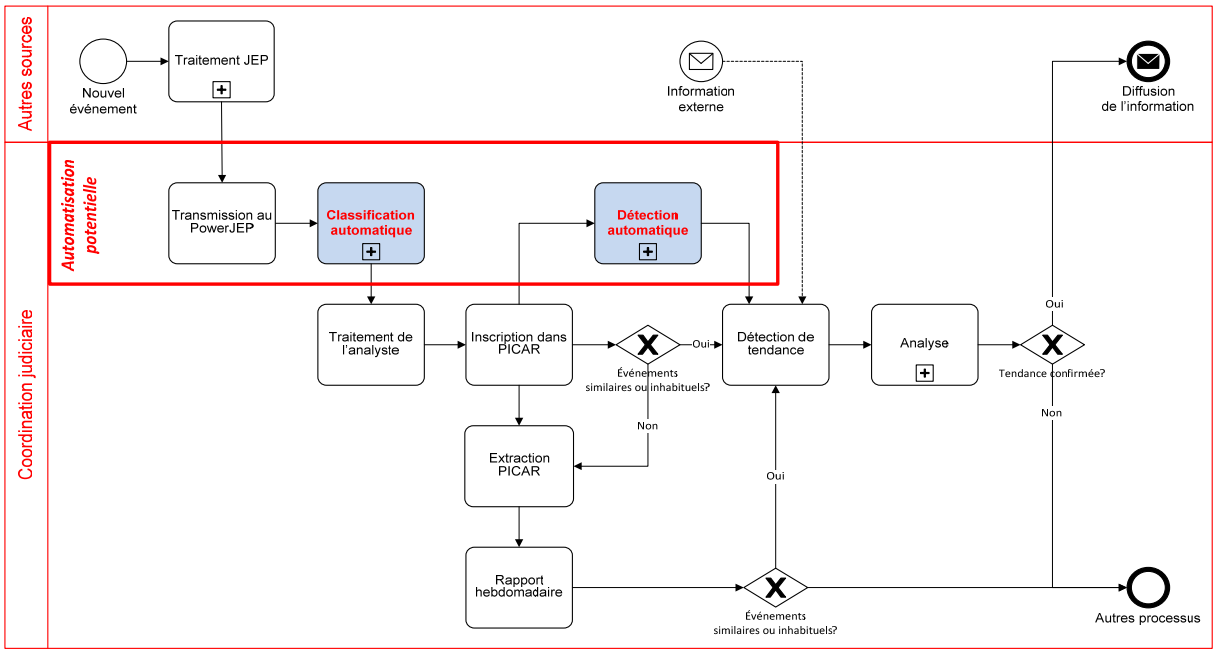


***Figure 28** : Processus BPMN de l'intégration d'événements dans PICAR et de la détection de tendances. Les tâches en rouge indiquent les étapes susceptibles d'être automatisées. La classification automatique permet un premier filtre avant l'intégration par les analystes, puis la détection automatique intervient en parallèle des autres formes de détection humaine en tant que support.*

Les résultats obtenus corroborent l'hypothèse principale de ce travail qui soutient la possibilité de détecter des problèmes dans la distribution spatio-temporelle des données de la criminalité. En effet, l'analyse de changement de point appliquée sur les différentes tendances permet de détecter les patterns de ruptures imprimés dans les données. Cependant, l'hypothèse d'une détection automatique plus complète, rapide et précise que la détection humaine n'est que partiellement corroborée. En effet, il apparaît que les capacités de l'algorithme varient selon le type de criminalité étudié. Dans le cas des cambriolages, il semble que l'algorithme automatique soit plus rapide que les analystes, mais c'est l'inverse qui se produit lorsque l'on considère les vols à l'astuce. L'hypothèse spécifique postulant que ces patterns de données reflètent les patterns d'activités criminelles se trouve corroborée par les différentes illustrations empiriques proposées. L'arrivée de nouveaux groupes d'auteurs, les changements d'environnement avec le passage à l'heure d'hiver ou encore l'activité d'un auteur sériel sont autant d'exemples imprimant des patterns particuliers dans les données collectées. Concernant la dernière hypothèse spécifique, une classification situationnelle des événements semble appropriée pour effectuer le processus de détection et l'hypothèse est corroborée. Le code CICOP est la variable cible la plus utilisée pour la détection, à la fois par l'algorithme et les analystes.

Les illustrations proposées montrent le grand potentiel des méthodes computationnelles en analyse et renseignement criminel lorsque leur intégration est réfléchie et guidée par les processus. La démarche méthodologique pour parvenir à ces résultats nécessite néanmoins de gagner en niveau analytique en formalisant une approche générale interdisciplinaire fondée sur la résolution de problème. La prochaine partie tente de faire un pas vers cet objectif en proposant un cadre de travail issu des travaux de cette thèse.

PARTIE IV : VERS UN CADRE DE TRAVAIL INTERDISCIPLINAIRE CENTRÉ SUR LES PROBLÈMES

L'expression de la démarche méthodologique exposée à la partie II et l'intégration de méthodes computationnelles dans les processus du CICOP à la partie III nous montrent que la complexité gérée par l'unité de renseignement criminel tend à devenir de plus en plus difficile à traiter. Il apparaît alors un besoin d'intégrer au sein de l'unité un solide composant dédié à la maîtrise de l'évolution de l'unité, de sa méthodologie et de ses outils. Les tâches d'un tel composant sont de :

- maintenir une vision globale sur la méthodologie et les outils disponibles ;
- éviter de passer abruptement d'une technologie à l'autre ;
- faire évoluer le dispositif de veille étape par étape en le confrontant à la réalité quotidienne le plus tôt possible (c.-à-d. généralement en intégrant un petit élément au sein de ce qui existe déjà) ;
- favoriser une démarche de recherche interdisciplinaire : l'analyse criminelle cherche à combiner les méthodes et techniques les plus adéquates issues des recherches de disciplines variées. De ce point de vue, l'analyse criminelle ne doit pas uniquement servir de simple fournisseur de données ou de champ d'expérimentation intéressant pour ces disciplines, qui souvent essaient de développer des modèles généraux pour leurs propres besoins.

La démarche exposée dans cette thèse et les résultats présentés permettent de franchir une étape vers ces objectifs et ces besoins. Cette dernière partie discute alors les implications de ces résultats de manière plus générale en proposant une synthèse sous la forme d'un cadre de travail interdisciplinaire (chapitre 11). Les limites et enjeux inhérents à cette recherche (chapitre 12) ainsi que les perspectives (chapitre 13) que celle-ci soulève sont finalement discutées.

11. Criminologie forensique computationnelle

La réflexion menée à la partie II permet d'identifier les différentes strates d'analyses et leur articulation au sein d'un processus général qui repose sur l'objet d'étude de la science forensique, la trace. En remplissant le rôle de pont entre la réalité observée et les raisonnements cognitifs qui en découlent, la trace diligente la manière dont sont construites *a posteriori* les différentes hypothèses qui visent à expliquer cette même réalité. La compréhension des problèmes, la représentation des activités sérielles et la reconstruction du crime reposent alors fondamentalement sur cette interface qu'est la trace. Être conscient de cette dépendance permet ensuite de nuancer et de mettre en perspective les différents raisonnements explicites ou implicites qui jalonnent l'analyse et le renseignement criminel. Il devient alors possible de mieux appréhender les trois niveaux d'analyses qui vont du niveau tactique au stratégique en passant par l'opérationnel. Chaque strate possédant ses propres objectifs, ses propres contraintes, sa propre temporalité et ses propres ressources, nous sommes d'avis que l'intégration de méthodes computationnelles doit être pilotée par la connaissance du domaine. Si la trace fait office de pont entre le réel et le raisonné, le pattern apparaît alors comme l'objet pivot en analyse criminelle. La compréhension des hypothèses et des mécanismes qui découlent de la notion de pattern joue un rôle crucial dans la sélection et l'implémentation de techniques computationnelles.

L'expérience réalisée sur la détection de pattern dans les tendances des activités criminelles a permis de proposer des liens simples entre les théories situationnelles en criminologie et certains aspects de la science forensique (renseignement forensique). Dans le même esprit, les méthodes computationnelles ont servi tant à traiter des données structurées au sein des situations criminelles, qu'à détecter des ruptures dans des données forensiques. Les résultats obtenus pour la détection de patterns dans les tendances des activités criminelles sont uniquement rendus possibles si ces trois disciplines, c.-à-d. la criminologie, la science forensique et les sciences de l'information, contribuent, en même temps et de manière équilibrée, au processus d'analyse criminelle. Dans la plupart des cas, sans l'information résultant des traces, une activité criminelle ne peut être précisément reconstruite ; sans les approches situationnelles en criminologie, les situations criminelles spécifiques ne peuvent pas être identifiées ; et sans les méthodes computationnelles, les données ne sont traitées que passivement.

Une approche méthodologique holistique émerge de ces combinaisons (Figure 29). Celle-ci nécessite néanmoins d'être formalisée de manière plus approfondie afin de consolider son utilisation harmonisée au sein de différentes juridictions, de la mettre à l'épreuve, et de faire évoluer le système sans démanteler les précédentes réalisations pour implémenter une nouvelle technologie.

Cela nous mène à l'importance de construire un cadre de travail interdisciplinaire plus ambitieux en sciences criminelles, qui sera susceptible de structurer l'approche de manière plus complète. Un tel cadre de travail, que l'on qualifiera de *Criminologie Forensique Computationnelle* (CFC) (Figure 30), vise à délivrer l'analyse et le renseignement criminel, en se basant sur les données de la criminalité générées par les traces, analysées avec les méthodes computationnelles, et expliquées/supportées par les théories criminologiques. En résumé, le composant forensique exploite l'information transmise par les traces résultant des activités criminelles à l'aide du renseignement forensique. Le composant criminologique est constitué des connaissances théoriques qui tentent d'expliquer la criminalité (ou la déviance) en étudiant l'acte criminel, l'auteur, la victime et/ou la réaction sociale à son égard. À ce titre, nous avons montré que le développement fulgurant de la criminologie environnementale et des approches situationnelles offre de nouveaux outils conceptuels pragmatiques pour atteindre un bon équilibre et lier les différents éléments dans la démarche d'ensemble.

Enfin, le composant computationnel permet de gérer la complexité et la quantité des données disponibles, tout en étant intégré dans les processus d'analyse et renseignement criminel.

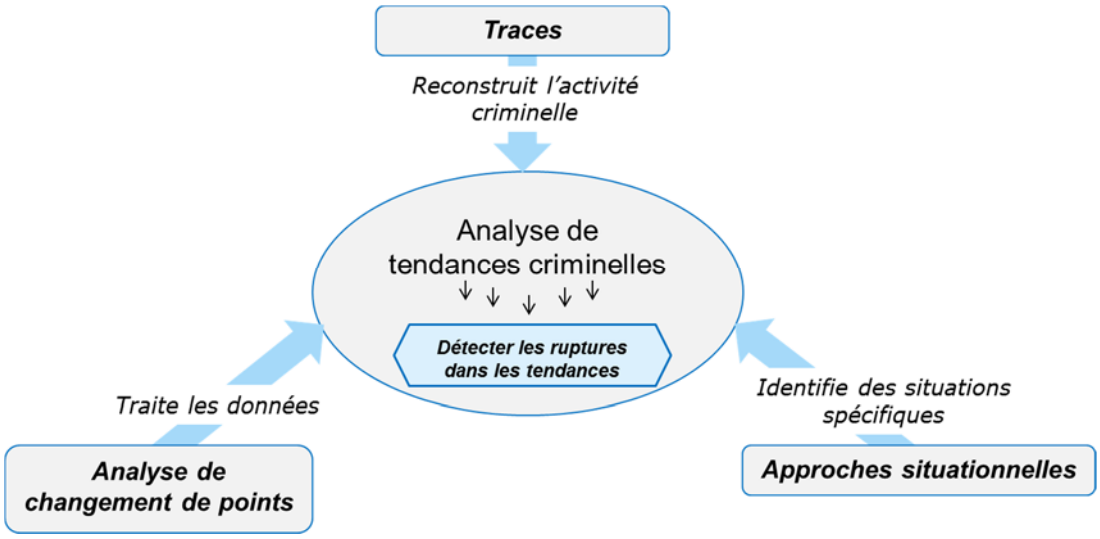


Figure 29 *: Cadre de travail interdisciplinaire dans l'analyse de tendances : Les traces laissées par les malfaiteurs permettent de reconstruire le crime, l'analyse de changement de points traite les données en vue d'identifier les ruptures dans les tendances, et les approches situationnelles en criminologie guident l'algorithme et interprètent les résultats en identifiant des situations spécifiques dans les activités criminelles.*

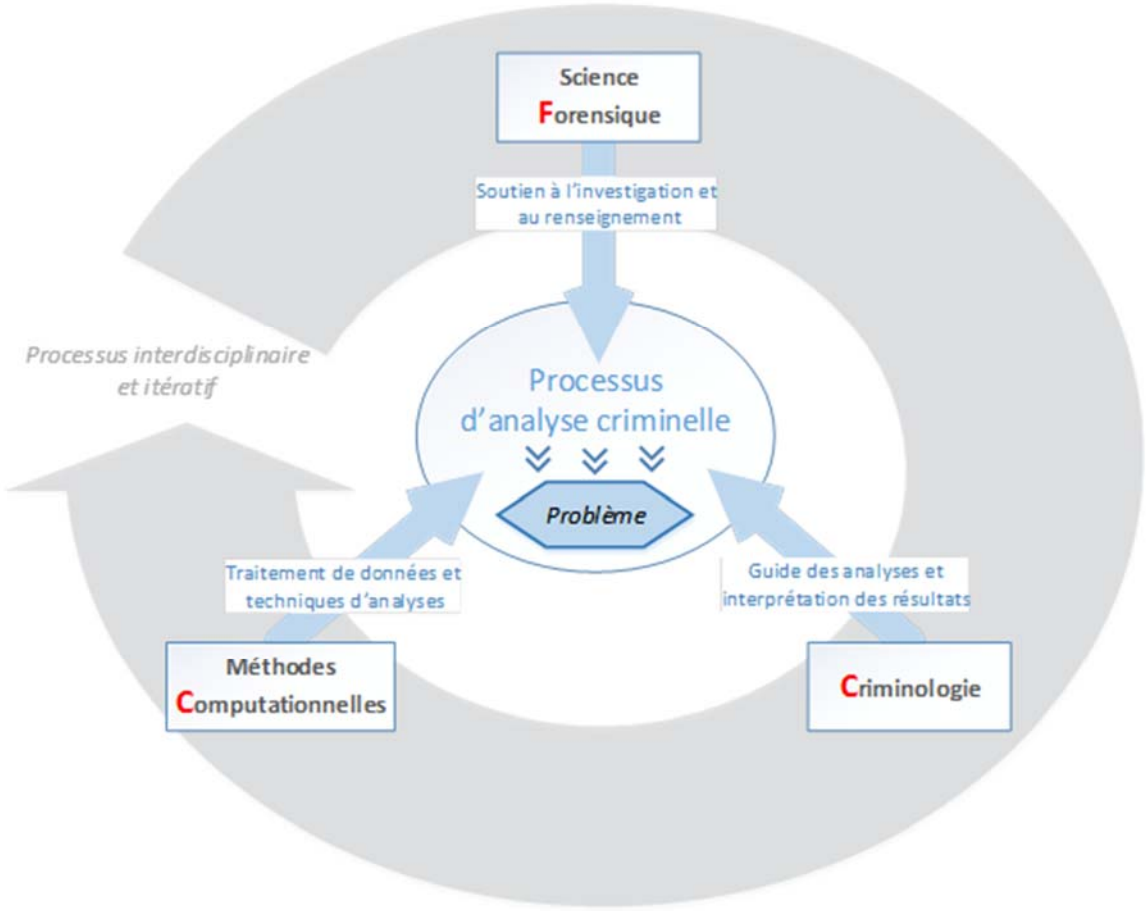


Figure 30 *: Criminologie Forensique Computationnelle (CFC) : Approche interdisciplinaire orientée vers la résolution de problèmes dans l'intégration des méthodes computationnelles en analyse criminelle. Cette intégration au sein des processus est alimentée par la science forensique qui offre un soutien à l'investigation et au renseignement. La criminologie permet de guider les analyses et d'interpréter les résultats obtenus par les algorithmes implémentés. La nature itérative du processus permet de conserver une souplesse d'adaptation nécessaire pour répondre aux évolutions de la criminalité.*

Les intégrations de méthodes computationnelles de classification de phénomènes et de détection de tendances à la partie III mettent également l'emphase sur les avantages de l'approche CFC en termes d'opérationnalisation des processus. Une telle approche favorise ainsi ces intégrations, notamment à travers :

- **La simplicité des méthodes.** Ces dernières font appel à des algorithmes intuitifs et nécessitent peu de paramètres dans leurs fonctionnements.

- **Le gain de temps.** Même si elles se révèlent semi-automatiques, les méthodes testées sont susceptibles d'opérer un premier filtre pour les analystes, permettant ainsi de transférer les ressources sur des tâches d'analyse.

- **La gestion multitâche.** La semi-automatisation des processus permet d'opérer sur différents niveaux d'agrégation spatiale en même temps et sur différents types de phénomènes criminels (p. ex. les codes CICOP).

- **L'adaptabilité.** Les méthodes proposées sont susceptibles de s'appliquer sur différents types d'activités (p. ex. cambriolages d'habitation, vols à l'astuce, *skimming*, etc.), différents types de traces (p. ex. traces de semelle, faux documents, etc.), différents types de tendances (p. ex. périodiques, temporaires, permanentes, etc.) et différents niveaux d'analyse (p. ex. stratégique, opérationnel, tactique, etc.).

Ainsi, des techniques pouvant être considérées comme basiques en sciences de l'information peuvent se révéler très efficaces. En revanche, la méthodologie entourant l'intégration de ces techniques n'apparaît pas aussi triviale et se pose comme un enjeu majeur. Comme l'ont montré les différentes expérimentations réalisées, la clé semble résider dans la capacité à identifier des sous-groupes de données avant la mise en œuvre de la méthode computationnelle. De la même manière qu'une base de données telle que PICAR en Suisse est accompagnée d'une approche méthodologique spécifique, le pilotage de techniques computationnelles demande à être encadré en amont et en aval par les connaissances théoriques (forensique et criminologique) et l'expérience accumulée des praticiens.

12. Enjeux et limites

Malgré ses contributions indéniables, l'approche CFC se heurte également à certaines limites et certains biais. Il est important de les considérer afin de mettre en perspective les résultats obtenus et de nuancer leur interprétation. Ces dernières sont relativement similaires aux limites généralement formulées dans la littérature sur la question (Maurushat, Bennett-Moses, & Vaile, 2015; Perry et al., 2013).

L'idéal souvent véhiculé par les techniques computationnelles nous guide vers une automatisation complète des différents processus. Bien que les avantages d'une méthode automatisée aient pu être démontrés lors de cette recherche, les processus de classification et de détection présentés restent néanmoins des approches semi-automatisées. Les données transmises à l'algorithme nécessitent un traitement humain au préalable et une fois les résultats obtenus, c'est à nouveau les analystes qui opèrent une analyse qualitative des données. Les méthodes computationnelles font plutôt office de couches-filtres qui permet de dégrossir le terrain pour les analystes. Elles préparent les données en vue d'accélérer la saisie des informations et d'améliorer sa fiabilité, de mettre en évidence ou suggérer des analogies pour déclencher des associations d'idées, ou d'attirer l'attention vers des jeux de données intéressants (détection). Les tâches d'analyse et de décision restent ensuite essentiellement l'apanage de l'expertise humaine.

La force de cette approche peut se révéler être une limite également. L'une des conditions pour le bon fonctionnement du système est l'injection de connaissances *a priori* dans le processus. Les deux illustrations empiriques proposées à la partie III (la classification automatique de phénomènes et la détection automatique de tendances dans les activités criminelles) sont basées sur l'existence de codes phénomènes. Il est donc important de bien comprendre et appréhender les différentes situations criminelles afin de guider l'intégration des données collectées suite aux interventions policières. Il peut se révéler rapidement illusoire de vouloir appliquer cette approche à tous les types de criminalité confondus. Même si l'approche CFC est susceptible d'être adaptée à différentes problématiques (voir chapitre 9.4.2), le travail préalable de compréhension du phénomène et l'adoption d'une méthodologie solide demeurent les pièces centrales de la réflexion.

Cet aspect nous ramène toutefois au risque de tunnel mental ou « l'effet tunnel » (Kahneman, 2013). Cet effet, qui dans le lexique médical, fait référence à une perte de la vision périphérique, peut se traduire dans notre contexte par une vision circonscrite et limitée de l'objet d'étude. Une partie des variables utilisées dans les analyses sont obtenues en décomposant les codes phénomènes du CICOP (voir chapitre 8.3). On pourrait alors craindre, notamment lors des essais de classification des événements, d'avoir une vision réduite et imposée des phénomènes. De même, pour la détection de tendances dans les activités criminelles, il existe un risque de n'identifier que les tendances d'un genre déjà connu et ayant déjà été détectées auparavant. La découverte de nouveaux phénomènes ou de nouvelles tendances représente ainsi un autre enjeu de l'approche CFC. Afin de nuancer cet effet tunnel, les techniques relevant de l'apprentissage non supervisé en data mining sont susceptibles de montrer un certain potentiel. Les techniques de *clustering* ou des règles d'association, capable d'effectuer des regroupements de patterns, sont à même de mettre l'accent sur des phénomènes éventuellement inconnus des analystes du CICOP.

En termes plus opérationnels, les enjeux liés aux faux positifs lors des détections ne sont pas négligeables. Le genre de système de détection tel que présenté au chapitre 9.3.3 est supposé soutenir le travail des analystes. Cependant, un trop fort taux de faux positifs dans les alertes générées par le programme est susceptible de produire un effet contraire à celui qui est recherché. Il existe alors un risque pour les analystes de passer plus de temps à traiter des alertes qui se révèlent non pertinentes au détriment de leurs autres tâches. Tout l'enjeu se situe alors à jauger le juste équilibre entre le taux de faux positifs et le taux de faux négatifs, sachant que le système ne devrait idéalement pas rater une tendance pertinente. Un travail d'optimisation de la méthode est encore nécessaire en vue d'atteindre cet objectif.

Finalement, il convient de rappeler que la partie III sur l'application empirique de méthodes computationnelles en analyse criminelle est une confrontation du modèle à une problématique opérationnelle. Il ne s'agit pas d'une évaluation stricto sensu de différentes techniques en vue de mesurer leurs performances dans un cadre précis (ou sur des jeux de données tests préparées). Nous avons vu par exemple que la technique de détection (*change point analysis*) utilisée ne répond pas de manière universelle à nos attentes en fonction de la variété des situations rencontrées : pour certains phénomènes déjà bien maîtrisés, la détection de tendances serait probablement plus

efficace en construisant, adaptant et exploitant des modèles qui permettent de reconnaître leur apparition plus rapidement. De tels approfondissements, ou hybridation de l'architecture, sont nécessaires à terme, par exemple au moment d'implémenter opérationnellement ces méthodes au sein d'une unité d'analyse de la police.

13. Perspectives

En établissant la construction d'un cadre interdisciplinaire tel que le CFC comme un objectif général et en considérant l'unité de renseignement criminel comme un objet de recherche susceptible d'être étudié sous l'angle de ce cadre, plusieurs axes de recherche s'ouvrent alors à nous.

Parmi ces directions possibles se pose la question de l'intégration des théories des opportunités à l'investigation de la scène de crime. La décomposition des phénomènes en situations criminelles, ainsi que la détection et l'analyse de patterns dans les données sont susceptibles d'alimenter en connaissance la recherche et la collecte de traces sur le champ d'investigation. Approfondir et formaliser cette relation permettrait d'aider l'investigation de scène de crime à l'aide du renseignement criminel.

L'implémentation d'un système de classification automatique de phénomènes et d'un système de détection automatique de tendances au sein de l'unité d'analyse du CICOP serait un ajout pertinent dans la construction itérative de leur méthodologie. Ces systèmes s'imbriqueraient dans le processus sans révolutionner l'ensemble des méthodes utilisées, mais en ajoutant un nouveau composant là où une lacune peut être comblée ou complétée. Dans cette optique, et malgré un affinage des paramètres de la méthode automatique réalisée sous la forme d'un travail de maîtrise au sein de l'École des sciences criminelles (Vivas Ramos, 2016), il reste encore du chemin à parcourir. L'objectif est de trouver un équilibre opérationnel entre les taux de faux positifs et de faux négatifs en adaptant les paramètres en fonction des phénomènes criminels.

La capacité d'adaptation de l'approche CFC ouvre également la voie à des applications d'autres types de données. Par exemple, il serait possible d'expérimenter la méthode automatique de détection de tendances sur différents types de traces (p. ex. traces digitales, traces de semelles, traces biologiques, etc.). La criminalité numérique offre également de grandes opportunités d'application, notamment sur la problématique des marchés illicites sur internet et des forums de discussions. Par exemple, la détection de patterns dans les tendances est susceptible d'éclairer l'évolution de la vente en ligne de produits stupéfiants (Broséus et al., 2016), de produits dopants (Pineau et al., 2016) ou de contrefaçons horlogères (Romerio Giudici, 2016).

Les enjeux liés à la découverte de phénomènes criminels non connus méritent aussi d'être étudiés en prenant l'approche CFC comme cadre de travail, de la même manière dont ont été traitées les étapes de classification et de détection au sein de cette thèse. Il s'agirait d'identifier et d'exprimer les processus dans lesquels s'intègre la découverte de phénomènes et de tester des méthodes computationnelles non supervisées adaptées au contexte opérationnel d'un environnement spécifique (unité d'analyse).

Finalement, l'approche CFC soulève la question de l'analyse stratégique. Cette thèse a mis en avant les avantages de l'approche au niveau opérationnel, mais le processus général dans lequel elle s'inscrit (voir chapitre 6) couvre également un volet plus stratégique. Les analyses de tendances en criminologie illustrent régulièrement ce potentiel (Nagin, Farrington, & Moffitt, 1995; Nagin et al., 1995; Weisburd, Bushway, Lum, & Yang, 2004; Weisburd, Groff, & Yang, 2012). Il serait alors prometteur de développer cet aspect en s'interrogeant notamment sur le rôle de la trace dans les processus de renseignement criminel stratégique, ainsi que sur l'expression des processus de production de renseignement à un niveau stratégique. L'avantage de l'approche CFC est de pouvoir raisonner par niveau d'analyse, chacun avec ses spécificités, ses contraintes et ses objectifs, mais tout en étant connecté les uns avec les autres. C'est véritablement la connaissance du processus complet et des différentes inférences en découlant qui permet de guider l'intégration de méthodes computationnelles en analyse et renseignement criminel.

CONCLUSION

Cette thèse a été réalisée en parallèle avec l'émergence du *predictive policing*. Ce mouvement naturellement issu de l'ère du *big data* a stimulé et questionné l'utilisation de plus en plus intensive des modèles computationnels et des technologies en analyse et renseignement criminel. L'élaboration d'une méthodologie réaliste, correctement formalisée et transparente s'impose alors comme un défi prioritaire. Cette problématique soulève des enjeux liés aux libertés publiques, mais également au besoin pour la prise de décisions en matière d'action de sécurité, d'être fondés sur des données probantes.

À travers ce travail, nous avons observé que les méthodes computationnelles semblent être adéquates pour soutenir l'analyse criminelle et ses dispositifs de veille, et ce, tout particulièrement concernant la détection de patterns dans les tendances dans les activités criminelles. Néanmoins, la formalisation de l'approche et des processus démontre que la production de renseignement criminel ne peut être réduite à une baguette magique qui serait capable d'extraire mystérieusement des connaissances pertinentes à partir des données à disposition. Cet idéal véhiculé par certaines approches en data mining n'apparaît pas réaliste en analyse criminelle. Il est recommandé d'injecter des connaissances *a priori* dans les processus, tout en restant ouvert à la découverte de nouveautés.

Cet équilibre subtil a été recherché, en prenant l'unité d'analyse d'un corps de police comme objet d'étude et en collaborant étroitement avec une équipe de recherche en sciences de l'information. Le développement d'une structure de veille à l'aide d'une approche itérative, étape par étape, contribue à consolider et à exprimer de manière plus approfondie sa dimension interdisciplinaire. Cela permet de stimuler la combinaison de connaissance criminologique, forensique et computationnelle au sein d'une approche centrée sur la résolution de problème. Notre réflexion propose un retour aux sources sur la plus élémentaire des pièces en termes d'information sur le crime, c.-à-d. la trace. Le développement subséquent se fonde alors sur les approches situationnelles en criminologie et constitue un guide interdisciplinaire destiné à orienter la recherche vers une intégration efficace et réaliste des modèles computationnels en analyse criminelle. Cependant, le cadre de travail proposé ne s'impose pas en unique solution dès lors qu'il résulte d'une situation spécifique. Il ne

doit pas favoriser l'émergence d'un effet tunnel susceptible d'entraver l'innovation et l'ouverture à de nouvelles perspectives. C'est réellement le processus ayant généré ce cadre de travail qui mérite d'être identifié comme un composant essentiel des sciences criminelles.

Cette problématique s'inscrit sur le long terme, dès lors que la traçabilité des activités criminelles change et augmente perpétuellement. Les promesses algorithmiques d'une automatisation parfaite en analyse criminelle séduisent naturellement, mais négligent souvent la corde raide sur laquelle elles se tiennent. L'équilibre à atteindre est subtil et évolue naturellement avec les innovations technologiques. C'est cette notion d'équilibre qui apparaît alors comme la pierre angulaire de ce travail et qui jalonne ses enjeux et défis à ses différents niveaux : l'équilibre méthodologique entre l'injection de connaissances et l'absence d'*a priori ;* l'équilibre théorique entre les activités litigieuses et les patterns observés dans les données ; l'équilibre pratique entre le potentiel des techniques computationnelles et les spécificités opérationnelles ; l'équilibre interdisciplinaire entre la science forensique, la criminologie et les méthodes computationnelles ; et l'équilibre holistique entre la trace et les connaissances.

« *Il est aussi noble de tendre à l'équilibre qu'à la perfection ; car c'est une perfection que de garder l'équilibre.* » (Jean Grenier)

BIBLIOGRAPHIE

Aepli, P., Ribaux, O., & Summerfield, E. (2011). *Decision Making in Policing: Operations and Management*. Lausanne, Suisse: EPFL Press.

Albertetti, F. (2016). *A Knowledge Extraction Framework for Crime Analysis: Unsupervised Methods in Uncertain Environments*. (Thèse de doctorat). Université de Neuchâtel, Neuchâtel, Suisse.

Albertetti, F., Cotofrei, P., Grossrieder, L., Ribaux, O., & Stoffel, K. (2013a). Crime linkage: A fuzzy MCDM approach. In *2013 IEEE International Conference on Intelligence and Security Informatics (ISI)* (p. 1‑3).

Albertetti, F., Cotofrei, P., Grossrieder, L., Ribaux, O., & Stoffel, K. (2013b). The CriLiM Methodology: Crime Linkage with a Fuzzy MCDM Approach. In *Intelligence and Security Informatics Conference (EISIC), 2013 European* (p. 67‑74).

Albertetti, F., Grossrieder, L., Ribaux, O., & Stoffel, K. (2016). Change points detection in crime-related time series: An on-line fuzzy approach based on a shape space representation. *Applied Soft Computing*, *40*, 441‑454.

Albertetti, F., & Stoffel, K. (2012). From Police Reports to Data Marts: a Step Towards a Crime Analysis Framework. Présenté à 5th International Workshop on Computational Forensics, Tsukuba, Japan: Springer.

Andrews, S., Akhgar, B., Yates, S., Stedmon, A., & Hirsch, L. (2013). Using Formal Concept Analysis to Detect and Monitor Organised Crime. In H. L. Larsen, M. J. Martin-Bautista, M. A. Vila, T. Andreasen, & H. Christiansen (Éd.), *Flexible Query Answering Systems* (p. 124‑133). Berlin: Springer Berlin Heidelberg.

Association Française de Normalisation. (1998). *Prestations de veille et prestations de mise en place d'un système de veille* (normalisation française No. XP X 50-053). Paris: AFNOR.

Azzola, A. (2010). *Délinquance des ressortissants géorgiens dans le canton de Vaud à partir de données policières entre 2000 et 2008* (Mémoire de maîtrise). University of Lausanne, Lausanne, Suisse.

Baechler, S. (2015). *Des faux documents d'identité au renseignement forensique Développement d'une approche systématique et transversale du traitement de la donnée forensique à des fins de renseignement criminel* (Thèse de doctorat). Université de Lausanne, Lausanne, Suisse.

Baechler, S., Cartier, D., Schucany, P., & Guéniat, O. (2015). Les interventions de la police scientifique suite à des cambriolages: quelle est la perception des lésés et y a-t-il lieu de s'en soucier? *Revue Internationale de Criminologie et de Police Technique et Scientifique*, *LXIX*(2), 228‑247.

Baechler, S., Morelato, M., Ribaux, O., Beavis, A., Tahtouh, M., Kirkbride, K. P., Esseiva, P., Margot, P., & Roux, C. (2015). Forensic intelligence framework. Part II: Study of the main generic building blocks and challenges through the examples of illicit drugs and false identity documents monitoring. *Forensic Science International*, *250*, 44‑52.

Baechler, S., Ribaux, O., & Margot, P. (2012). 2012 Student Paper: Toward a Novel Forensic Intelligence Model: Systematic Profiling of False Identity Documents. *Forensic Science Policy & Management: An International Journal*, *3*(2), 70‑84.

Barclay, D. (2009). Using forensic science in major crime inquiries. In J. Fraser & R. Williams (Éd.), *Handbook of Forensic Science* (p. 337‑358). Cullompton: Willan.

Baud, J. (2015). *Les patterns dans les données criminelles* (Mémoire de maîtrise). Université de Lausanne, Lausanne, Suisse.

Bennett, R. R. (1991). DEVELOPMENT AND CRIME: A Cross-National, Time-Series Analysis of Competing Models. *Sociological Quarterly*, *32*(3), 343‑363.

Berry, M. J. A., & Linoff, G. S. (2004). *Data Mining Techniques For Marketing, Sales, and Customer Relationship Management*. Indiana, USA: Wiley.

Birkett, J. (1989). Scientific scene linking. *Journal of the Forensic Science Society*, *29*(4), 271‑284.

Birrer, S. (2010). *Analyse systématique et permanente de la délinquance sérielle: Place des statistiques criminelles; apport des approches situationnelles pour un système de classification; perspectives en matière de coopération* (PhD thesis). University of Lausanne, Lausanne, Suisse.

Bitzer, S., Albertini, N., Lock, E., Ribaux, O., & Delémont, O. (2015). Utility of the clue — From assessing the investigative contribution of forensic science to supporting the decision to use traces. *Science and Justice*, *55*(6), 509‑513.

Bitzer, S., Ribaux, O., Albertini, N., & Delémont, O. (2016). To analyse a trace or not? Evaluating the decision-making process in the criminal investigation. *Forensic Science International*, *262*, 1‑10.

Bloch, A. (1999). *L'intelligence économique* (2e éd.). Paris: Economica.

Boba, R. (2009). *Crime analysis with crime mapping* (2ème). Californie: Sage.

Borg, A., Boldt, M., Lavesson, N., Melander, U., & Boeva, V. (2014). Detecting serial residential burglaries using clustering. *Expert Systems with Applications*, *41*(11), 5252‑5266.

Box-Steffensmeier, J. M., Freeman, J. R., Hitt, M. P., & Pevehouse, J. C. W. (2014). *Time Series Analysis for the Social Sciences*. New York: Cambridge University Press.

Braga, A. A., & Pierce, G. L. (2004). Linking crime guns: the impact of ballistics imaging technology on the productivity of the Boston Police Department's Ballistics Unit. *Journal of Forensic Sciences*, *49*(4), 701‑706.

Brantingham, P. J., & Brantingham, P. L. (1990). Situational crime prevention in practice. *Canadian Journal of Criminology*, *32*, 17‑40.

Brantingham, P. L., & Brantingham, P. J. (1993). Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology*, *13*(1), 3‑28.

Brantingham, Paul J., & Brantingham, P. L. (1978). Theoretical model of crime site selection. In *Crime. law and sanctions* (Sage). Beverly Hills, CA: M. D. Krohn & R. L. Akers.

Brantingham, Paul J., & Brantingham, P. L. (1981). *Environmental Criminology* (Sage). Beverly Hills, CA.

Bratton, W. J., & Malinowski, S. W. (2008). Police Performance Management in Practice: Taking COMPSTAT to the Next Level. *Policing: A Journal of Policy and Practice*, *2*(3), 259‑265.

Broséus, J., Rhumorbarbe, D., Mireault, C., Ouellette, V., Crispino, F., & Décary-Hétu, D. (2016). Studying illicit drug trafficking on Darknet markets: Structure and organisation from a Canadian perspective. *Forensic Science International*, *264*, 7‑14.

Brown, C., Ross, A., & Attewell, R. G. (2014). Benchmarking Forensic Performance in Australia–Volume Crime. *Forensic Science Policy & Management: An International Journal*, *5*(3‑4), 91‑98.

Bruce, C. W. (2008). The Patrol Route Monitor: a Modern Approach to Threshold Analysis. *International Association of Crime Analysts*.

Cantor, D., & Land, K. C. (1985). Unemployment and Crime Rates in the Post-World War II United States: A Theoretical and Empirical Analysis. *American Sociological Review*, *50*(3), 317‑332.

Cao, L. (2008). Domain Driven Data Mining. In *IEEE International Conference on Data Mining Workshops* (p. 74‑76). F. Bonchi et al. Consulté à l'adresse http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=4733896

Cao, L., Yu, P. S., Zhang, C., & Zhao, Y. (2010). *Domain Driven Data Mining*. New York, USA: Springer.

Chen, H., Chung, W., Qin, Y., Chau, M., Xu, J. J., Wang, G., … Atabakhsh, H. (2003). Crime Data Mining: An Overview and Case Studies. In *Proceedings of the 2003 Annual National Conference on Digital Government Research* (p. 1–5). Boston, MA: Digital Government Society of North America. Consulté à l'adresse http://dl.acm.org/citation.cfm?id=1123196.1123231

Chilvers, M., & Weatherburn, D. (2001). Operation and crime review panels: Their impact on break and enter. *Crime and Justice Statistics Bureau Brief*.

Chopin, J. (2017). *La gestion des liens entre les crimes sexuels de prédation□: repenser ViCLAS sous la perspective du paradigme situationnel* (Thèse de doctorat). Université de Lausanne, Lausanne, Suisse.

Clarke, R. V., & Eck, J. E. (2005). *Crime Analysis for Problem Solvers In 60 Small Steps*. USA: Center for Problem Oriented Policing.

Clarke, R. V. G. (1997). *Situational Crime Prevention: Successful Case Studies* (2nd Revised edition). USA: Criminal Justice Press.

Cleland, C. E. (2011). Historical science, experimental science, and the scientific method. *Geology*, *29*(11), 987‑990.

Cohen, L. E., & Felson, M. (1979). Social Change and Crime Rate Trends: A Routine Activity Approach. *American Sociological Review*, *44*(4), 588‑608.

Committee on Facilitating Interdisciplinary Research. (2004). *Facilitating Interdisciplinary Research*. Washington DC: THE NATIONAL ACADEMIES PRESS. Consulté à l'adresse http://www.nap.edu/catalog/11153/facilitating-interdisciplinary-research

Cornish, D. B. (1994). The Procedural Analysis of Offending and its Relevance for Situational Prevention. In R. V. Clarke (Éd.), *Crime Prevention Studies* (Vol. 3, p. 151‑196). Monsey, NY: Criminal Justice Press.

Cornish, D. B., & Clarke, R. V. (1986). *The Reasoning Criminal: Rational Choice Perspectives on Offending*. Secaucus, NJ: Springer-Verlag.

Cotofrei, P., & Stoffel, K. (2011). Fuzzy extended BPMN for modelling crime analysis processes. In *Proceedings of First Int. Symposium on Data - Driven Process Discovery and Analysis* (p. 13‑28).

Crispino, F. (2006). *Le principe de Locard est-il scientifique□? Ou analyse de la scientificité des principes fondamentaux de la criminalistique*. (Thèse de doctorat). Université de Lausanne, Lausanne, Suisse.

Crispino, Frank. (2008). Nature and place of crime scene management within forensic sciences. *Science and Justice*, *48*(1), 24‑28.

Cukier, K. N., & Mayer-Schoenberger, V. (2013). The Rise of Big Data: How It's Changing the Way We Think About the World. *Foreign Affairs*, *May/June 2013*.

Cusson, M. (2005). *La délinquance, une vie choisie: entre plaisir et crime*. Montréal: Hurtubise HMH.

Cusson, M. (2008). Répétitions criminelles, renseignements et opérations coup-de-poing. *Problèmes actuels de science criminelle*, (21), 37‑52.

Cusson, M., & Ribaux, O. (2015). Vers une méthode commune à la police scientifique et à la criminologie. *Revue Internationale de Criminologie et de Police Technique et Scientifique*, *68*(3), 266‑283.

Dégardin, K., Roggo, Y., & Margot, P. (2015). Forensic intelligence for medicine anti-counterfeiting. *Forensic Science International*, *248*, 15‑32.

Dessimoz, D., & Champod, C. (2016). A dedicated framework for weak biometrics in forensic science for investigation and intelligence purposes: The case of facial information. *Security Journal*, *29*(4), 603‑617.

Dupont, B. (2016). Des effets perturbateurs de la technologie sur la criminologie. *Revue Internationale de Criminologie et de Police Technique et Scientifique*, *69*(3), 305‑322.

Egger, S. A. (1984). Working Definition of Serial Murder and the Reduction of Linkage Blindness. *Journal of Police Science and Administration*, *12*(3), 348‑357.

Elhay, S., Golub, G., & Kautsky, J. (1991). Updating and Downdating of Orthogonal Polynomials with Data Fitting Applications. *SIAM Journal on Matrix Analysis and Applications*, *12*(2), 327‑353.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, *17*, 37‑54.

Felson, M., & Clarke, R. V. (1998). *Opportunity makes the thief: Practical theory for crime prevention*. Londres: Home Office.

Filipov, V., Mukhanov, L., & Shchukin, B. (2008). Transaction aggregation as strategy for credit card fraud detection. In *Cybernetic Intelligent Systems CIS 2008*. London.

Franke, K., & Srihari, S. N. (2008). Computational Forensics: An Overview. In S. N. Srihari & K. Franke (Éd.), *Computational Forensics* (p. 1‑10). Berlin: Springer Berlin Heidelberg.

Friend, Z. (2013). Predictive Policing: Using Technology to Reduce Crime. *FBI Law Enforcement Bulletin*. Consulté à l'adresse: http://leb.fbi.gov/2013/april/predictive-policing-using-technology-to-reduce-crime

Fuchs, E., Gruber, T., Nitschke, J., & Sick, B. (2010). Online Segmentation of Time Series Based on Polynomial Least-Squares Approximations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(12), 2232‑2245.

Greenberg, D. F. (2001). Time Series Analysis of Crime Rates. *Journal of Quantitative Criminology*, *17*(4), 291‑327.

Grossrieder, L., Albertetti, F., Stoffel, K., & Ribaux, O. (2013). Des données aux connaissances, un chemin difficile: réflexion sur la place du data mining en analyse criminelle. *Revue Internationale de Criminologie et de Police Technique et Scientifique*, *66*(1), 99‑116.

Grubesic, T. H. (2006). On The Application of Fuzzy Clustering for Crime Hot Spot Detection. *Journal of Quantitative Criminology*, *22*(1), 77‑105.

Gustafsson, F. (1996). The marginalized likelihood ratio test for detecting abrupt changes. *IEEE Transactions on Automatic Control*, *41*(1), 66‑78.

Hane, T. (2015). La veille, une autre forme d'acquisition et de mise en valeur de l'information. In *L'intelligence économique au service de la lutte contre le blanchiment de capitaux et le financement du terrorisme* (p. 250‑269). Strasbourg, France: Université de Strasbourg.

Hazard, D. (2014). *La pertinence en science forensique: une (en)quête épistémologique et empirique* (Thèse de doctorat). Université de Lausanne.

Heaton, R. (2000). The prospects for intelligence-led policing: Some Historical and quantitative considerations. *Policing and Society*, *9*(4), 337‑355.

Hido, S., Idé, T., Kashima, H., Kubo, H., & Matsuzawa, H. (2008). Unsupervised Change Analysis Using Supervised Learning. In T. Washio, E. Suzuki, K. M. Ting, & A. Inokuchi (Éd.), *Advances in Knowledge Discovery and Data Mining* (p. 148‑159). Berlin: Springer Berlin Heidelberg.

Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, *332*(6025), 60‑65.

Hofmann, H. F., Pfeifer, R., & Vinkhuyzen, E. (1993). *Situated Software Design*.

Holmes, M. P., Wang, Y., & Ziedins, I. (2009). *The application of data mining tools and statistical techniques to identify patterns and changes in fire events* (Report). Consulté à l'adresse: https://researchspace.auckland.ac.nz/handle/2292/9810

Inman, K., & Rudin, N. (2000). *Principles and Practice of Criminalistics: The Profession of Forensic Science* (1 edition). Boca Raton, Fla: CRC Press.

Kahneman, D. (2013). *Thinking, Fast and Slow* (Reprint). New York: Farrar Straus Giroux.

Kawahara, Y., & Sugiyama, M. (2009). Change-Point Detection in Time-Series Data by Direct Density-Ratio Estimation. In *Proceedings of the 2009 SIAM International Conference on Data Mining* (Vol. 1–0, p. 389‑400). Society for Industrial and Applied Mathematics. Consulté à l'adresse: http://epubs.siam.org/doi/abs/10.1137/1.9781611972795.34

Kedma, G., Guri, M., Sela, T., & Elovici, Y. (2013). Analyzing users' web surfing patterns to trace terrorists and criminals (p. 143‑145). Présenté à International Conference on Intelligence and Security Informatics (ISI), IEE.

Kempf, K. (1986). Offense Specialization: Does It Exist? In D. B. Cornish & R. V. Clarke (Éd.), *The Reasoning Criminal: Rational Choice Perspectives on Offending* (p. 186‑201). New Brunswick, USA: Transaction Publishers.

Kind, S. (1987). *The Scientific Investigation of Crime*. Harrogate, UK: Forensic Science Service.

la Conférence latine des Chefs des Départements de Justic et Police (CLDJP). (2013, mars 25). Les cantons romands renforcent leur coopération policière. Consulté à l'adresse: http://www.cldjp.ch/data/police/com_presse-police-13.pdf

Lammers, M. (2013). Are Arrested and Non-Arrested Serial Offenders Different? A Test of Spatial Offending Patterns Using DNA Found at Crime Scenes. *Journal of Research in Crime and Delinquency*, 0022427813504097.

Lammers, M., Bernasco, W., & Elffers, H. (2012). How Long Do Offenders Escape Arrest? Using DNA Traces to Analyse when Serial Offenders Are Caught. *Journal of Investigative Psychology and Offender Profiling*, *9*(1), 13‑29.

Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity, and Variety*. META Group. Consulté à l'adresse: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Li, S.-T., Kuo, S.-C., & Tsai, F.-C. (2010). An intelligent decision-support model using FSOM and rule extraction for crime prevention. *Expert Systems with Applications*, *37*(10), 7108‑7119.

Liu, S., Yamada, M., Collier, N., & Sugiyama, M. (2013). Change-point detection in time-series data by relative density-ratio estimation. *Neural Networks*, *43*, 72‑83.

Locard, E. (1920). *L'enquête criminelle et les méthodes scientifiques*. E. Flammarion.

Maguire, M. (1982). *Burglary in a Dwelling: The OffenCe, the Offender and the Victim*. London: Heinemann Educational Books.

Margot, P. (2003). *Cours de science forensique* (Institut de Police Scientifique, Université de Lausanne). Lausanne, Suisse.

Margot, P. (2011). Commentary on the Need for a Research Culture in the Forensic Sciences. *UCLA Law Review*, *58*, 795‑801.

Maurushat, A., Bennett-Moses, L., & Vaile, D. (2015). Using « Big » Metadata for Criminal Intelligence: Understanding Limitations and Appropriate Safeguards. In *Proceedings of the 15th International Conference on Artificial Intelligence and Law* (p. 196–200). New York, NY, USA: ACM.

Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. E. (2011). Self-Exciting Point Process Modeling of Crime. *journal of the American Statistical Association*, *106*(493), 100‑108.

Morelato, M., Baechler, S., Ribaux, O., Beavis, A., Tahtouh, M., Kirkbride, P., … Margot, P. (2014). Forensic intelligence framework—Part I: Induction of a transversal model by comparing illicit drugs and false identity documents monitoring. *Forensic Science International*, *236*, 181‑190.

Nagin, D. S., Farrington, D. P., & Moffitt, T. E. (1995). Life-Course Trajectories of Different Types of Offenders*. *Criminology*, *33*(1), 111‑139.

Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, *50*(3), 559‑569.

Nissan, E. (2012). An overview of data mining for combating crime. *Applied Artificial Intelligence: An international Journal*, *26*(8), 760‑786.

Pareto, V. (1965). *Écrits sur la courbe de répartition de la richesses* (version original, 1896). Genève, Suisse: Droz.

Pearsall, B. (2010). Predictive Policing: The Future of Law Enforcement? *National Institute of Justice*, *266*. Consulté à l'adresse: http://www.nij.gov/journals/266/pages/predictive.aspx

Pease, K. (2010). Crime Science. In S. Shoham, P. Knepper, & M. A. Kett (Éd.), *International Handbook of Criminology*. Boca Raton, FL: CRC Press.

Peirce, C. S. (1935). *Collected Papers of Charles Sanders Peirce*. Cambridge: Harvard University Press.

Perry, W. L., McInnis, B., Price, C. C., Smith, S. C., & Hollywood, J. S. (2013). *PREDICTIVE POLICING: The Role of Crime Forecasting in Law Enforcement Operations* (RAND Corporation research report series No. 2010- IJ- CX-K007). USA: National Institute of Justice. Consulté à l'adresse: https://www.ncjrs.gov/pdffiles1/nij/grants/243830.pdf

Phelps, C. (2010). IBM predictive analytics help slash crime rates in Memphis. *Public Safety IT*.

Phillips, P., & Lee, I. (2012). Mining co-distribution patterns for large crime datasets. *Expert Systems with Applications*, *39*(14), 11556‑11563.

Pineau, T., Schopfer, A., Grossrieder, L., Broséus, J., Esseiva, P., & Rossy, Q. (2016). The study of doping market: how to produce intelligence from Internet forums. *Forensic Science International*.

Ratcliffe, J. H. (2008). *Intelligence-Led Policing*. Cullompton: Willan Publishing.

Ratcliffe, J. H. (2016). *Intelligence-Led Policing* (2nd éd.). Londres: Routledge.

Ratcliffe, Jerry H. (2011). Intelligence-led policing: Anticipating risk and influencing action. In *Criminal Intelligence for the 21st Centur* (p. 206‑220). Wright, R, Morehouse, B, Peterson, MB & Palmieri, L.

Ratle, F., Gagné, C., Terrettaz-Zufferey, A.-L., Kanevski, M., Esseiva, P., & Ribaux, O. (2008). Advanced clustering methods for mining chemical databases in forensic science. *Chemometrics and Intelligent Laboratory Systems*, *90*(2), 123‑131.

Reardon, B., Nance, K., & McCombie, S. (2012). Visualization of ATM Usage Patterns to Detect Counterfeit Cards Usage. In *2012 45th Hawaii International Conference on System Science (HICSS)* (p. 3081‑3088).

Ribaux, O, & Margot, P. (1999). Inference structures for crime analysis and intelligence: the example of burglary using forensic science data. *Forensic Science International*, *100*(3), 193‑210.

Ribaux, Olivier. (1997). *La recherche et gestion des liens dans l'investigation criminelle: Le cas particulier du cambriolage* (PhD thesis). University of Lausanne, Lausanne, Suisse.

Ribaux, Olivier. (2014). *Police scientifique : Le renseignement par la trace*. Lausanne: PPUR Presses Polytechniques.

Ribaux, Olivier, Baylon, A., Lock, E., Delémont, O., Roux, C., Zingg, C., & Margot, P. (2010). Intelligence-led crime scene processing. Part II: Intelligence and crime scene examination. *Forensic Science International*, *199*(1–3), 63‑71.

Ribaux, Olivier, Baylon, A., Roux, C., Delémont, O., Lock, E., Zingg, C., & Margot, P. (2010). Intelligence-led crime scene processing. Part I: Forensic intelligence. *Forensic Science International*, *195*(1–3), 10‑16.

Ribaux, Olivier, & Birrer, S. (2008). Système de suivi et d'analyse des cambriolages appliqué dans des polices suisses. In C. Schwarzenegger & J. Müller (Éd.), *Erstes Zürcher Präventionsforum: Kommunale Kriminalprävention Crime Mapping Einbruchskriminalität* (p. 189‑205). Europa Institut Zürich.

Ribaux, Olivier, & Birrer, S. (2010). Iterative Development of Co-operation within an Increasingly Complex Environment. Example of a Swiss Regional Analysis Centre. In F. Lemieux (Éd.), *International Police Cooperation: Emerging Issues, Theory and Practice* (p. 81‑100). Cullompton: Willan.

Ribaux, Olivier, Crispino, F., Delémont, O., & Roux, C. (2015). The Progressive Opening of Forensic Science Towards Criminological Concerns. *Security Journal*.

Ribaux, Olivier, Crispino, F., & Roux, C. (2015). Forensic intelligence: deregulation or return to the roots of forensic science? *Australian Journal of Forensic Sciences*, *47*(1), 61‑71.

Ribaux, Olivier, Genessay, T., & Margot, P. (2011). Les processus de veille opérationnelle et science forensique. In S. Leman-Langlois (Éd.), *Sphères de surveillance* (p. 137‑158). Montréal: Les Presses de l'Université de Montréal.

Ribaux, Olivier, & Margot, P. (2003). Case based reasoning in criminal intelligence using forensic case data. *Science & Justice*, *43*(3), 135‑143.

Ribaux, Olivier, & Margot, P. (2007). La trace matérielle, vecteur d'information au service du renseignement. In M. Cusson, B. Dupont, & F. Lemieux, *Traité de sécurité intérieure* (p. 300‑321). Québec, Canada: Cahiers du Québec.

Ribaux, Olivier, Taroni, F., & Margot, P. (1995). La recherche et la gestion des liens dans l'investigation criminelle: une étape vers l'exploitation systématique des données de police. *Revue internationale de Criminologie et de Police Technique*, (2), 229‑242.

Ribaux, Olivier, Walsh, S. J., & Margot, P. (2006). The contribution of forensic science to crime analysis and investigation: Forensic intelligence. *Forensic Science International*, *156*(2‑3), 171‑181.

Ritter, N. (2008). DNA Solves Property Crimes (But Are We Ready for That?). *NIJ Journal*, *261*, 2‑12.

Rodrigues, S. (2012). *The detection of series of burglaries by analysis of spatiotemporal shoemark patterns.* (Master thesis not published). University of Lausanne, Lausanne, Switzerland.

Romerio Giudici, C. (2016). *Analyse de forums Internet: une source de renseignement sur la contrefaçon horlogère* (Mémoire de maîtrise). Université de Lausanne, Lausanne, Suisse.

Rossy, Q. (2011). *Méthodes de visualisation en analyse criminelle: approche générale de conception des schémas relationnels et développement d'un catalogue de patterns* (Thèse de doctorat). Université de Lausanne, Lausanne, Suisse.

Rossy, Q., Ioset, S., Dessimoz, D., & Ribaux, O. (2013). Integrating forensic information in a crime intelligence database. *Forensic Science International*, *230*(1), 137‑146.

Rossy, Q., & Ribaux, O. (2014). A collaborative approach for incorporating forensic case data into crime investigation using criminal intelligence analysis and visualisation. *Science and Justice*, *54*(2), 146‑153.

Rouach, D. (2010). *La veille technologique et l'intelligence économique* (5e éd.). Paris: Presses Universitaires de France - PUF.

Santos, R. B. (2014). The Effectiveness of Crime Analysis for Crime Reduction Cure or Diagnosis? *Journal of Contemporary Criminal Justice*, *30*(2), 147‑168.

Schroeder, J., Xu, J., & Chen, H. (2003). CrimeLink Explorer: Using Domain Knowledge to Facilitate Automated Crime Association Analysis. In H. Chen, R. Miranda, D. D. Zeng, C. Demchak, J. Schroeder, & T. Madhusudan (Éd.), *Intelligence and Security Informatics* (p. 168‑180). Berlin: Springer Berlin Heidelberg.

Schroeder, J., Xu, J., Chen, H., & Chau, M. (2007). Automated criminal link analysis based on domain knowledge. *Journal of the American Society for Information Science and Technology*, *58*(6), 842‑855.

Schuliar, Y., & Crispino, F. (2013). Semiotics, Heuristics, and Inferences Used by Forensic Scientists. In J. Siegel & P. Saukko (Éd.), *Encyclopedia of forensic Sciences* (2nd edition, Vol. 3, p. 310‑313). Waltham: Academic Press.

Sghaier, M. (2015). *Science forensique et data mining* (Mémoire de maîtrise). Université de Lausanne, Lausanne, Suisse.

Sherman, L. W., Gartin, P. R., & Buerger, M. E. (1989). Hot Spots of Predatory Crime: Routine Activities and the Criminology of Place. *Criminology*, *27*(1), 27‑56.

Soergel, D. (1994). Indexing and retrieval performance: The logical evidence. *Journal of the American Society for Information Science*, *45*(8), 589‑599.

Sorensen, D. (2003). *The nature and prevention of residential burglary: A review of the international literature with an eye toward prevention in Denmark*. Copenhagen, Denmark: Justitsministeriet.

Stember, M. (1991). Advancing the social sciences through the interdisciplinary enterprise. *The Social Science Journal*, *28*(1), 1‑14.

Stoffel, K., Cotofrei, P., & Han, D. (2010). Fuzzy methods for forensic data analysis. In *2010 International Conference of Soft Computing and Pattern Recognition (SoCPaR)* (p. 23‑28). Paris.

Sutherland, E. H., & Cressey, D. R. (1966). *Principles of criminology*. Philadelphia: Lippincott. Consulté à l'adresse: http://catalog.hathitrust.org/api/volumes/oclc/264420.html

Terrettaz-Zufferey, A.-L., Ratle, F., Ribaux, O., Esseiva, P., & Kanevski, M. (2006). Assessment of data mining methods for forensic case data analysis. *Journal of Criminal Justice and Security*, *Special issue no 3-4*, 350‑355.

Tremblay, P. (2010). *Le délinquant idéal: performance, discipline, solidarité*. Montreal, CA: Liber.

Tufféry, S. (2007). *Data Mining et statistique décisionnelle: L'intelligence des données*. Paris: TECHNIP.

Vivas Ramos, M. (2016). *Analyse de tendances de phénomènes criminels par data mining* (Mémoire de maîtrise). Université de Lausanne, Lausanne, Suisse.

Vlahos, J. (2012). The Department of Pre-Crime. *Scientific American*, *306*, 62‑67.

Vollmer, A. (1919). Revision of the Atcherley Modus Operandi System. *Journal of the American Institute of Criminal Law and Criminology*, *10*(2), 229.

Weisburd, D., Bushway, S., Lum, C., & Yang, S.-M. (2004). Trajectories of Crime at Places: A Longitudinal Study of Street Segments in the City of Seattle*. *Criminology*, *42*(2), 283–322.

Weisburd, D., Groff, E. R., & Yang, S.-M. (2012). *The criminology of place*. New York, USA: Oxford university Press.

Weisel, D. L. (2005). *Analyzing Repeat Victimization* (Center for Problem Oriented Policing). USA.

Whitrow, C., Hand, D. J., Juszczak, P., Weston, D., & Adams, N. M. (2009). Transaction aggregation as a strategy for credit card fraud detection. *Data Mining and Knowledge Discovery*, *18*(1), 30‑55.

Wolfgang, M. E. (1987). *Delinquency in a Birth Cohort*. University of Chicago Press.

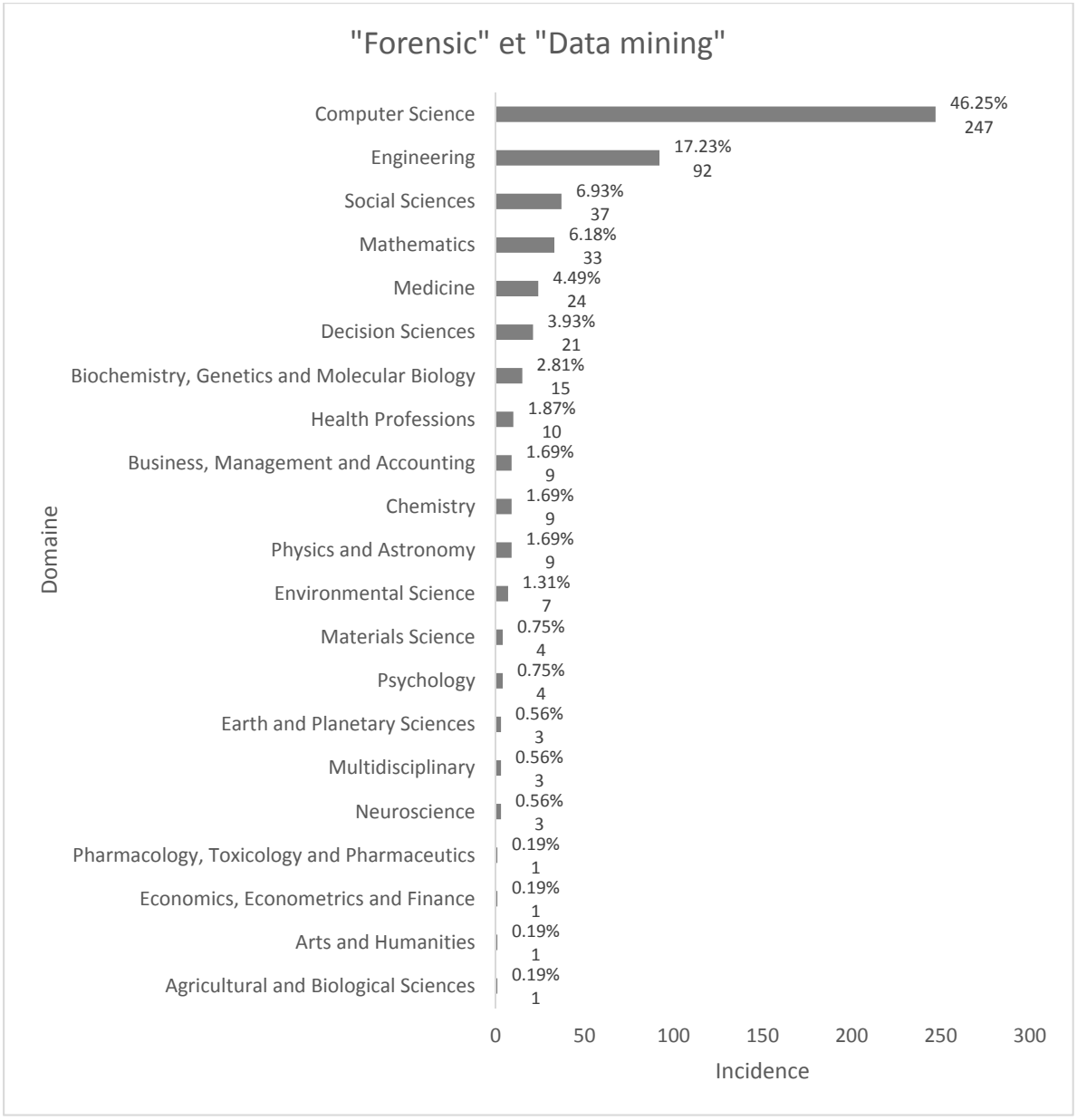
Yamada, M., Kimura, A., Naya, F., & Sawada, H. (2013). Change-point detection with feature selection in high dimensional time-series data. In *Proceedings of International Joint Conference on Artificial Intelligence* (p. 1827‑1833). Beijing, China.

Young, W. T., Goldberg, H. G., Memory, A., Sartain, J. F., & Senator, T. E. (2013). Use of Domain Knowledge to Detect Insider Threats in Computer Activities. In *2013 IEEE Security and Privacy Workshops (SPW)* (p. 60‑67).

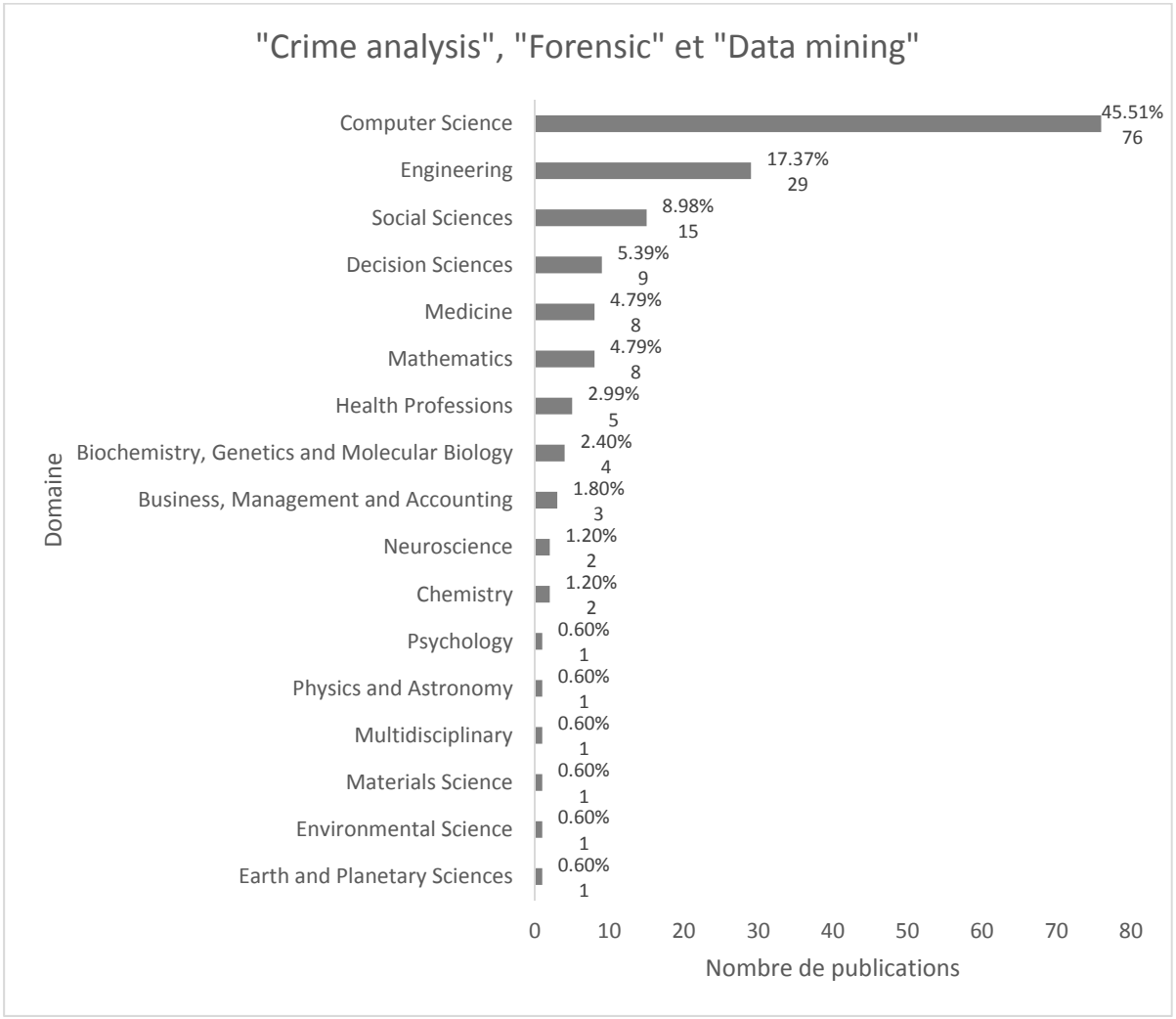
ANNEXES

Annexes A: Analyse documentaire - Graphiques complémentaires

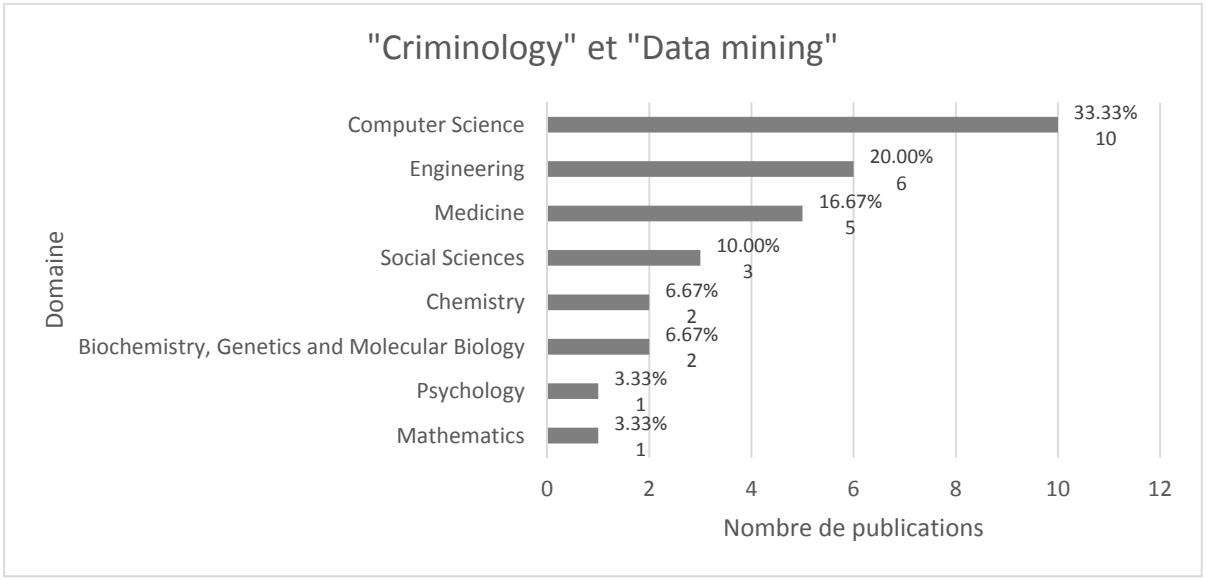
Annexe 1 *: Taux d'incidence des termes « forensic » et « data mining » dans les titres, mots-clés et résumé des publications scientifiques répertoriées sur SCOPUS en fonction du domaine de publication (n= 534). Source: www.scopus.com, état au 04.08.2015.*



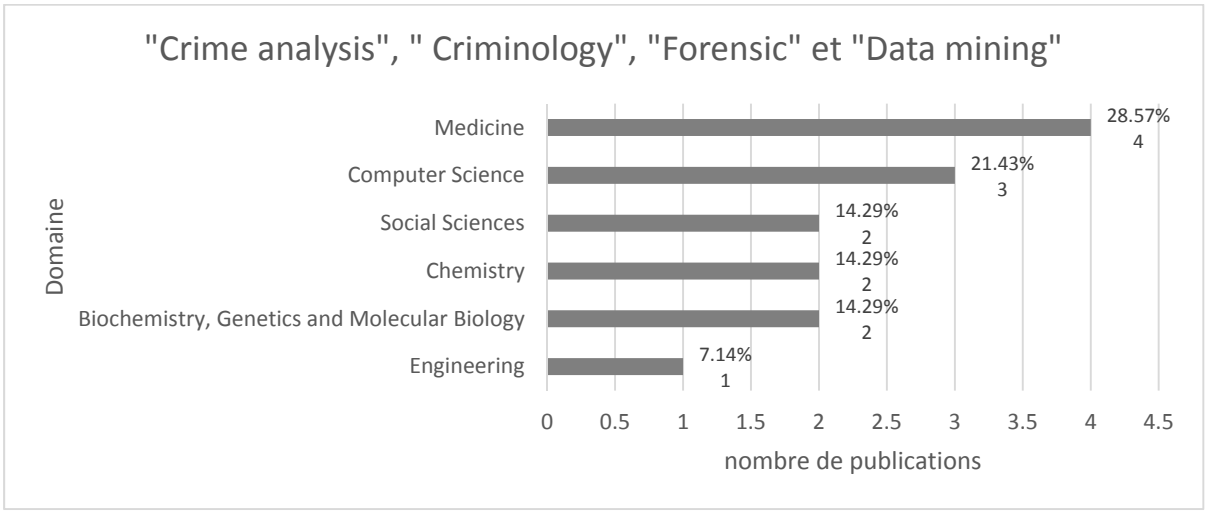
***Annexe 2** : Taux d'incidence des termes « crime analysis », « forensic » et « data mining » dans les titres, mots-clés et résumé des publications scientifiques répertoriées sur SCOPUS en fonction du domaine de publication (n= 167). Source: www.scopus.com, état au 04.08.2015.*



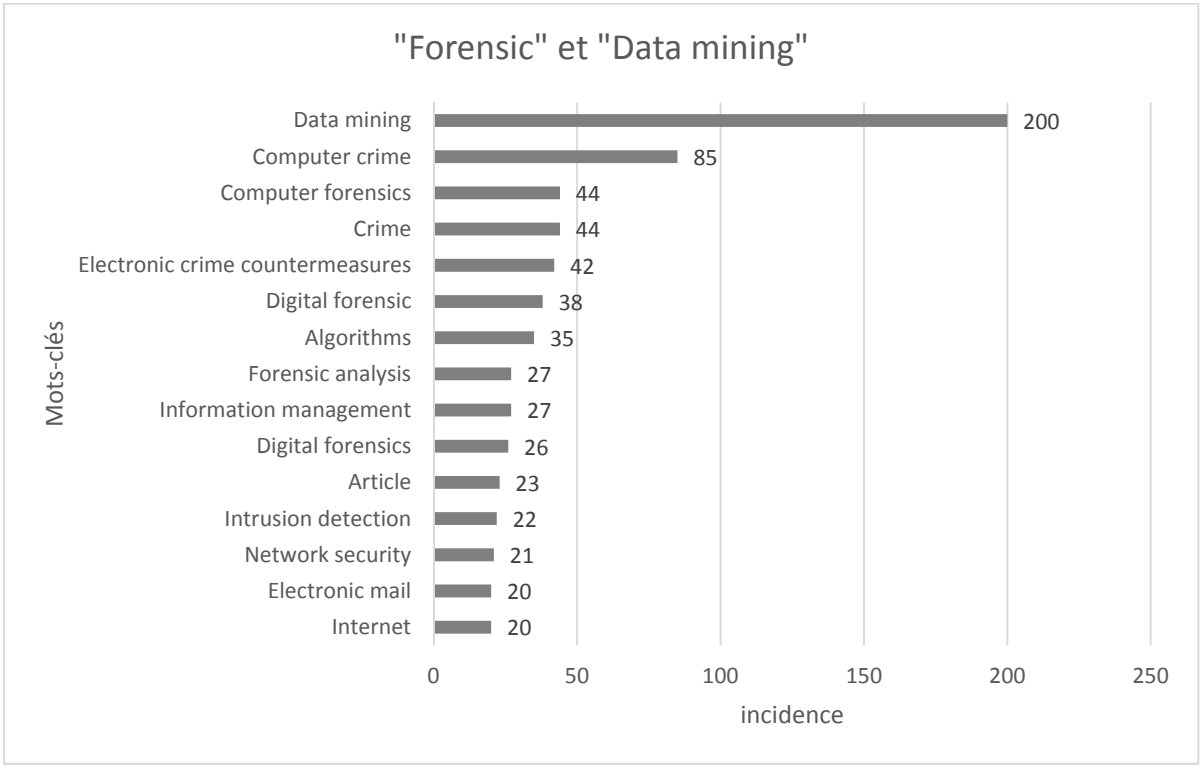
Annexe 3 *: Taux d'incidence des termes « criminology » et « data mining » dans les titres, mots-clés et résumé des publications scientifiques répertoriées sur SCOPUS en fonction du domaine de publication (n= 30). Source: www.scopus.com, état au 04.08.2015.*



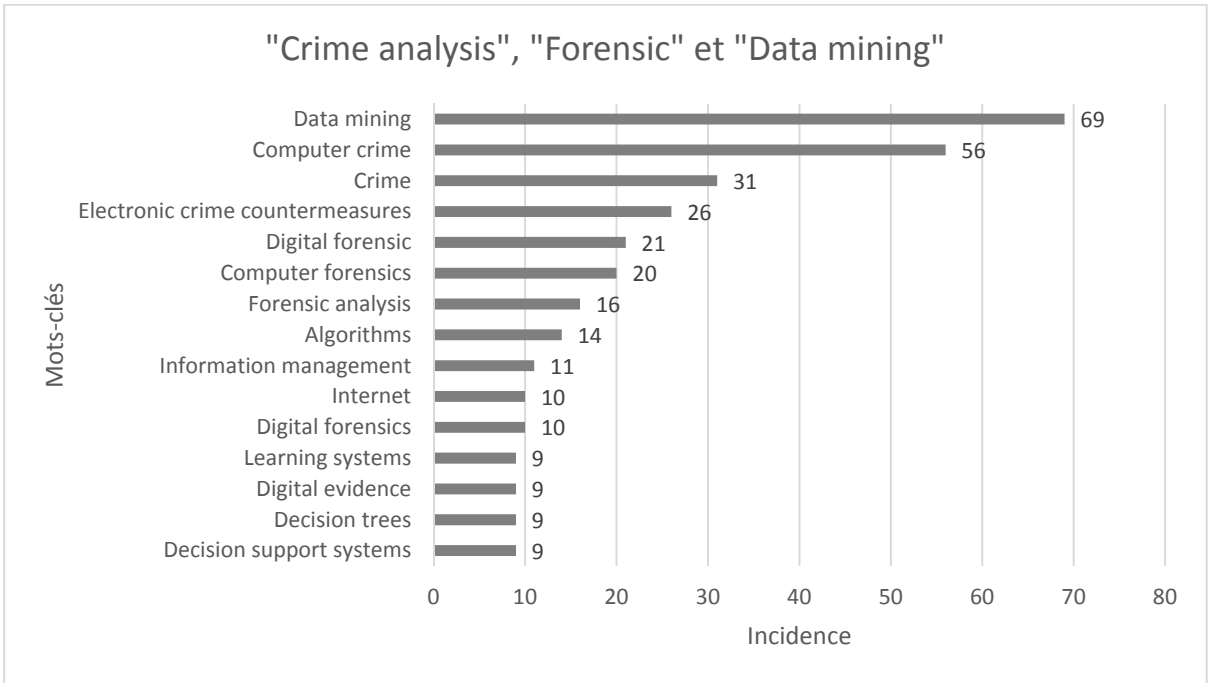
Annexe 4 *: Taux d'incidence des termes « crime analysis », « criminology », « forensic » et « data mining » dans les titres, mots-clés et résumé des publications scientifiques répertoriées sur SCOPUS en fonction du domaine de publication (n= 14). Source: www.scopus.com, état au 04.08.2015.*



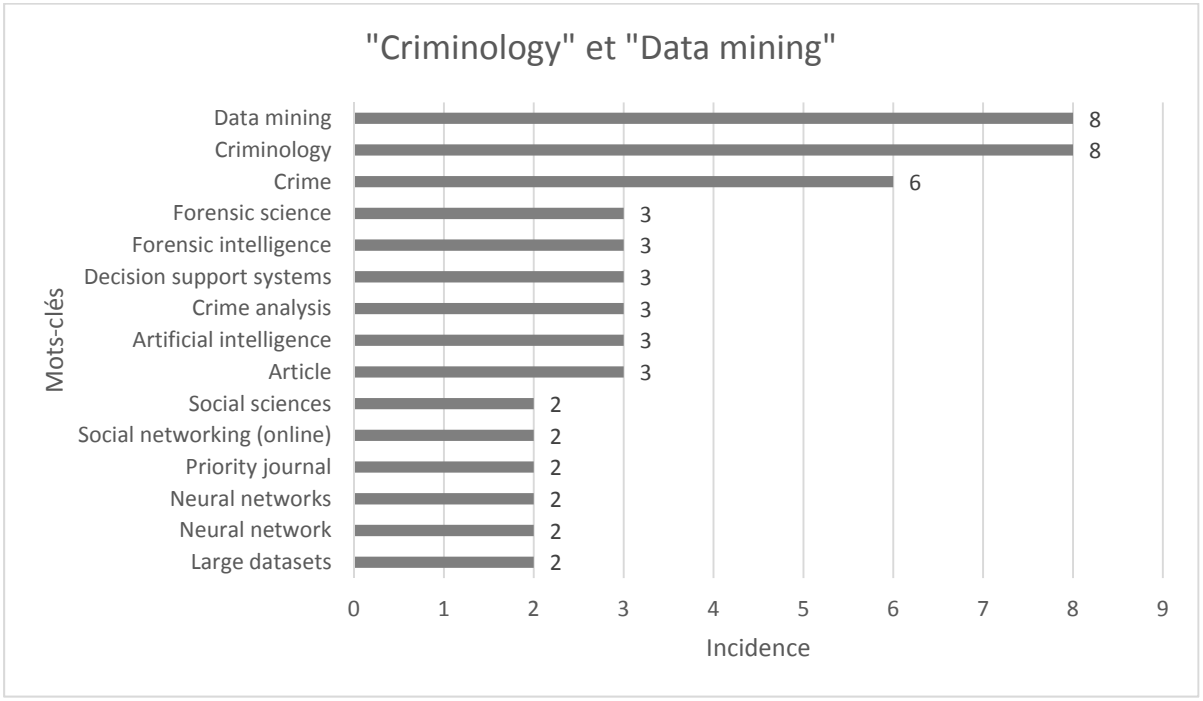
***Annexe 5** : Taux d'incidence des termes « forensic » et « data mining » dans les titres, mots-clés et résumé des publications scientifiques répertoriées sur SCOPUS en fonction des mots-clés (n = 1'773). Classement des 15 scores les plus élevés (n= 674; 38.01%). Source: www.scopus.com, état au 04.08.2015.*



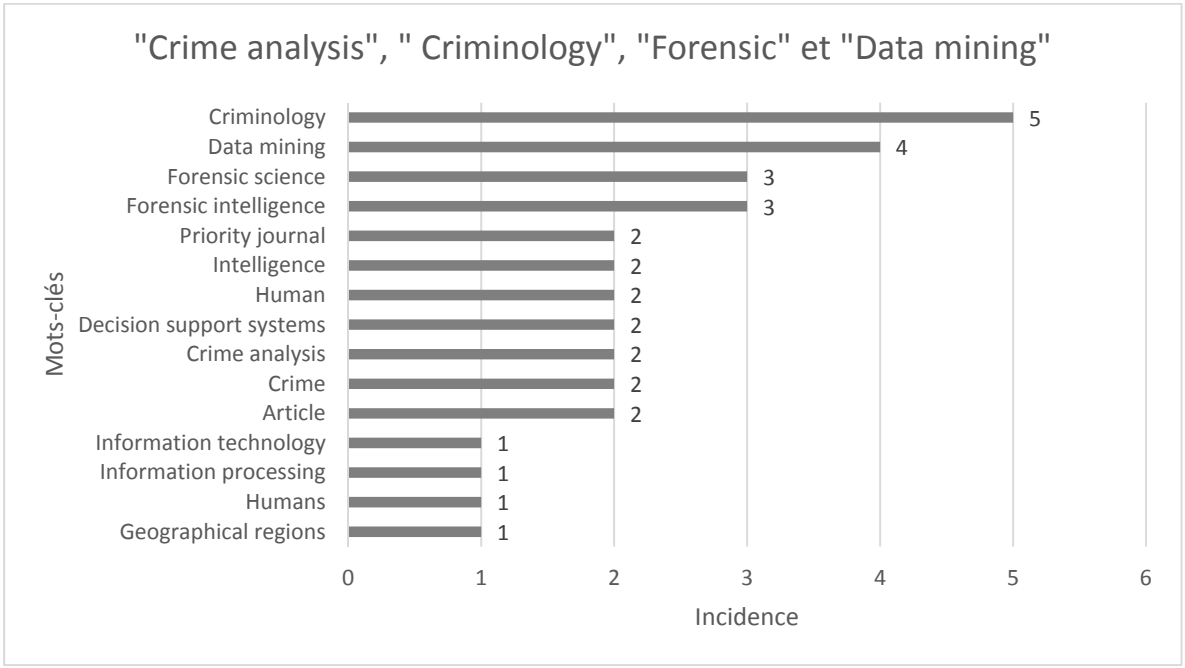
***Annexe 6** : Taux d'incidence des termes « crime analysis », « forensic » et « data mining » dans les titres, mots-clés et résumé des publications scientifiques répertoriées sur SCOPUS en fonction des mots-clés (n= 818). Classement des 15 scores les plus élevés (n= 320; 39.12%). Source: www.scopus.com, état au 04.08.2015.*



***Annexe 7** : Taux d'incidence des termes « criminology » et « data mining » dans les titres, mots-clés et résumé des publications scientifiques répertoriées sur SCOPUS en fonction des mots-clés (N = 205). Classement des 15 scores les plus élevés (n= 52; 25.37%). Source: www.scopus.com, état au 04.08.2015.*



***Annexe 8** : Taux d'incidence des termes « crime analysis », « criminology », « forensic » et « data mining » dans les titres, mots-clés et résumé des publications scientifiques répertoriées sur SCOPUS en fonction des mots-clés (n= 58). Classement des 15 scores les plus élevés (n= 33; 56.90%). Source: www.scopus.com, état au 04.08.2015.*



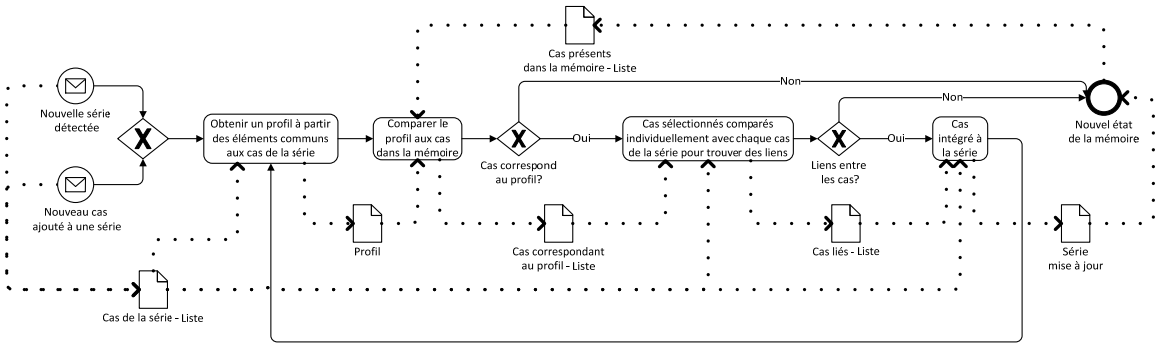
Annexes B : Utilité du pattern – Exemple

Annexe 9 *: Affiche de campagne de prévention contre les cambriolages du soir (source : http://mediapolice.ch/fr/telechargements/category/prevention-contre-les-cambriolages-au-crepuscule-sera-fr).*

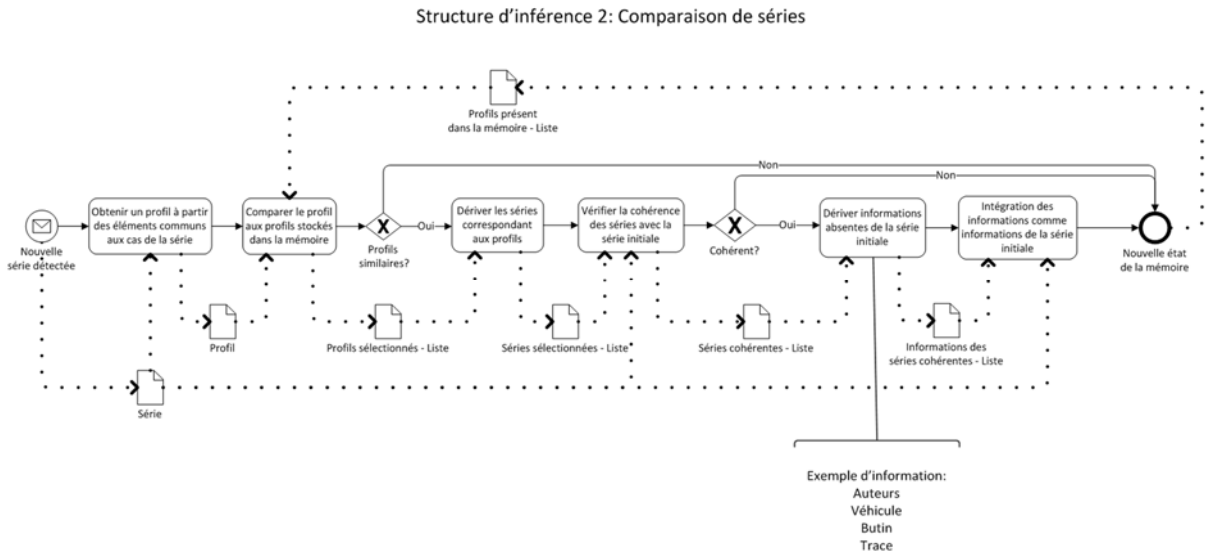
Annexes C: Structures d'inférences en renseignement criminel – processus BPMN

Annexe 10 *: Structure d'inférence 1 : classification à travers le profilage.*

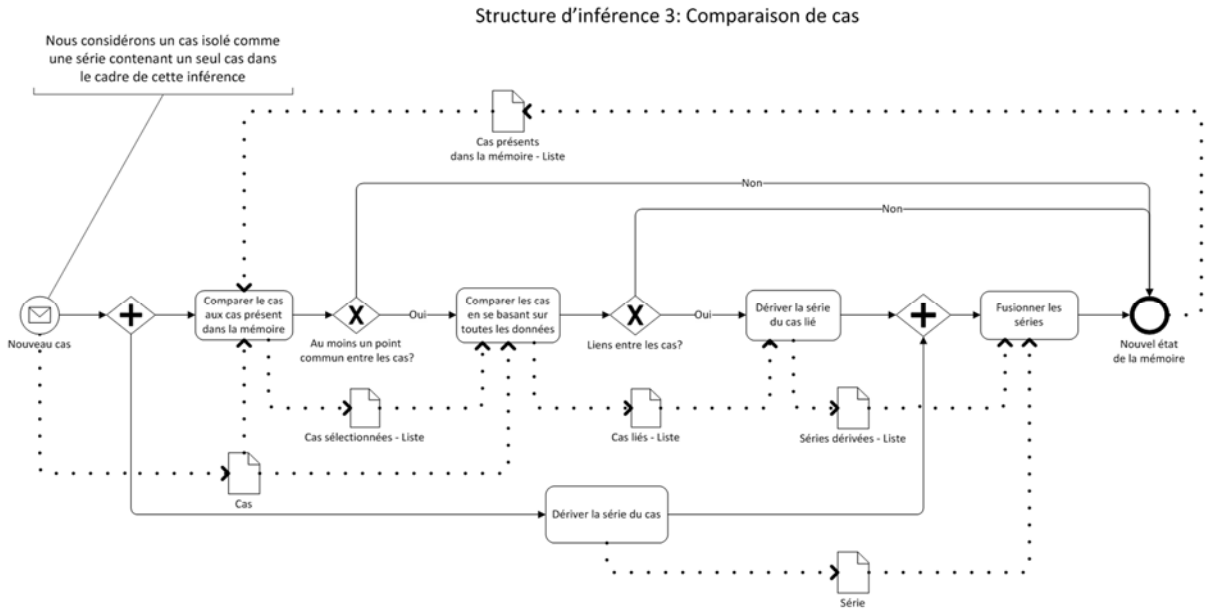
Structure d'inférence 1: Classification à travers le profilage



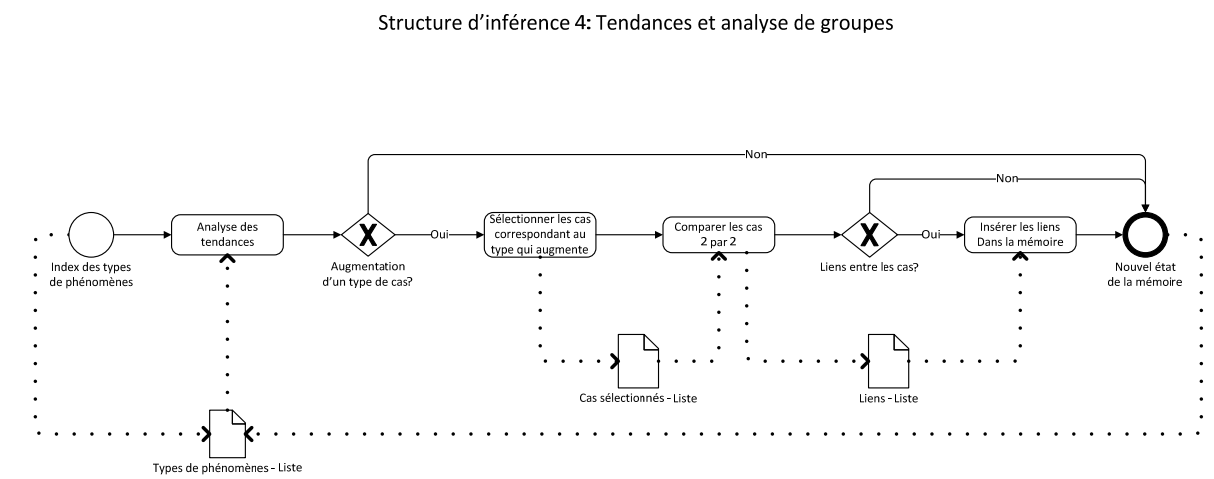
Annexe 11 *: Structure d'inférence 2 : comparaison de séries.*



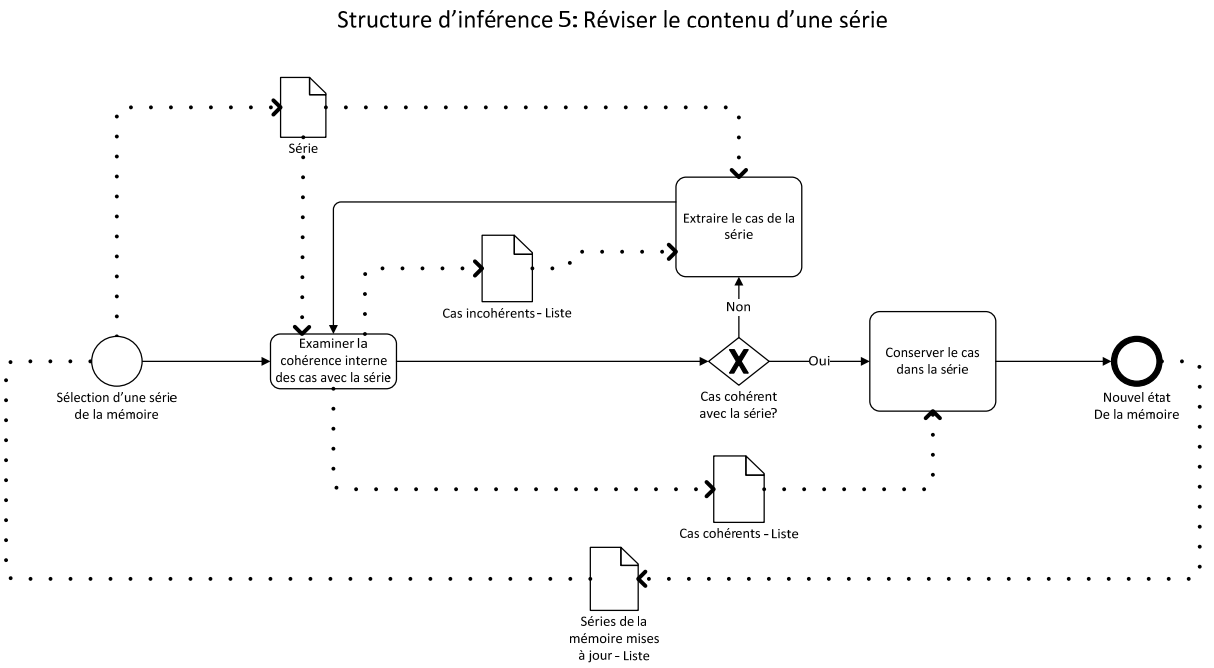
Annexe 12 *: Structure d'inférence 3 : comparaison de cas.*



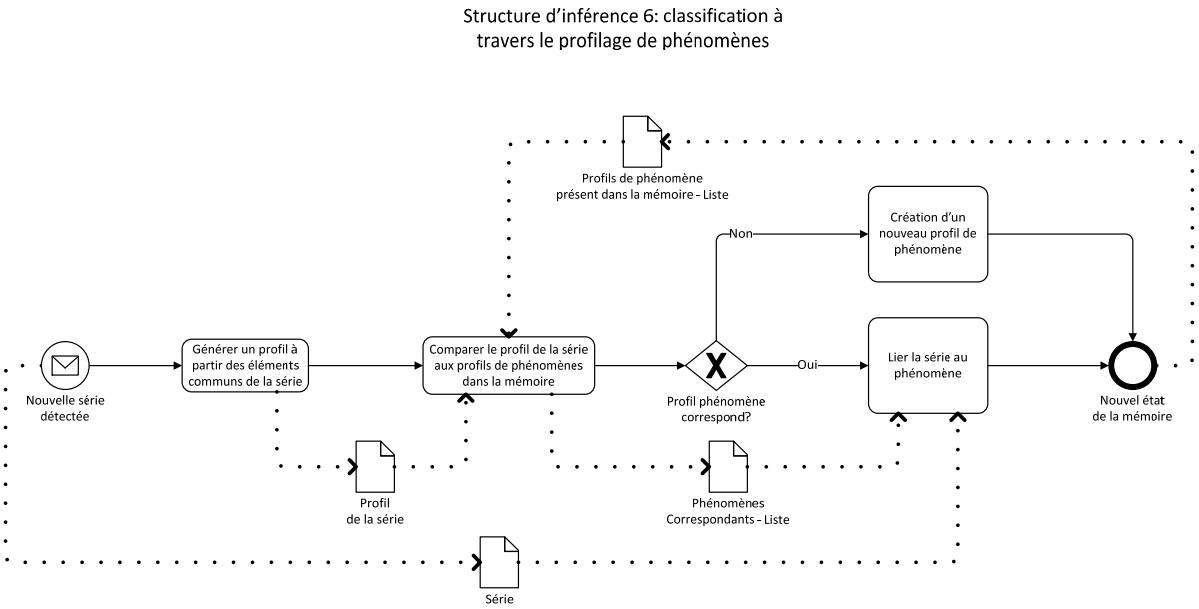
***Annexe 13** : Structure d'inférence 4 : tendances et analyse de groupes.*



***Annexe 14** : Structure d'inférence 5 : révision du contenu d'une série.*



Annexe 15 *: Structure d'inférence 6 : classification à travers le profilage de phénomènes.*



Annexes D: Codifications dans PICAR - tableaux complémentaires

Annexe 16 *: les types d'événements enregistrés dans la base de données PICAR.*

Type d'événements dans PICAR
Agression - Bagarre - Racket
Brigandage/Hold-Up
Cambriolage
Découverte plaque/objets/coffre
Découverte véhicule
Dommages à la propriété
Effraction compteur/automate
Effraction vestiaire/casier
Escroquerie
Fausse monnaie
Filouterie d'auberge
Incendie - Explosion
Individu/véhicule suspect
Interpellation - Arrestation
Meurtre
Mœurs
Renseignement/info
Utilisation frauduleuse d'un ordinateur
Vol à l'arraché
Vol à l'astuce
Vol à l'étalage
Vol à la tire
Vol bourse sommelière
Vol dans un véhicule
Vol de plaques
Vol de véhicules
Vol simple

Annexe 17 *: Codes CICOP pour les cambriolages d'habitation dans la base de données PICAR.*

Code CICOP pour les cambriolages d'habitation	
GIORNO	Cambriolages diurnes d'habitations sans détail
GIORNO CILINDRO	Cambriolages diurnes d'habitations par arrachage du cylindre de la porte palière
GIORNO EPAULEE	Cambriolages diurnes d'habitations par effraction porte à coups d'épaule, de pieds ou de vive force
GIORNO FINESTRA	Cambriolages diurnes d'habitations par effraction ou introduction clandestine fenêtre ou porte-fenêtre
GIORNO PIATTO	Cambriolages diurnes par effraction porte palière avec outil plat genre tournevis
GIORNO CHIAVE	Cambriolage d'habitations la journée, où la clé est trouvée, par exemple dans la boîte aux lettres
HALL	Cambriolages diurnes dans les habitations (en général introduction par porte non verrouillée) suivi du vol de numéraire ou autre, souvent dans le hall d'entrée
NOTTE	Cambriolages nocturnes d'habitations durant le sommeil des lésés
NOTTE CHIGNOLE	Cambriolages nocturnes d'habitations durant le sommeil des lésés par percement montant fenêtre ou porte-fenêtre
NOTTE CILINDRO	Cambriolages nocturnes d'habitations durant le sommeil des lésés par arrachage du cylindre
NOTTE FINESTRA	Cambriolages nocturnes d'habitations durant le sommeil des lésés par effraction fenêtre / porte-fenêtre ou par fenêtre / porte-fenêtre entrouverte
SERA	Cambriolages d'habitations le soir (tombée nuit)
SERA BLOKO	Cambriolages d'habitations le soir (tombée nuit) en bloquant la porte d'entrée

Annexe 18 *: Type de lieu détaillé pour les habitations dans PICAR.*

Type de lieu détaillé pour les habitations dans PICAR (2012)
Appartement
Autre habitation
Cave/buanderie/grenier
Chalet
Corridor/escalier/ascenseur
Ferme
Villa

Annexe 19 *: Voie d'entrée dans PICAR pour les cambriolages d'habitation.*

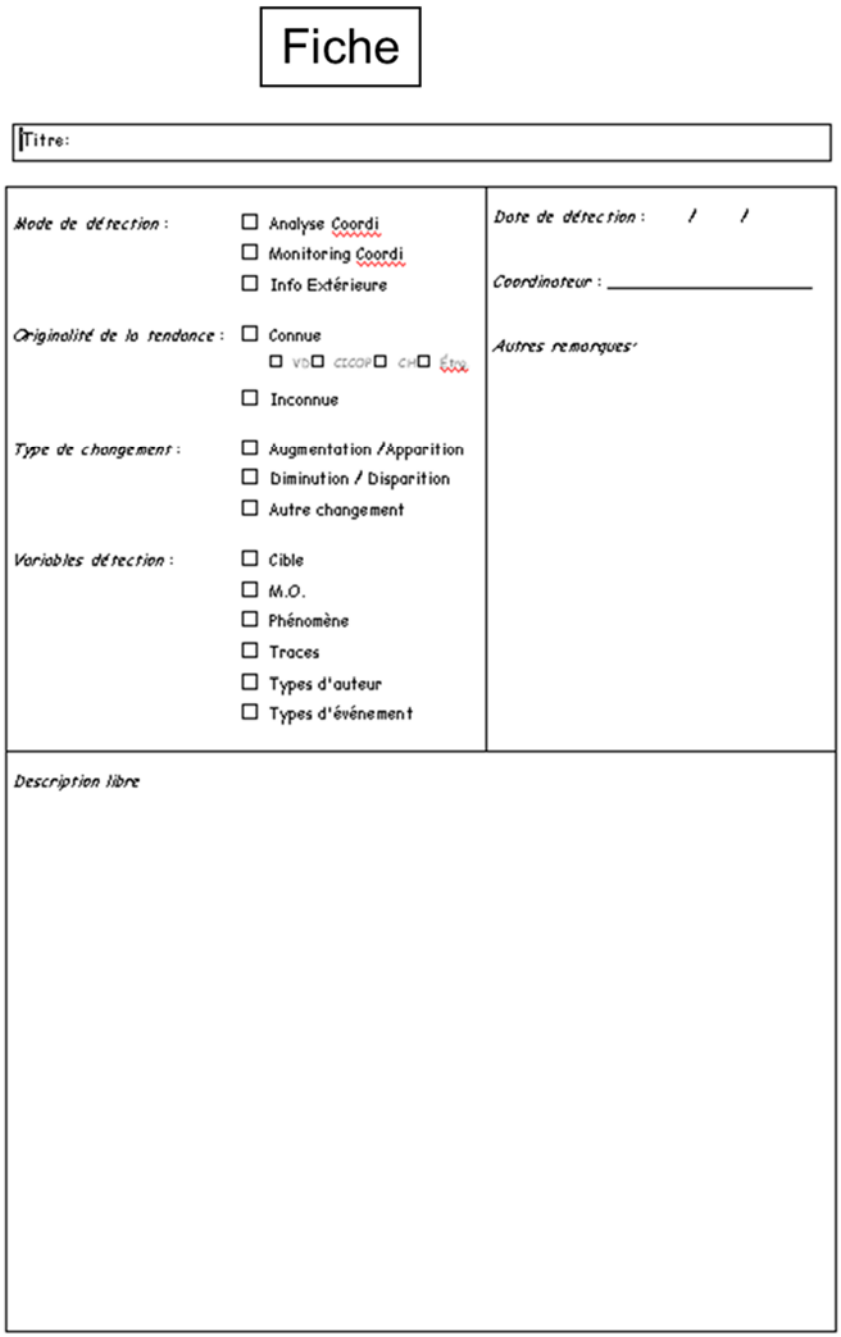
Voie d'entrée dans PICAR pour les cambriolages d'habitation
Fenêtre/porte-fenêtre
Grille/saut-de-loup
Imposte/vasistas
Porte
Toit

***Annexe 20** : Mode opératoire recensé pour les cambriolages d'habitation dans PICAR.*

Mode opératoire recensé pour les cambriolages d'habitation dans PICAR
Arme à feu
Bloquer porte d'entrée
Boîte aux lettres
Bris de vitre
Chatière
Chignole
Clé volée/trouvée/fabriquée
Coffre-fort emporté
Coffre-fort forcé
Coffre-fort meulé
Coup de pied
Cylindre arraché
Cylindre arraché- emporté
Cylindre arraché- remis
Escalade
Introduction clandestine
Jet de pierre/pavé
Judas masqué/enlevé
Neutraliser l'alarme
Outil
Store/volet soulevé/forcé
Utilisation frauduleuse d'un ordinateur
Vive force

Annexes E : Collecte de tendances - Documents complémentaires

Annexe 21 *: Modèle de fiche destinée aux analystes de la coordination judiciaire de la police cantonale vaudoise: Les analystes remplissent cette fiche et la remettent au coordinateur désigné qui remplit la base de données.*



Annexes F: Détection de tendances - Résultats complémentaires

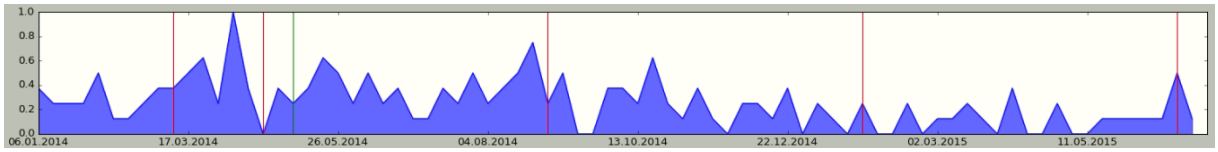
***Annexe 22** : Description des tendances détectées par l'unité d'analyse du CICOP (n= 27). En italique et en gras figurent les tendances non retenues dans l'analyse.*

ID	Nom	Date de détection	Type d'événements	Codes CICOP
7	Cambriolages de fermes avec vols d'outillage	06/05/2014	Cambriolage	
8	Vols à l'astuce - Bijoux de pacotille	06/05/2014	Vol à l'astuce	BIJOUX PACOTILLE
9	SACOM - FALSO POLIZIA	12/05/2014	Vol à l'astuce	FALSO POLIZIA SACOM
10	Vol à l'astuce dans les établissements publics	16/05/2014	Vol à l'astuce	
11	Vol à l'arraché de colliers lors de festivals	02/06/2014	Vol à l'arraché	NORDAFRICA
12	Recrudescence effraction automate à billets	18/06/2014	Effraction compteur/automate	AUTOMATE
13	***Arrivée d'auteurs ukrainiens***	***03/07/2014***	***Vol à l'étalage***	***RUSSIA***
14	***Vol à l'astuce en rue visant les gens d'origine portugaise***	***24/06/2014***	***Vol à l'astuce***	***IBERICA***
15	Effraction automate TL	11/08/2014	Effraction compteur/automate	AUTOMATE
16	SERA tardifs	11/08/2014	Cambriolage	NORDAFRICA
17	Cambriolages dans les hôtels	11/08/2014	Cambriolage	RATS D'HOTEL
19	Vols dans les parkings de centres commerciaux - SACOM	29/08/2014	Vol à l'astuce	SACOM
21	MONETA - OBSERVO	01/09/2014	Vol à l'astuce	MONETA OBSERVO
22	ENKELTRICK	16/10/2014	Vol à l'astuce	ENKELTRICK GITANA
23	BIGLIETTO - Vol au rendez-moi	14/11/2014	Vol à l'astuce	BIGLIETTO ROMANIA
24	Baisse momentanée des GIORNO CILINDRO	21/11/2014	Cambriolage	GIORNO CILINDRO
25	FALSO - Faux plombiers - contrôle radiateur/courant d'eau	14/02/2015	Vol à l'astuce	FALSO
26	BIJOUX PACOTILLE - ROMANIA. Escroqueries aux faux bijoux visant les automobilistes sur les grands axes routiers.	24/03/2015	Escroquerie	BIJOUX PACOTILLE ROMANIA
27	BANCOMAT - NORDAFRICA. Vol à l'astuce lors de retraits au bancomat.	28/03/2015	Vol à l'astuce	BANCOMAT NORDAFRICA
28	BIGLIETTO - Vol au rendez-moi	27/01/2015	Vol à l'astuce	BIGLIETTO GITANA ROMANIA
29	***ZIGANA - Contrôles, interpellations, mises en cause pour cambriolage***	***16/01/2015***	***Cambriolage***	***GIORNO FINESTRA GIORNO PIATTO ZIGANA***
30	Cambriolages de cabinet vétérinaire.	27/02/2015	Cambriolage	
31	COCO - Cambriolages avec vols de cigarettes	06/02/2015	Cambriolage	COCO
32	Cambriolages à la Vallée de Joux - Habitations et commerces	03/02/2015	Cambriolage	ROMANIA ZIGANA
34	Vol à l'astuce faux plombier	27/02/2015	Vol à l'astuce	FALSO
35	SACOM - FALSO POLIZIA	14/04/2015	Vol à l'astuce	FALSO POLIZIA SACOM
36	Vols dans les véhicules SACO, suivi de retraits frauduleux	04/07/2014	Vol dans un véhicule	SACO

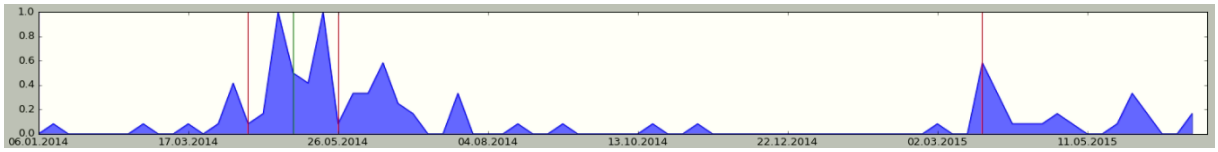
***Annexe 23** : Variables filtres utilisées pour constituer les jeux de données correspondant à chaque tendance détectée.*

ID	Date de détection	Étendue géographique	Type de lieux général	Type de lieux détail	Type d'événement	Code phénomène	Mode opératoire
7	06.05.14			Ferme	Cambriolage		
8	06.05.14					BIJOUX PACOTILLE	
9	12.05.14					FALSO POLIZIA	
10	16.05.14	VD	Etabliss publics		Vol à l'astuce		
11	02.06.14	District de la Riviera-Pays-d'Enhaut			Vol à l'arrachée		
12	18.06.14	District de Lausanne				AUTOMATE	
15	11.08.14	District de Lausanne				AUTOMATE	
16	11.08.14	District de Nyon, Morges, Ouest lausannois				SERA	Escalade
17	11.08.14					RATS D'HOTEL	
19	29.08.14					SACOM	
21	01.09.14					MONETA	
22	16.10.14					ENKELTRICK	
23	14.11.14					BIGLETTO	
24	21.11.14	VD				GIORNO CILINDRO	
25	14.02.15					FALSO	Faux plombier
26	24.03.15					BIJOUX PACOTILLE	
27	28.03.15					BANCOMAT	
28	27.01.15					BIGLETTO	
30	27.02.15			Cabinet médical			
31	06.02.15					COCO	
32	03.02.15	District du Jura Nord-vaudois			Cambriolage		
34	27.02.15					FALSO	
35	14.04.15					FALSO POLIZIA	
36	04.07.14	VD				SACO	

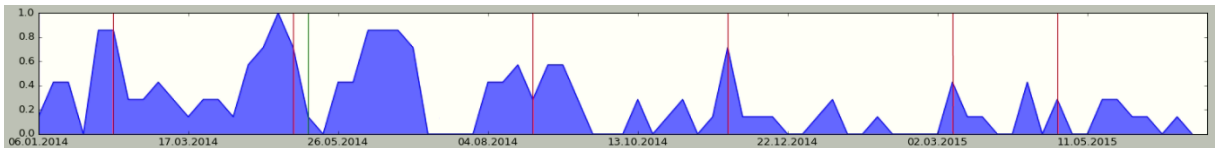
***Annexe 24** : Résultats de l'analyse de changement de points sur les tendances détectées. Les traits rouges correspondent aux détections automatiques de l'algorithme, le trait vert à la détection humaine des analystes et le trait jaune indique une correspondance parfaite entre la détection automatique et humaine. Les données sont agrégées en semaines entre janvier 2014 et juin 2015. (Paramètres utilisés : SEG=1 ; SSS=3 ; DPU=0.09 ; K=6).*



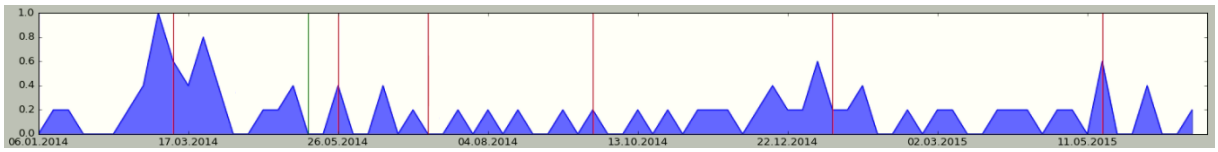
Tendance 07 (n=160)



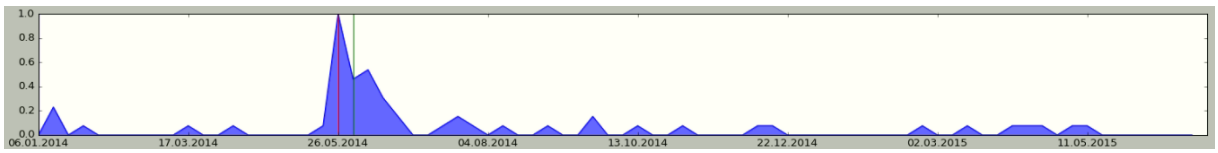
Tendance 08 (n=103)



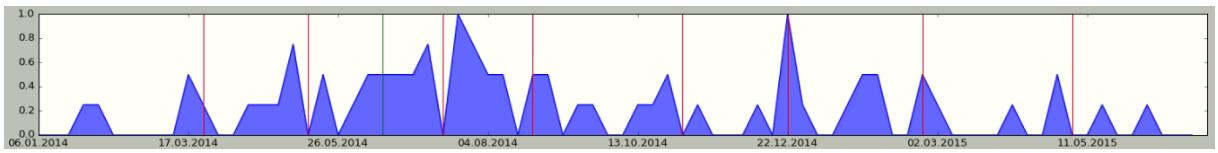
Tendance 09 (n=142)



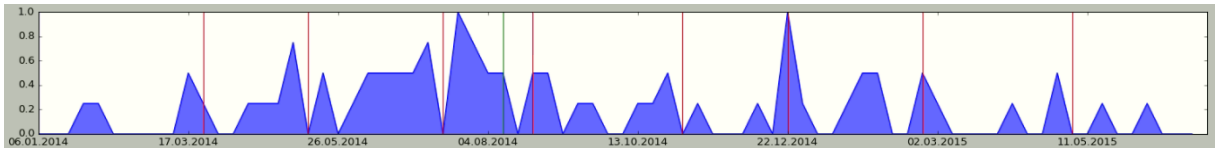
Tendance 10 (n=66)



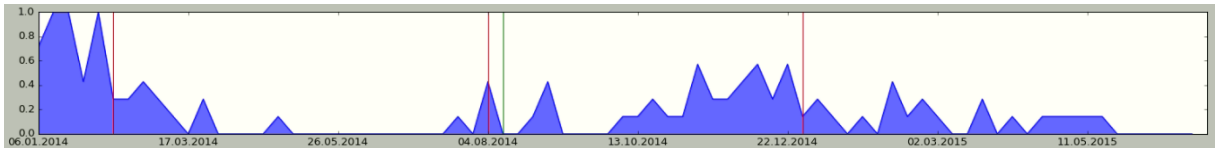
Tendance 11 (n=58)



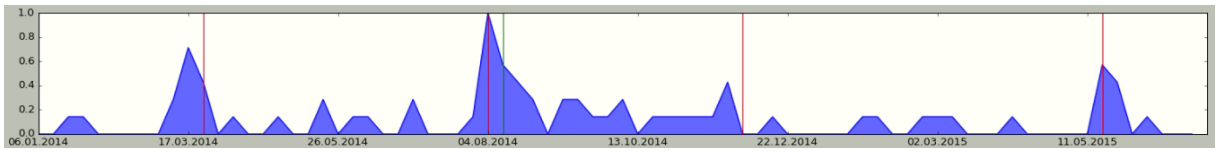
Tendance 12 (n=66)



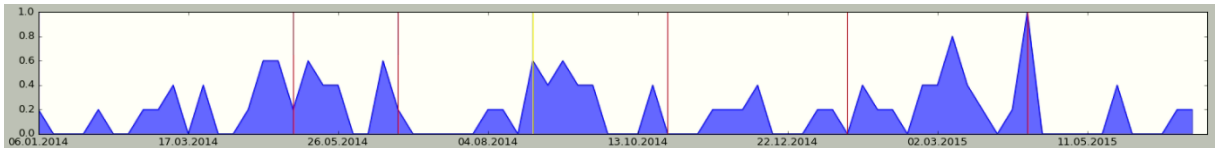
Tendance 15 (n=66)



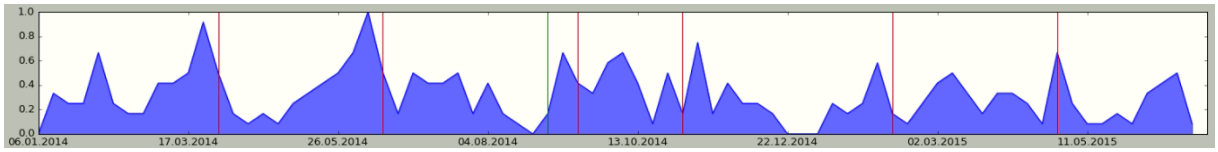
Tendance 16 (n=97)



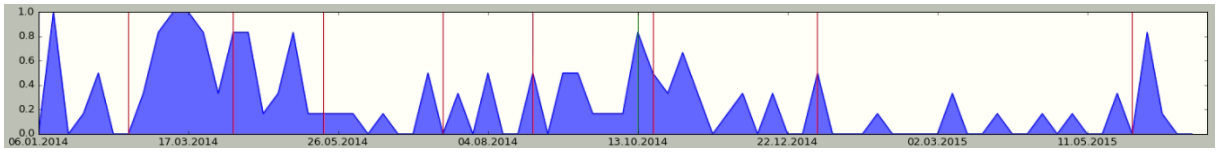
Tendance 17 (n=68)



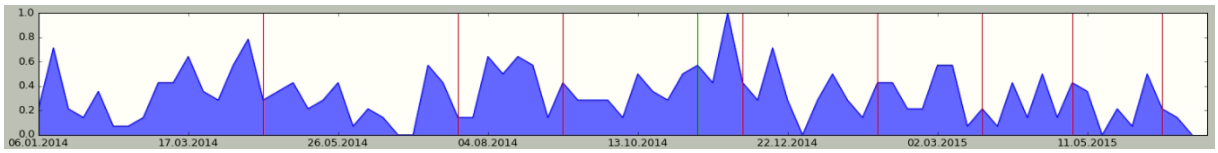
Tendance 19 (n=75)



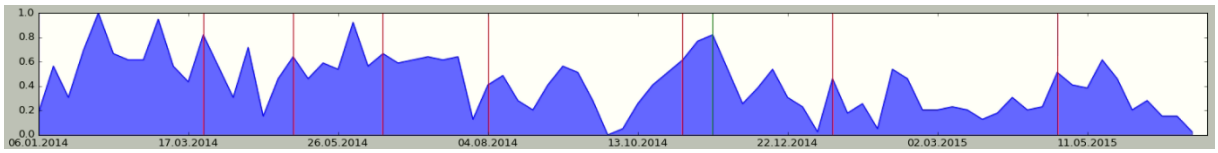
Tendance 21 (n=296)



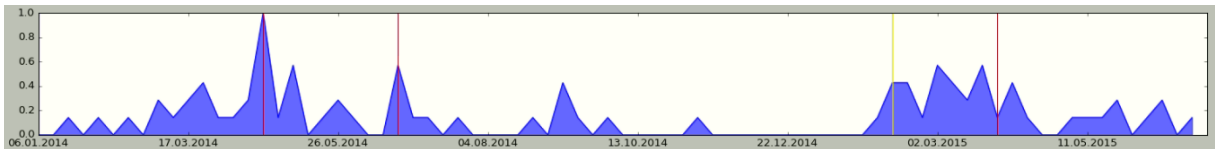
Tendance 22 (n=117)



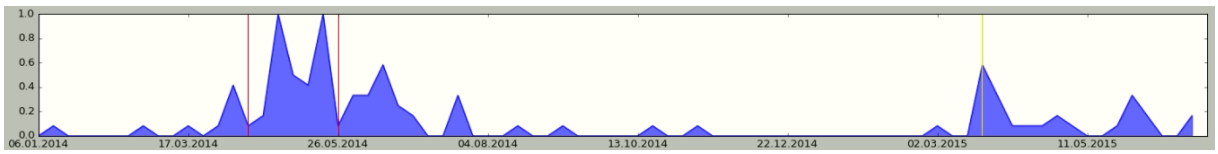
Tendance 23 (n=358)



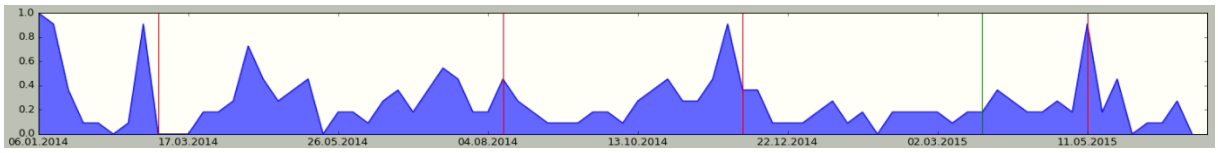
Tendance 24 (n=1449)



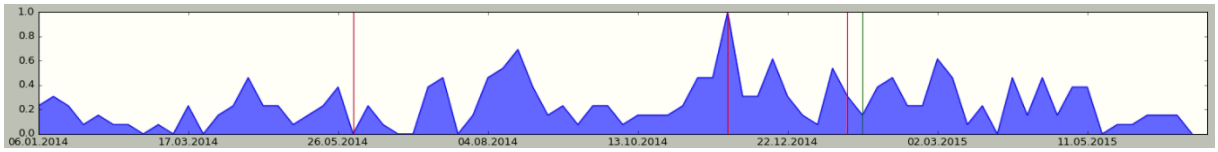
Tendance 25 (n=80)



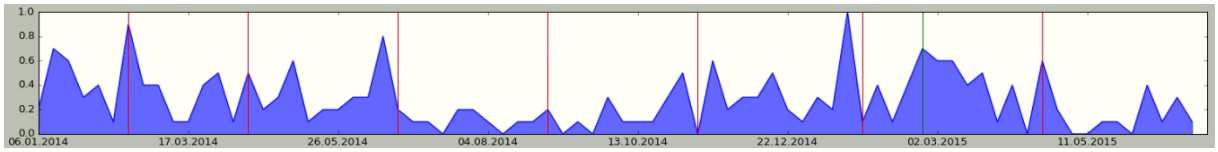
Tendance 26 (n=103)



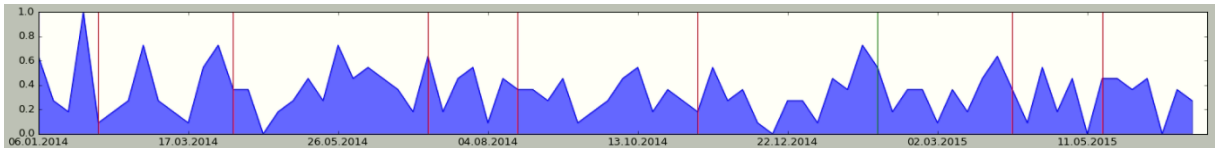
Tendance 27 (n=224)



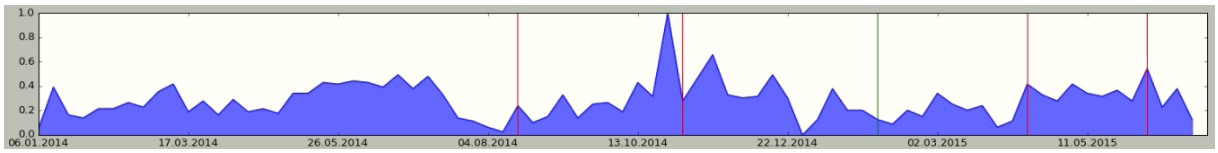
Tendance 28 (n=243)



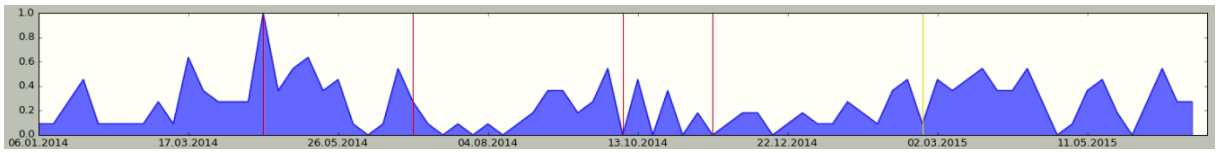
Tendance 30 (n=2310)



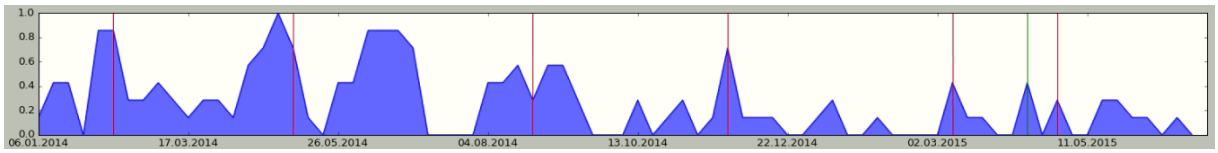
Tendance 31 (n=292)



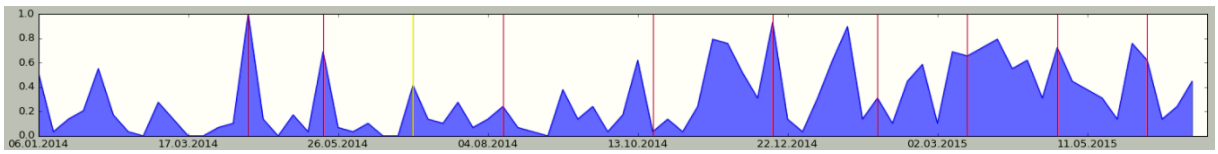
Tendance 32 (n=295)



Tendance 34 (n=3147)



Tendance 35 (n=212)



Tendance 36 (n=142)

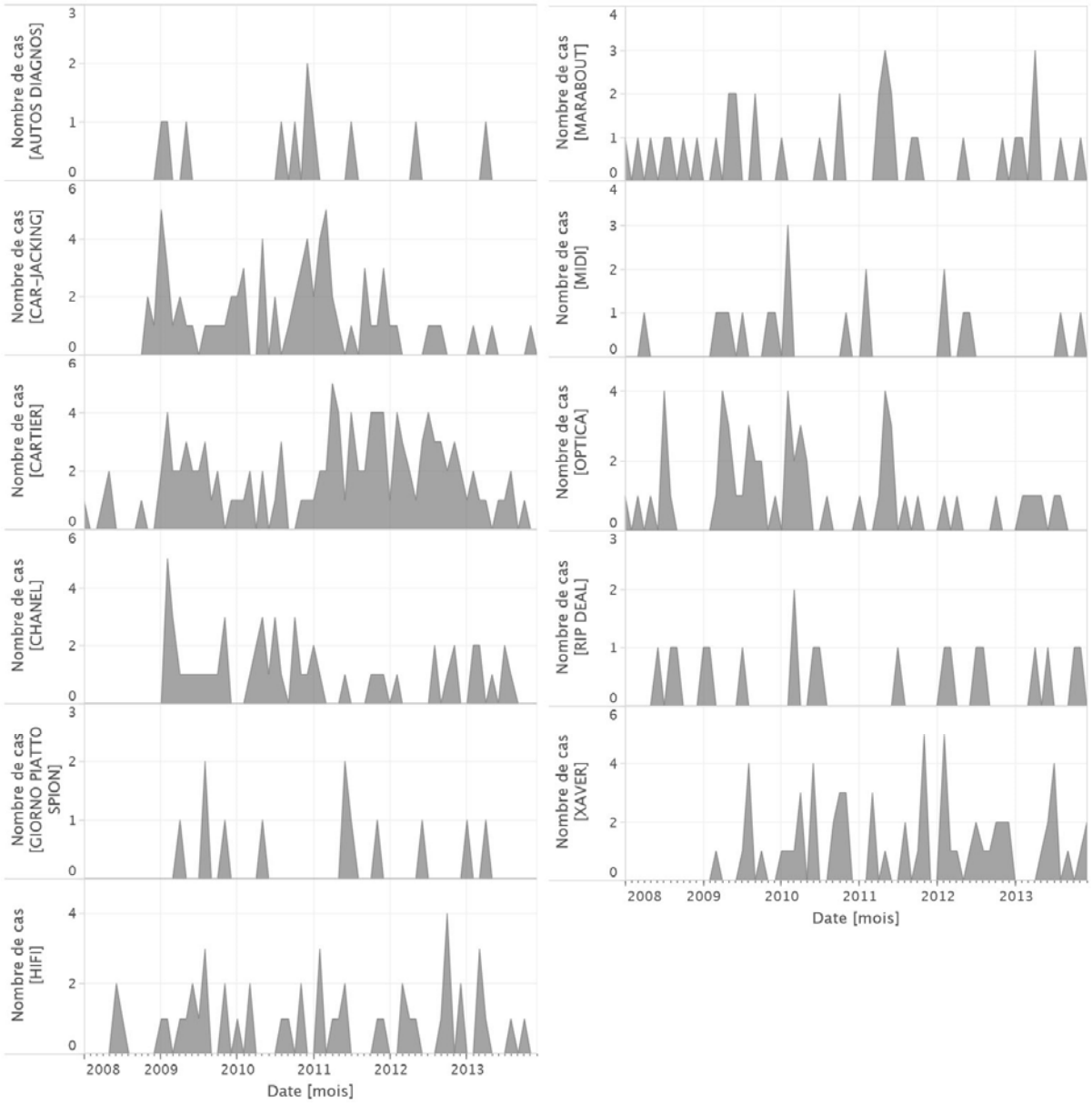
Annexes G: Paramétrisation- Résultats complémentaires

***Annexe 25** : Description des codes phénomènes enregistrés dans PICAR entre 2009 et 2013. Les types d'auteurs ne sont pas inclus.*

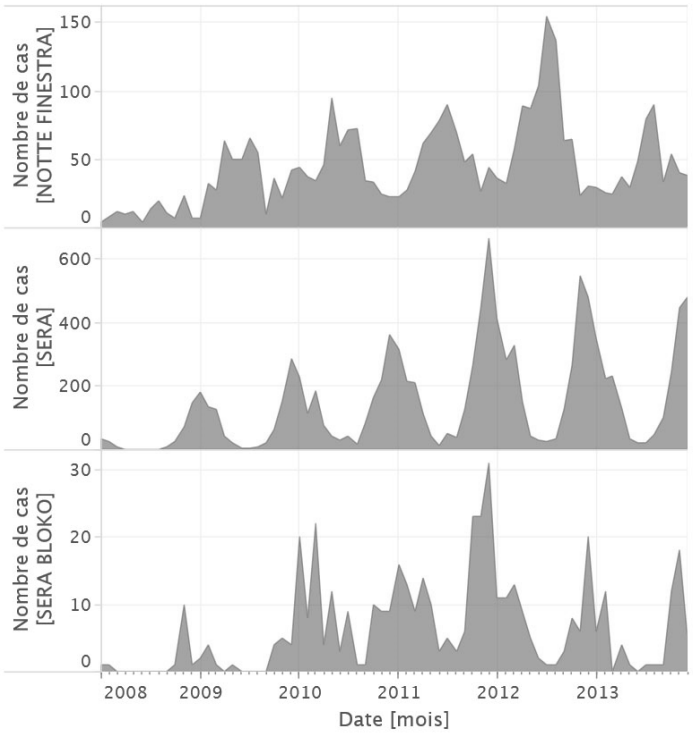
Code phénomène	***Description***
AIUTO	Vols à l'astuce, les auteurs aident la victime à monter dans le train et lui subtilisent son argent
ALU	Vol de métaux
ASTUCE COMMERCES	Vols à l'astuce dans les commerces
ASTUCE HABITATIONS	Vols à l'astuce en général dans les habitations, souvent au préjudice des personnes relativement âgées et qui ont souvent été suivies depuis un magasin ou la poste
ASTUCE PNEU	Vol à l'astuce en crevant le pneu de la voiture afin de dérober le sac à main dans l'habitacle
ASTUCE RUE	Vols astuce dans la rue
ASTUCE TRAIN	Vol à l'astuce dans les trains, sauf cas AIUTO où de l'aide est proposée
ASTUCE VHC	Vols astuce dans la rue en lien avec la voiture du lésé (sauf pneu crevé)
AUTOMATE	Cambriolage d'appareils à prépaiement (lavage de station-service, etc..)
AUTOS	Vol de véhicule de valeur (neuf), souvent dans les garages
AUTOS DIAGNOS	Vol des appareils de diagnostic dans les garages
BANCOMAT	Vols à l'astuce en lien avec des distributeurs automatiques de billets
BATELO	Vols en série de moteurs de bateau
BIGLIETTO	Vol au rendez-moi avec grosse coupure
BIJOUX PACOTILLE	Vente de bijoux de pacotille, parfois avec usures (faire croire que or)
CADDIE	Vols à la tire ou vols simples au préjudice de clients(es) des grands magasins qui mettent leur sac dans le caddie
CARBURO	Tout type de vol de carburant et autres combustibles (siphonnage, par effraction, etc.)
CAR-JACKING	Vol de véhicule avec violence
CARTIER	Vols à l'astuce dans les bijouteries
CHANEL	Cambriolage avec vols de parfums en grande quantité
COCO	Cambriolages de magasins où on vol des grandes quantités de cigarettes et de spiritueux
ENKELTRICK	Astuce auprès de personnes âgées en se faisant passer pour une connaissance, demande des sommes d'argent importantes
FALSO	Vols à l'astuce en général dans les habitations, souvent au préjudice des personnes relativement âgées par des malfaiteurs se présentant sous de fausses qualités (plombiers, etc.)
FALSO POLIZIA	Vols à l'astuce en général dans les habitations, souvent au préjudice des personnes relativement âgées avec intervention d'un faux policier.
FURTIF	Vol par introduction clandestine dans des entreprises, commerces, bureaux, ..., pendant les heures d'ouverture. Emporte le porte-monnaie, le sac à main, l'ordinateur portable, également considéré comme vols simples.
GIORNO	Cambriolages diurnes d'habitations sans détail
GIORNO CHIAVE	Cambriolage d'habitations la journée, où la clé est trouvée, par exemple dans la boîte aux lettres
GIORNO CILINDRO	Cambriolages diurnes d'habitations par arrachage du cylindre de la porte palière
GIORNO EPAULEE	Cambriolages diurnes d'habitations par effraction porte à coups d'épaule, de pieds ou de vive force

GIORNO FINESTRA	Cambriolages diurnes d'habitations par effraction ou introduction clandestine fenêtre ou porte-fenêtre
GIORNO PIATTO	Cambriolages diurnes par effraction porte palière avec outil plat genre tournevis
GIORNO PIATTO SPION	Cambriolages diurnes par effraction porte palière avec outil plat genre tournevis, masquer le judas des voisins
HALL	Cambriolages diurnes dans les habitations (en général introduction par porte non verrouillée) suivi du vol de numéraire ou autre, souvent dans le hall d'entrée
HIFI	Cambriolages de commerces spécialisés dans la vente en matériel radio TV vidéo, vente de téléphones portables, etc.
HOLD-UP	Attaques à main armée dans banques, postes, transport de fonds
LASSO	Arraché bancomat
MARABOUT	Escroqueries et vols par des pseudo mages ou sorciers africains
MIDI	Cambriolages de magasins entre 1200 et 1400, pour voler essentiellement de l'argent
MODA	Cambriolages de magasins ou boutiques de mode avec vol important de vêtements
MŒURS ADULTE	Tout ce qui touche à ce domaine au préjudice des adultes
MŒURS ENFANTS	Tout ce qui touche à ce domaine au détriment des enfants plus pédophilie
MONETA	Vol à l'astuce en demandant de la monnaie sur quelques francs
NOTTE	Cambriolages nocturnes d'habitations durant le sommeil des lésés
NOTTE CHIGNOLE	Cambriolages nocturnes d'habitations durant le sommeil des lésés par percement montant fenêtre ou porte-fenêtre
NOTTE CILINDRO	Cambriolages nocturnes d'habitations durant le sommeil des lésés par arrachage du cylindre
NOTTE FINESTRA	Cambriolages nocturnes d'habitations durant le sommeil des lésés par effraction fenêtre / porte-fenêtre ou par fenêtre / porte-fenêtre entrouverte
OBSERVO	Vols à l'astuce à la sortie des banques ou du distributeur au préjudice de clients venant de retirer de l'argent
OMEGA	Cambriolages de bijouteries (souvent vitrines) par coups de masse ou d'objets lourds, ou le coup du bélier
OPTICA	Cambriolages de magasins d'optique avec vol important de lunettes
RADIO	Effractions voitures pour voler autoradio et accessoires, appareils électroniques
RAPINA BIJOUX	Braquages à main armée dans des bijouteries
RATS D'HOTEL	Vol dans des hôtels, au préjudice de clients.
RIP DEAL	Escroqueries ou vols à l'astuce par des malfaiteurs de type gitan YU qui appâtent les victimes avec des taux de change très intéressants pour diverses transactions.
SACO	Effractions voitures pour dérober sacs à main
SACOM	Vol pendant qu'on range le sac à commission dans la voiture
SERA	Cambriolages d'habitations le soir (tombée nuit)
SERA BLOKO	Cambriolages d'habitations le soir (tombée nuit) en bloquant la porte d'entrée
SKIMMING	Duplication de carte de crédit ou EC
SUBTIL	Vol du porte-monnaie/natel dans les restaurants soit dans poche de la veste ou dans sac à main, généralement par le coup du dos-à-dos
TRESORO	Cambriolages de commerces ou entreprises où on s'attaque au coffre-fort, parfois emporté, etc.
XAVER	Brigandages dans des maisons familiales/de maîtres, essentiellement la nuit, au cours desquels les victimes sont souvent molestées

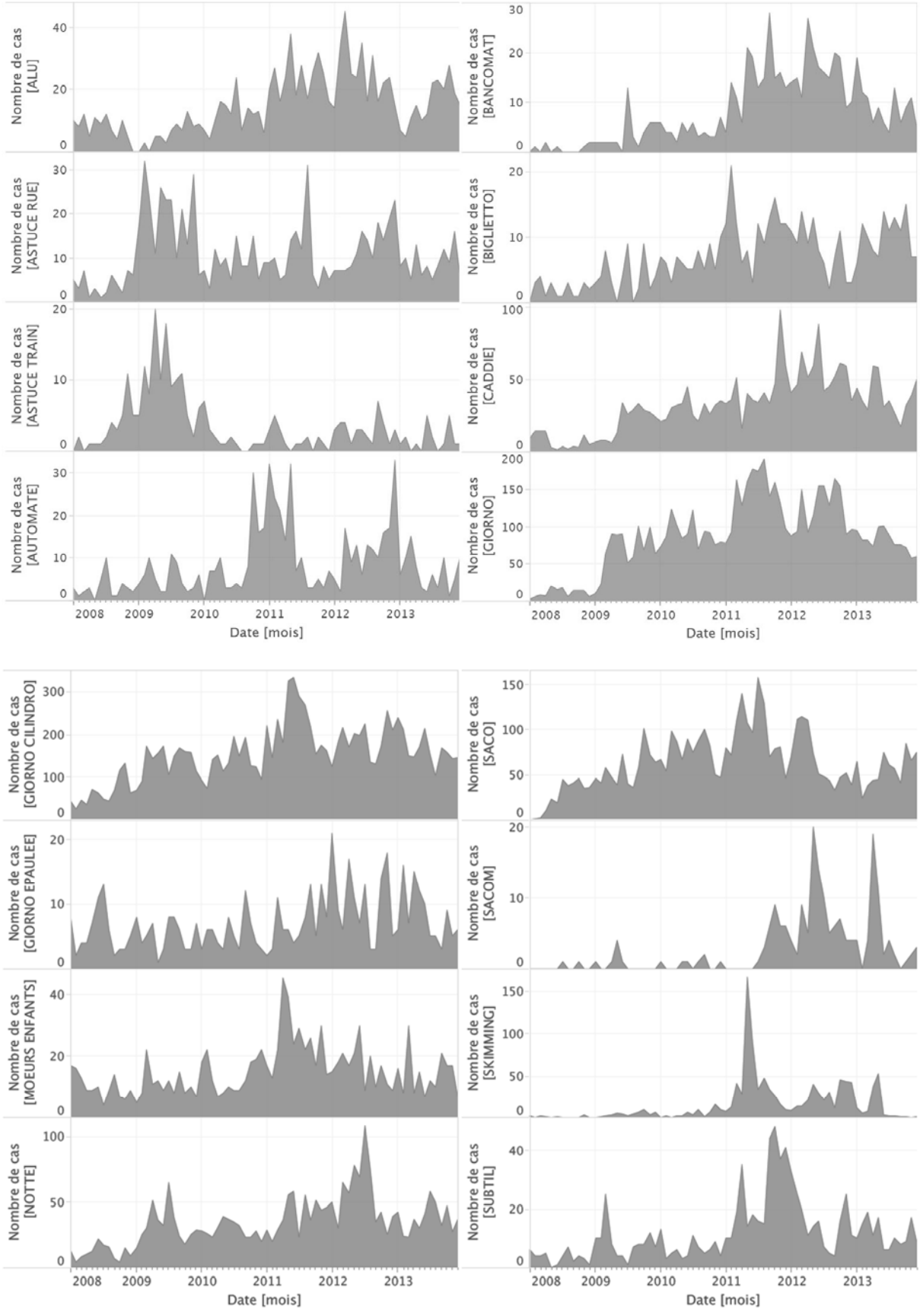
Annexe 26 *: Phénomènes CICOP classifiés dans la catégorie de ruptures de tendances « Peu de cas » (n= 11).*



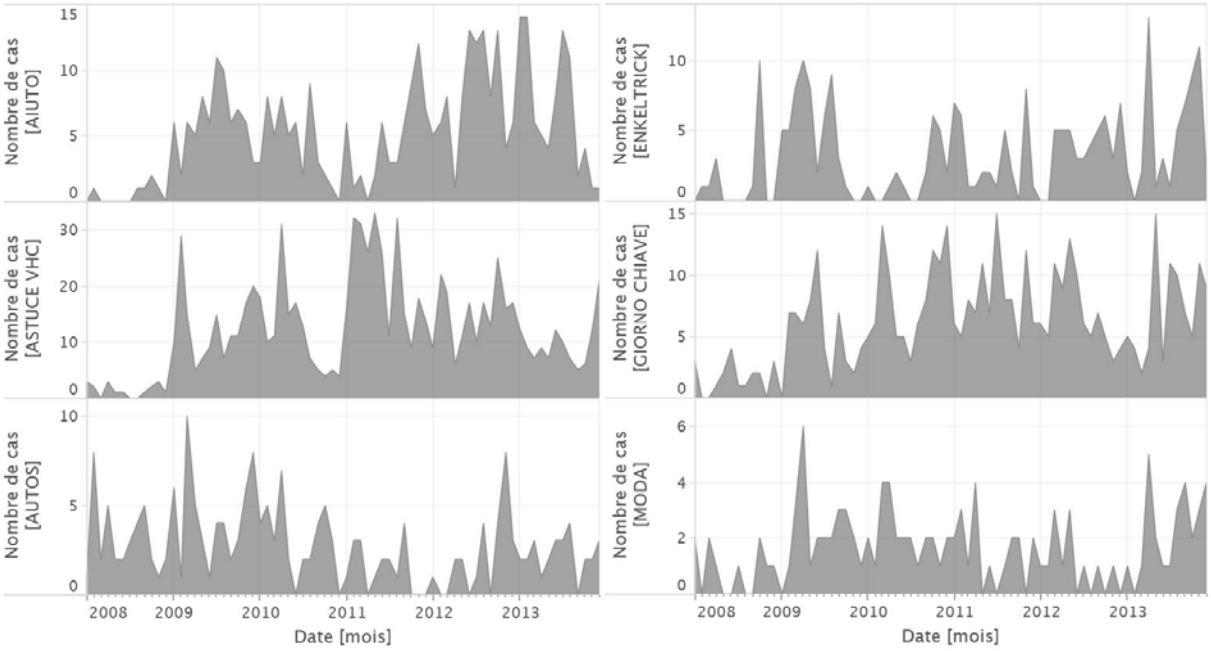
Annexe 27 *: Phénomènes CICOP classifiés dans la catégorie de ruptures de tendances « Pattern de rupture périodique » (n= 3).*



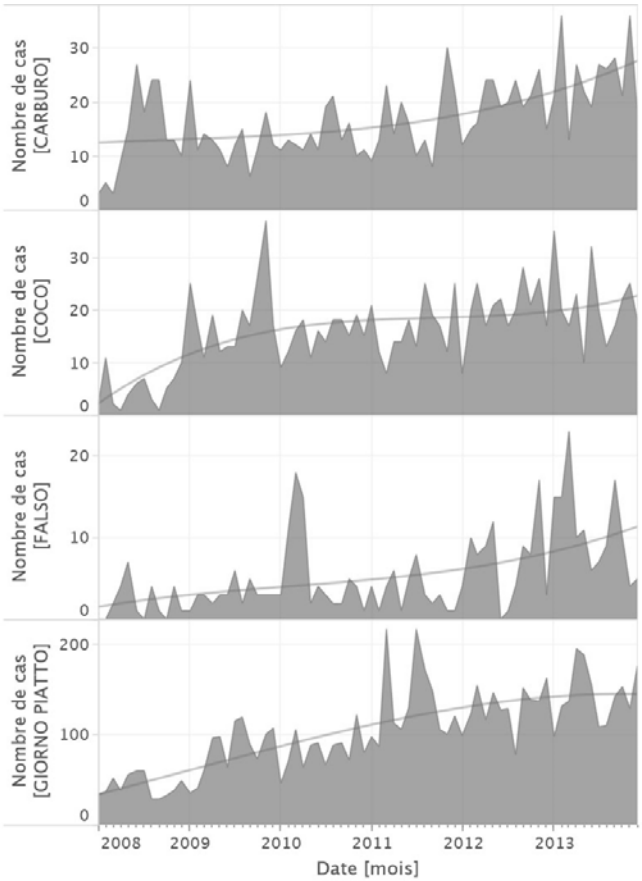
Annexe 28 *: Phénomènes CICOP classifiés dans la catégorie de ruptures de tendances « Pattern de rupture progressive temporaire » (n= 8).*



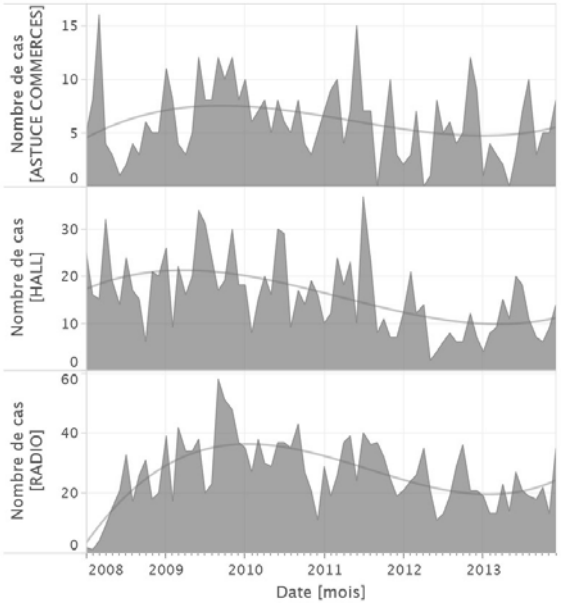
***Annexe 29** : Phénomènes CICOP classifiés dans la catégorie de ruptures de tendances « Pattern de rupture régressive temporaire » (n= 6).*



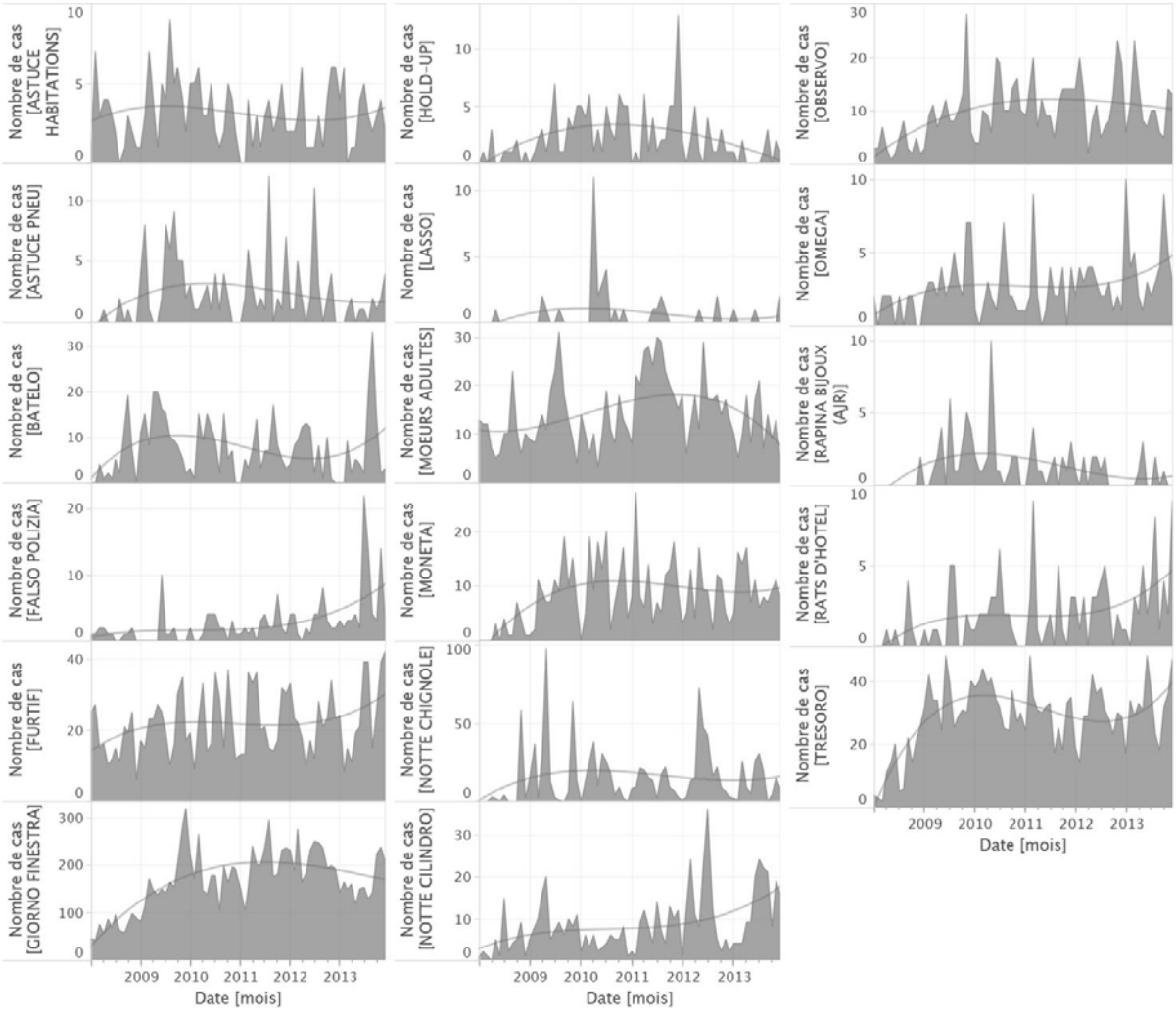
***Annexe 30** : Phénomènes CICOP classifiés dans la catégorie de ruptures de tendances « Pattern de rupture progressive permanente » (n= 4).*



Annexe 31 *: Phénomènes CICOP classifiés dans la catégorie de ruptures de tendances « Pattern de rupture régressive permanente » (n= 3).*



Annexe 32 *: Phénomènes CICOP non classifiés dans une catégorie de ruptures de tendances (n= 17).*



TABLES DES ILLUSTRATIONS

Index des tableaux

Index des graphiques

Index des figures

Index des annexes

ISBN 2-940098-79-4