

## METHODOLOGIE POUR LA DEFINITION DE SCENARIOS HUMANITAIRES REALISTES ET REPRESENTATIFS

---

L'étude approfondie d'un réseau logistique suppose de s'intéresser aux flux, aux profils de demande, aux coûts des structures, au dimensionnement des ressources, à la rotation de stocks, aux objectifs de service, et à beaucoup d'autres facteurs encore. Ces éléments sont particulièrement difficiles à obtenir dans le cadre humanitaire. Le cas de la « demande » est notamment très critique tant l'incertitude y est grande. De prime abord, on peut imaginer qu'il est difficile, voire impossible, de prévoir l'emplacement et l'amplitude de la prochaine catastrophe humanitaire... Et même si l'on pouvait s'en approcher, le contexte serait tellement changeant que les informations deviendraient rapidement erronées ou incomplètes. Pourtant, les acteurs du monde humanitaire ont tout à gagner à pouvoir bénéficier d'outils d'aide à la décision capable de les aider à réaliser leurs missions.

Charles (2010) a démontré qu'une grande majorité des outils disponibles, notamment pour concevoir les réseaux logistiques, sont développées au travers d'approches de type recherche opérationnelle et optimisation. Martel et al. (2013) ont par ailleurs souligné le fait qu'une grande partie de ces propositions étaient déterministes et ne considéraient que trop simplement le caractère incertain de l'humanitaire. Plus récemment, une tendance forte au développement d'approches stochastiques ou floues est apparue. La limite de ces travaux porte souvent sur l'utilisation de données de demande très simplistes qui ne permettent pas de véritablement valider les bénéfices et limites des propositions faites et de réellement donner confiance aux utilisateurs potentiels (Cassidy, 2003 : Murray, 2005 : Charles, 2010 : Martel et al. 2013). C'est pour pallier cette difficulté que plusieurs auteurs ont suggéré de s'intéresser à la réalisation de prévisions de demande en contexte humanitaire et à l'élaboration de scénarios réalistes (Charles, 2010 : Peres et al. 2012 : Braman et al. 2013).

# 1. Système de prévision des catastrophes et de leurs conséquences

Le sujet même de nos travaux de recherche invite à se poser la question de comment anticiper les futures catastrophes et leurs impacts potentiels. Si cette question a peu de sens dans le cadre de catastrophes majeures de type tsunami en Indonésie en 2004, tremblement de terre en Haïti en 2010, tsunami au Japon en 2011 ou typhon aux Philippines en 2013, le sujet prend tout son sens dans le cadre de crises récurrentes du type de celles que nous souhaitons étudier.

Comme les travaux de (Charles, 2010), (Kovács et al., 2007) ou (Peres et al., 2012) l'ont démontré, l'occurrence de catastrophes récurrentes n'est pas si imprévisible que ce que le sens commun peut laisser croire. Sur cette base, et dans le cas exclusif des catastrophes récurrentes, nous émettons l'hypothèse qu'un système de prévisions doit pouvoir être établi. L'objectif étant de pouvoir disposer d'une estimation réaliste de la future « demande » à laquelle les chaînes logistiques humanitaires devront répondre.

Pour alimenter cette réflexion, nous proposons dans cette section un rapide tour d'horizon du sujet de l'estimation de la demande d'abord par le prisme traditionnel de l'industrie, puis par celui de la gestion de crises. Dans un deuxième temps, nous proposerons une méthodologie instrumentée susceptible de concrétiser l'établissement de prévisions de besoin pour le cas des catastrophes récurrentes.

## 1.1. Revue de littérature

### 1.1.1. *Les prévisions dans le monde commercial*

La plupart des entreprises doivent planifier leurs activités en anticipant les besoins de leurs clients. En particulier, les entreprises qui assemblent à la commande ou celles qui livrent sur stock déclenchent leurs approvisionnements et leurs productions sur du prévisionnel. Pour y arriver, elles s'appuient sur des systèmes de planification. Ces systèmes se déclinent le plus souvent sur trois niveaux dits « stratégique », « tactique » et « opérationnel ».

Le niveau stratégique a pour objectif de positionner l'entreprise sur le marché dans son ensemble (dans le cadre de sa stratégie générale). Les étapes essentielles en sont (Bourbonnais et Usunier, 2001 ; Hadj-Hamou, 2004 ; Dupont et Lauras, 2008) :

- Prévoir ce que sera la demande sur le long terme ;
- Considérer les risques et opportunités de l'environnement (concurrence, fournisseurs, évolution des législations et des technologies...) ;
- Déterminer les parts de marché visés.
- Fixer les types d'activité : faire ou faire-faire (*make or buy*) ;
- Choisir le type d'approvisionnement (mono ou multi sources) et sélectionner les fournisseurs ;
- Dessiner le réseau logistique (implantation et dimensionnement des divers sites de production et stockage, réseau de distribution et transport).

Le niveau tactique correspond aux décisions à moyen terme dans le respect des décisions définies par la planification stratégique (Plan Industriel et Commercial). Les étapes essentielles sont (Bourbonnais et Usunier, 2001 : Hadj-Hamou, 2004 : Dupont et Lauras, 2008) :

- Prévoir ce que sera la demande et/ou les ventes sur le moyen terme ;
- Déterminer les plans de production permettant de satisfaire au mieux la demande compte tenu des capacités de production et de sous-traitance, de la disponibilité des ressources matérielles et humaines, des stocks existant ;
- Répartir la production entre les sites (CRP : dimensionnement des capacités de production et des effectifs), affecter les demandes aux centres de distribution et planifier les transports (DRP : plan de distribution) ;
- Déclencher les approvisionnements.

Le niveau opérationnel traite des problèmes d'affectation et de contrôle des flux physiques sur le court terme. Il repose essentiellement sur le Programme Directeur de Production (PDP) qui précise pour chaque produit fini, les quantités à produire, période par période (Bourbonnais et Usunier, 2001 : Hadj-Hamou, 2004 : Dupont et Lauras, 2008).

Comme on peut le constater, la prévision de la demande et/ou du marché est la toute première étape des niveaux stratégiques et tactiques. La prévision est une étape essentielle d'un processus de planification de la chaîne logistique (cf. figure suivante). La performance globale de la chaîne logistique dépend directement de la qualité de ces prévisions. Elles conditionnent dans de nombreux cas la performance, la fiabilité et la productivité générale du système. Les principaux bénéfices à attendre d'une bonne prévision sont :

- La réduction des niveaux de stock de sécurité ;
- Un ajustement des stocks de produits finis sur la politique commerciale voulue (taux de service) ;
- La réduction des risques d'obsolescence des stocks ;
- Une meilleure productivité et réactivité des unités de production ;
- L'optimisation de la chaîne de distribution.

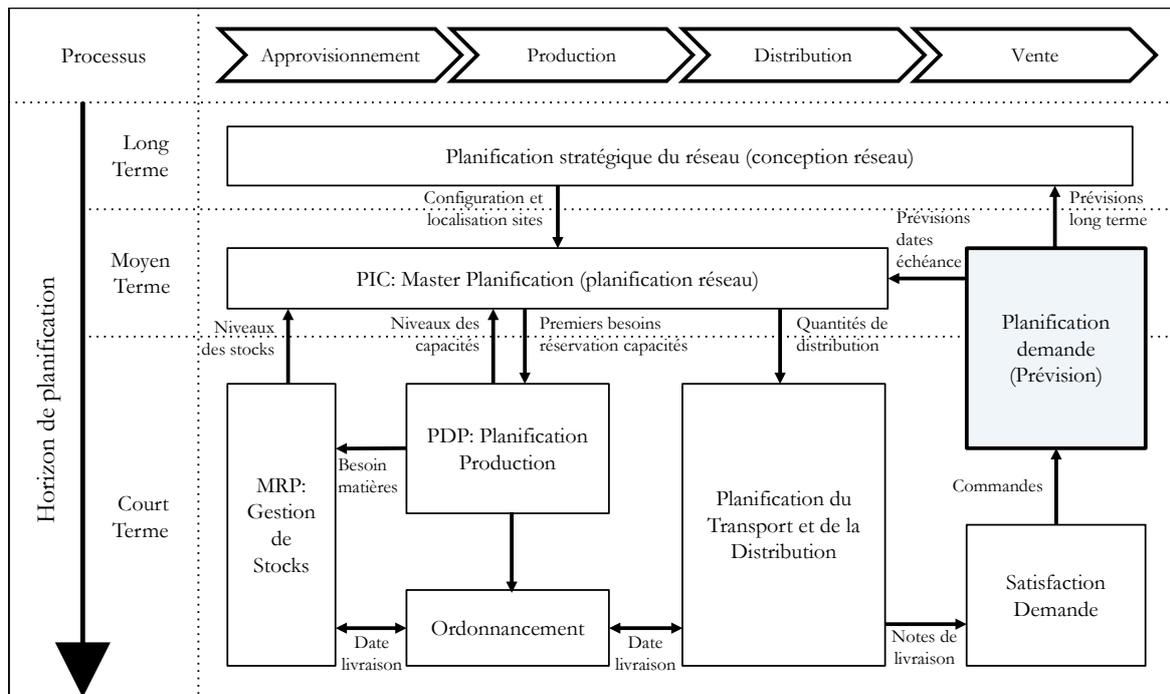


Figure III-1 : La prévision dans le processus de planification des chaînes logistiques.

Source : Hadj-Hamou (2004).

A l'évidence, les prévisions de demande sont utilisées avec des finalités différentes par des acteurs différents de la chaîne logistique. Dès lors lesdits acteurs devront prendre garde à (Bourbonnais et Usunier, 2001 ; Dupont et Lauras, 2008) :

- Ne pas confondre objectif et prévision ;
- Ne pas confondre prévision de demande et prévision de vente ;
- Adapter le niveau de réalisation des prévisions au besoin d'utilisation ;
- Déterminer judicieusement le profil et les compétences des prévisionnistes.

Il existe diverses méthodes pour obtenir des prévisions. Ces méthodes reposent implicitement sur l'idée qu'il est possible de prévoir la demande future du produit étudié en se basant sur un certain référentiel de connaissance. Les méthodes de prévision dépendent de la manière de construire ce référentiel. Pour un produit totalement nouveau ou récent sur le marché, on utilise l'avis d'expert ou les informations recueillies en testant le produit sur un panel restreint d'utilisateurs potentiels. Lorsque le produit (ou des produits similaires) sont déjà présents sur le marché, on utilise la connaissance du passé pour extrapoler le futur. Deux phases sont alors nécessaires pour établir une prévision (Bourbonnais et Usunier, 2001) :

- La modélisation du passé ;
- L'extrapolation.

Ainsi, pour estimer des prévisions de demande, une entreprise doit utiliser d'une part les connaissances qu'elle a des marchés et de leurs évolutions (informations externes) et d'autre part, les connaissances qu'elle a d'elle-même et de son évolution (informations internes). La figure suivante illustre cette démarche.

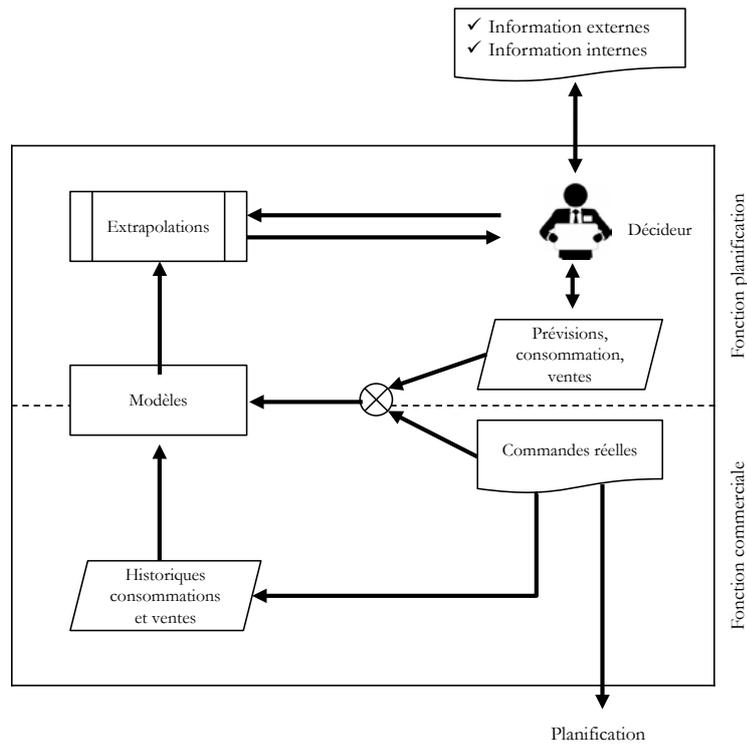


Figure III-2 : De la donnée à la prévision.

D'après Bourbonnais et Usunier (2001).

La figure suivante présente les grandes étapes d'un système de prévision des ventes dans le monde commercial. Avant toute chose, il convient de collecter les données qui serviront à la modélisation du passé. Ces données devront être fiables et en nombre suffisant. Ensuite, vient la méthode de prévision à proprement parlée quelle qu'elle soit. Le traitement est alors informatisé et les données éventuellement mises à jour. La prévision peut alors être définie et validée avant d'être diffusée auprès de l'ensemble des acteurs de la chaîne logistique qui travaillent avec.

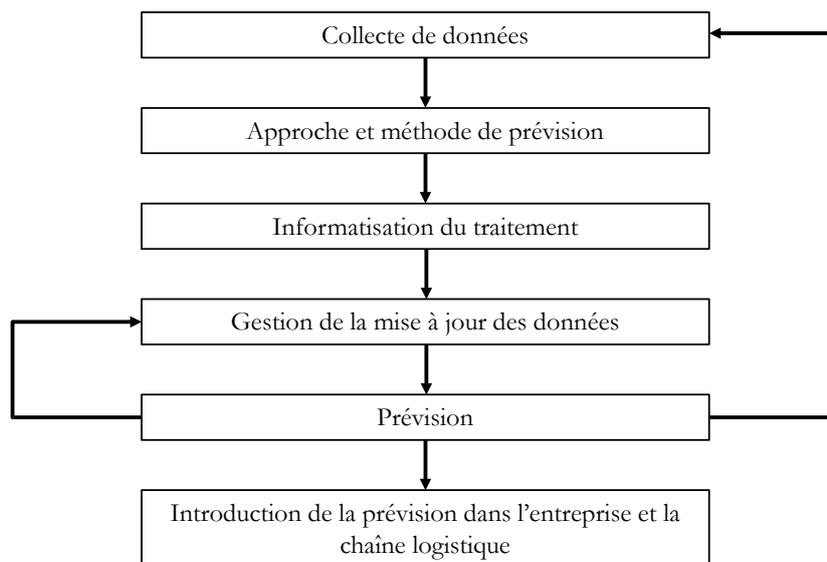


Figure III-3 : Le processus de prévision.

D'après Hadj-Hamou (2004).

Les informations utilisées peuvent être quantitatives (constituées de données chiffrées) ou qualitatives (basées sur des opinions). Les techniques et outils sur lesquels reposent les méthodes de prévisions dépendent bien évidemment de la nature de ces données.

Les méthodes quantitatives partent de données chiffrées. Ces données peuvent être internes à l'entreprise : historique de ventes mensuelles, chiffre d'affaire réalisé au cours des dernières années. On parle alors de méthodes quantitatives endogènes. Mais elles peuvent aussi provenir de l'environnement extérieur (indices des prix, indices économiques, ventes d'autres produits complémentaires ...). On parle alors de méthodes quantitatives exogènes. Ces méthodes sont faciles à mettre en place (un tableur est suffisant pour les calculs). Elles sont simples et rapides à utiliser lorsque le modèle a été développé. Les principaux logiciels du marché les proposent. Les données nécessaires existent le plus souvent dans les systèmes d'information des entreprises (ex. ventes des mois passés) ou sont faciles à trouver (ex. indices économiques).

Lorsqu'on dispose d'un historique suffisant, on va extrapoler cet historique pour obtenir les prévisions. On parle alors d'analyse des séries chronologiques. Les principales méthodes d'analyse des séries chronologiques en production manufacturière sont la moyenne mobile, le lissage exponentiel simple, double (modèle de Holt) ou triple (modèle Holt-Winters) et les méthodes de Box-Jenkins. D'autres méthodes issues du monde économique existent également telles que les modèles économétriques *Inputs/Outputs* (dont *the Regional Economic Modeling System*, REMI et *the Computable Equilibrium Models*). Elles forment le gros des méthodes de prévision proposées par les logiciels du marché. Leur inconvénient principal est leur faible capacité à prendre en compte des facteurs «nouveaux» ou imprévus. La longueur de l'historique est importante. Par exemple, si les ventes sont saisonnières sur une année, il faut disposer d'au moins deux années d'historiques. Dans le cas contraire, les prévisions relèvent du domaine de l'expertise. Pour certains produits, il est possible de relier la demande aux ventes réalisées antérieurement sur d'autres produits (les pièces de rechange par exemple) ou sur des faits passés (la vente de petits pots bébé est liée aux naissances des mois précédents). Lorsqu'il est possible de corrélérer la demande avec des phénomènes sur lesquels on dispose de données statistiques fiables, on utilise des méthodes de corrélation et l'économétrie.

Les méthodes quantitatives s'appuient sur des données. Or, les données dont on dispose, bien qu'elles soient vraies et reflètent la réalité, ne sont pas forcément exploitables telles quelles dans les calculs. Pour un certain nombre de raisons, les données brutes ne sont pas toujours homogènes ou induisent des biais. Il faut donc procéder à un nettoyage préliminaire des données avant de les utiliser (ou en termes plus précis : redresser les chroniques). Inversement, il faudra parfois redresser les prévisions issues des calculs pour tenir compte d'évènements à venir connus (promotions, vacances...).

Dans les méthodes dites qualitatives (parce qu'on n'y traite pas des chiffres et des quantités), on utilise des données subjectives. Ces données dépendent du jugement, de l'expérience et de l'expertise de certains acteurs. Elles sont les seules envisageables lorsqu'il existe très peu de données sur le produit étudié (introduction d'un nouveau produit ou pénétration d'un nouveau

marché, entreprise en démarrage). Les méthodes les plus connues sont la méthode du sondage ou la méthode Delphi. Ces méthodes se basent sur :

- Les opinions d'experts ;
- Les opinions des consommateurs (enquête d'intention ou marché test) ;
- Les opinions des vendeurs et des commerciaux.

Le principal avantage de ces méthodes est de pouvoir prendre en compte de facteurs intangibles. Bien qu'elles soient moins rigoureuses et mathématiques que les méthodes quantitatives, elles peuvent se révéler aussi extrêmement fécondes par les informations qu'elles apportent. Le *brainstorming* (le déballage d'idée) qu'elle provoque peut permettre, non seulement d'aboutir à une prévision, mais aussi de faire naître des idées sur l'évolution du marché, les causes d'achat ou de non-achat du produit étudié, des pistes d'amélioration... A l'inverse, elles comportent des inconvénients tels que :

- La durée du processus de consultation (étude de marché) ;
- Le risque d'obtenir une prévision biaisée ou arbitraire (bon vouloir des commerciaux) ;
- Un coût souvent élevé (consultation d'experts) ;
- Elles sont en général peu précises.

Le choix d'une méthode est fonction du contexte : existence ou non de données chiffrées, stabilité du contexte socio-économique sur l'horizon d'étude. Lorsque c'est possible, il est intéressant de panacher les deux approches qualitatives et quantitatives. De plus, l'éventail des méthodes quantitatives étant assez large et les logiciels du marché en donnant un large éventail, il est utile de recouper les prévisions issues de plusieurs de ces méthodes.

Malgré tous les efforts, les prévisions restent une science inexacte. Pour mettre toutes les chances de son côté, les prévisions doivent être recoupées, voire modifiées par les informations connues par ailleurs : on se renseignera par exemple sur les campagnes publicitaires prévues au service marketing, les négociations en cours au service commercial, les nouveautés attendues chez les concurrents, le lancement d'un nouveau produit pouvant concurrencer le produit étudié. Les prévisions doivent être partagées et validées par les divers services de l'entreprise : logistique, commercial ou marketing. On parle alors de prévisions collaboratives. Ces prévisions collaboratives peuvent inclure les partenaires extérieurs et intégrer des fournisseurs ou des clients.

### *1.1.2. Les prévisions dans le monde humanitaire*

La notion de prévision renvoie au sujet de l'objet considéré et de la demande associée. Dans le cadre d'une catastrophe, l'objet concerne avant toute chose les victimes du désastre. La demande, elle, porte sur la quantité de matériels, produits et de services qu'il convient d'apporter pour soulager lesdites victimes (Quarantelli, 2005). Parler de prévision dans le monde de la gestion de crise, revient à s'intéresser à deux choses :

- La probabilité de survenue d'un événement perturbateur (tremblement de terre, inondation, cyclone, etc.) ;
- L'estimation des conséquences potentielles de la survenue de cet événement (impact sur les infrastructures, victimes, etc.).

Si depuis longtemps on essaie d'anticiper la survenue de catastrophes (notamment par l'étude de phénomènes climatologiques ou géologiques), un rapide tour d'horizon de l'état de l'art sur la gestion de crises montre que peu ou pas de travaux se sont intéressés à l'évaluation *a priori* des conséquences d'une catastrophe. La plupart des travaux existants s'intéressent à l'estimation des conséquences d'une crise une fois que celle-ci est survenue (McEntire et Cope, 2004 ; Clower, 2007 ; Xu et al., 2010 ; Zhang et al., 2013). On parle alors d'évaluation *a posteriori*. Les principales méthodes existantes pour réaliser ce genre d'évaluation sont (Clower, 2007) :

- Les enquêtes : elles présentent l'avantage de capter des informations directement auprès de ceux qui ont subi la catastrophe (UNISDR, 2013). Les supports utilisés sont très variables et peuvent traiter tout autant de données matérielles, sensorielles, économiques ou sociales. Les enquêtes se déclinent également sur des niveaux de granularité plus ou moins importants selon les besoins. Le plus grand problème avec ces approches porte sur le biais évident des réponses et sur la représentativité des échantillons retenus. Rose et Liao (2005) proposent une étude critique de ce genre d'approche notamment en s'intéressant aux enquêtes utilisées pour évaluer les impacts des inondations du centre-ouest des Etats-Unis en 1993 et du tremblement de terre de Northridge en 1994 ;
- La comptabilité économique : il s'agit d'établir le bilan économique complet d'une catastrophe (Zimmerman et al. 2005). Si certaines conséquences sont assez évidentes à chiffrer, d'autres sont beaucoup plus délicates à estimer (Mileti, 1999 ; Cochrane, 2004). C'est le cas évidemment de la dimension humaine (personnes décédées, blessées, etc.) mais également d'éléments comme la perte de monuments historiques, de biens culturels (Mileti, 1999) ou l'intervention de bénévoles dans le cadre de la réponse à la crise (McEntire et Dawson, 2007).

Dans les deux cas, le débat porte alors souvent sur l'estimation des coûts d'une catastrophe et notamment sur la distinction entre coûts directs et coûts indirects. Rose (2004) par exemple aborde cette question en s'intéressant aux conséquences d'une catastrophe qui impacte une centrale électrique. Au total, il est possible de considérer que les effets directs sont ceux qui sont directement visibles après la survenue de la catastrophe (Cochrane, 2004) alors que les effets indirects sont ceux qui seront visibles sur le long terme. De ces premiers éléments, nous pouvons en conclure que la notion de prévision dans un contexte de crise pourra distinguer les deux types de conséquences. Dans nos travaux, nous nous focaliserons uniquement sur l'estimation des conséquences directes.

D'après Clower (2007), il existe également des techniques d'évaluation *a priori* qui sont susceptibles d'être utilisées pour évaluer les effets directs et indirects d'une catastrophe. Et quelques expériences ont d'ailleurs déjà été menées. Ainsi, Nury et al. (2013) utilisent une approche de type Box Jenkins pour prévoir les futurs phénomènes climatologiques (variation de température) du Bangladesh. Greenberg et al. (2007) pour leur part suggèrent d'utiliser une approche économétrique pour anticiper la survenue de catastrophes et d'anticiper les conséquences économiques potentielles même s'ils précisent qu'il n'y a pas aujourd'hui de cas

documentés dans la littérature. Pourtant, de nombreuses institutions publiques américaines se laissent tenter par ces approches et achètent des logiciels proposant ce genre de modèles.

Le sujet de la prévision de demande en cas de crise humanitaire pose question. Ortuño et al. (2013) mentionnent que la nature même des crises, y compris récurrentes, rend délicat toute forme d'anticipation que ce soit pour prévoir l'occurrence ou les conséquences. Pourtant, si on s'intéresse un instant aux catastrophes passées, il est possible de tirer de nombreux enseignements qui mettent à mal ce point de vue si répandu. Nous nous sommes notamment intéressés à la base de données *International Disaster Database*, EM-DAT (<http://www.emdat.be/>), qui contient des données sur les caractéristiques et les effets de plus de 18000 catastrophes survenues dans le monde depuis 1900 jusqu'à nos jours.

Un premier constat concerne le fait que la donnée sur l'occurrence des crises (humanitaires notamment) est disponible et globalement bien renseigné (en tout cas sur une période d'un siècle environ). Comme indiqué précédemment, de nombreux travaux ont d'ores et déjà tenté d'exploiter cette connaissance pour établir des modèles de prévision capables de déterminer les futures évènements climatologiques ou géologiques (inondation, éruption volcanique, tremblement de terre, tempêtes, etc.) susceptibles d'être à l'origine de catastrophes humanitaires. Les travaux du (WGCEP, 2008) sur la prévision de futurs tremblements de terre sont à cet égard particulièrement représentatifs de ce qu'il est aujourd'hui possible de faire. La précision de l'estimation tant en matière de localisation, de magnitude que de date est assez remarquable. Dans le même esprit on retiendra les travaux de (Li et al. 2012 ; Braman et al. 2013) sur les inondations ou ceux de (Tatham et al. 2012) sur les cyclones. Une autre idée couramment évoquée porte sur le fait qu'il y aurait de plus en plus de catastrophes et que la nature même des catastrophes est en train de changer. Il serait présomptueux d'aller à l'encontre de cette affirmation dans un contexte de bouleversement climatique avéré. Néanmoins les proportions ne sont sans doute pas celles que l'on croit. Ainsi, les travaux de (IFRC, 2007 ; IPCC, 2007 ; Charles et al. 2010) démontrent que s'il y a bien une tendance à la démultiplication des catastrophes, leur ampleur moyenne, elle, a tendance à diminuer. Au travers d'une étude détaillée sur une période de 10 ans, (Charles, 2010) démontre que les événements ne sont pas si aléatoires qu'on pourrait le penser (fréquence et intensité globalement constantes sur la période). D'autres chercheurs tels que Peres et al. (2012) ou Kovács et al. (2007) considèrent que pour les petites et moyennes catastrophes, les futures occurrences seront globalement similaires aux précédentes. Nous retenons de cette partie de notre analyse qu'il est raisonnable de considérer que des méthodes de prévision de l'occurrence de futures catastrophes existent ou sont susceptibles d'exister, dès lors que nous nous limitons à l'étude de crises de nature récurrente. Dans la pratique, les analyses des données des catastrophes passées permettent de fournir des informations précieuses sur les tendances à venir (localisation, intensité, typologie et la saisonnalité, etc.).

Pour autant, le sujet va bien au-delà de la simple estimation de l'événement à venir. La question des conséquences potentielles associées audit événement doit également être traitée. Et comme cela a été dit dans le chapitre précédent, cette question fait référence à de nombreuses notions telles que la vulnérabilité et la résilience des territoires. Or, bien plus que l'occurrence des catastrophes, ce sont les caractéristiques des territoires impactés qui évoluent au cours du temps

(Charles, 2010). L'enjeu de la prévision de demande dans un contexte de catastrophe humanitaire semble donc porter sur cet aspect des choses et sur la capacité à estimer *a priori* l'impact associé à la survenue de telle ou telle catastrophe (Comes et al. 2014). Plusieurs auteurs ont proposé des solutions permettant d'adresser le sujet de la prévision dans le monde humanitaire. On trouve notamment les travaux de :

- Beamon et Balcik (2008), dans le cadre d'un sujet sur la conception de chaîne logistique humanitaire, estiment la demande future à partir d'une extrapolation sur la densité de population de la zone concernée. Les auteurs eux-mêmes évoquent les limites de cette approche qui, à défaut d'être robuste, permet au moins de considérer une dynamique temporelle des situations à traiter dans le futur ;
- Dans le cadre d'un travail sur la logistique d'urgence, Sheu (2007) propose de son côté une approche linéaire d'estimation de la demande future basée sur une fonction temps. Ici encore, l'hypothèse de départ est très restrictive et l'approche ne permet en aucun cas de tenir compte des spécificités des territoires et de leur évolution au cours du temps ;
- Lodree et Taskin (2009) abordent le problème de l'organisation logistique des supermarchés face aux surtensions de la demande causées par les ouragans. Les auteurs supposent que la demande est directement proportionnelle à la vitesse prévue du vent des ouragans. De leur côté, Taskin et Lodree (2010) développent un modèle stochastique de programmation multi-période pour gérer les décisions de planification de la production avant la saison des ouragans. La demande de produits est fonction de la quantité des ouragans prévus pour la prochaine saison. Cette estimation est notamment déterminée à partir d'un modèle de taux de probabilité utilisant les Chaînes de Markov. Dans les deux études, les approches préconisées sont très orientées industrie et ne permettent pas d'envisager le dimensionnement et la gestion des chaînes logistiques humanitaires ;
- Rawls et Turnquist (2010) ont proposé un modèle de prévision basé sur l'étude des ouragans ayant frappés les Etats-Unis par le passé. A l'image des travaux de (Charles, 2010 : Salmeron et Apte, 2010) et de beaucoup d'autres, le modèle de prévision se contente ici de reproduire le passé. C'est sans doute l'approche la plus répandue aujourd'hui dans la littérature. L'ensemble des éléments que nous avons développé précédemment indiquent que ces approches, loin d'être dénuées de sens, sont sans doute insuffisantes, notamment parce qu'elles ne considèrent pas suffisamment des critères tels que l'évolution de la vulnérabilité et la résilience des territoires ;
- Kumar et al. (2012) ont analysés non pas les catastrophes passées, mais les conséquences des catastrophes passées. Ils ont ainsi démontré que, à l'échelle de la planète, il était possible de modéliser la répartition des victimes par une distribution de probabilité de Weibull. L'utilisation de cette distribution permet alors aux auteurs de prévoir le nombre de personnes qui seront affectées par des catastrophes naturelles dans les années à venir. Si le résultat de cette recherche semble assez probant, le caractère très macroscopique (loi des grands nombres) de l'approche empêche toute transposition réaliste à des cas plus précis ;
- Davis et al. (2013) explorent quant à eux le terrain des prévisions court terme pour l'aide au pilotage des chaînes logistiques humanitaires. Ils s'intéressent notamment au

cas des ouragans. Ces travaux, bien que tout à fait convaincants, sont en décalage par rapport au sujet de cette thèse qui s'intéresse à l'établissement de décision d'ordre stratégique telle que la conception de chaîne logistique humanitaire.

Au final, il ressort de cette section que les outils et démarches actuels considèrent qu'il n'est pas possible de prévoir les conséquences des futures catastrophes humanitaires. Or, pour certaines d'entre elles (catastrophes récurrentes de petites et moyennes envergures), l'analyse des historiques montrent que l'on répond aux hypothèses de fonctionnement des systèmes de prévision traditionnels. Une transposition est donc sans doute envisageable à la condition que l'on prenne garde à travailler de manière concomitante sur la prévision d'occurrence de l'événement déclencheur et sur la prévision des conséquences (impacts) associées à cette catastrophe. L'objet de la prochaine section est de proposer une méthodologie instrumentée permettant d'atteindre cet objectif.

## **1.2. Contribution n° 1 : modèle de prévision de la demande pour des catastrophes récurrentes**

### *1.2.1. Principes généraux*

En cohérence avec ce que nous venons de développer, rappelons que l'objectif poursuivi est de donner de la consistance aux outils d'aide à la décision que tel ou tel est susceptible de proposer en les éprouvant sur des données quantifiées et réalistes. Des auteurs tels que (Charles, 2010 : Braman et al. 2013) ont démontré récemment les bénéfices potentiels qui seraient associées à l'obtention de telles prévisions de demande sans pour autant proposer de méthode concrète de mise en œuvre.

Conformément aux travaux de (IFRC, 2007 : IPCC, 2007 : Kovács et al. 2007 : Charles, 2010 : EM-DAT, 2011 : Peres et al. 2012 : Braman et al. 2013), nous émettons l'hypothèse qu'il est possible de prévoir les catastrophes futures (de type récurrentes uniquement) en fonction des catastrophes passées. En termes statistiques, il s'agit de considérer que les catastrophes récurrentes répondent aux hypothèses des modèles de prévisions de type endogène. En fonction des types de phénomènes étudiés (tremblement de terre, inondations, cyclones, etc.), de la localisation et des niveaux de granularité visés (temporel, géographique, etc.), l'estimation de l'occurrence peut varier du plus simple au plus complexe. Le modèle le plus simple consiste souvent à considérer que les événements futurs reprendront peu ou prou les événements passés. C'est cette hypothèse que nous retrouvons aujourd'hui dans la très grande majorité des travaux de recherche publiés (Charles, 2010 : Peres et al. 2012 : Galindo et Batta, 2013). Cette hypothèse peut être considérée comme valide dès lors que les informations sont gérées et utilisées de manière relativement agrégée, c'est à dire pour alimenter des décisions d'ordre stratégique ou tactique. Par exemple, dans les travaux de (Charles, 2010), il s'agit de travailler à l'échelle de la planète et sur un horizon de temps pluriannuel. Dans d'autres cas, nécessitant plus de précision (décision opérationnelle), des spécialistes de tel ou tel phénomène géologique ou climatique sont en capacité de proposer des modèles de prévision d'occurrence spécifiques à un terrain d'application et sur un horizon de temps plus fin. C'est notamment le cas des travaux de (Li et al. 2012 : Braman et al. 2013) sur les inondations, du (WGCEP, 2008) sur les tremblements de terre ou ceux de (Tatham et al. 2012) sur les cyclones. Dans le cadre de notre étude, nous nous

intéressons aux décisions d'ordre stratégique, dans un contexte de catastrophes récurrentes quelque soit leur nature (inondations, tremblements de terre, cyclones, etc.). Aussi, nous retiendrons pour la suite l'hypothèse d'un modèle de prévision des occurrences futures basée sur la reconduction de phénomènes passés.

Néanmoins l'objet de notre proposition est bien d'aller au-delà de la prévision d'occurrence en s'intéressant notamment à l'établissement d'une prévision d'impact de la future catastrophe. Pour ce faire, nous proposons une démarche en 4 étapes telles que décrites dans la figure suivante :

1. Identifier l'ensemble des facteurs d'influence qui sont susceptibles d'expliquer le lien entre la survenue d'une catastrophe et ses conséquences attendues ;
2. Sélectionner, parmi cet ensemble, les facteurs discriminants pour un type de catastrophe et un territoire donnés ;
3. Etablir la fonction de corrélation associée au contexte particulier étudié ;
4. Valider le modèle de prévision.

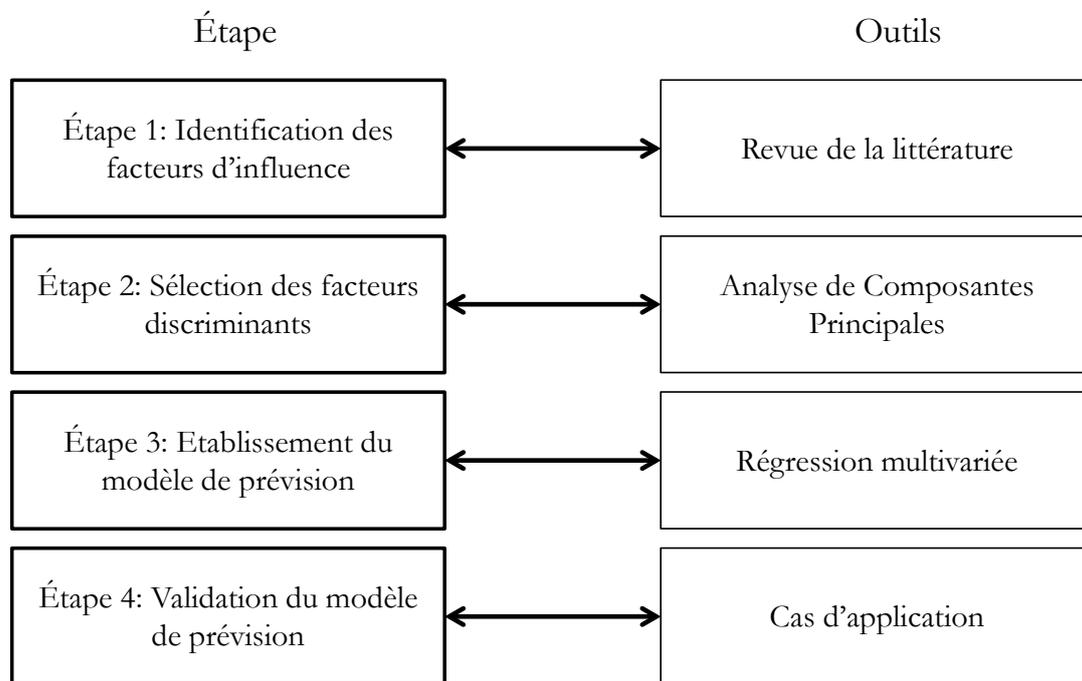


Figure III-4 : Méthodologie de la prévision en cas de crises.

#### 1.2.1.1. Etape 1 : Identification des facteurs d'influence.

Cette étape s'appuie sur une revue de la littérature permettant de comprendre les liens existant entre la survenue d'une catastrophe et ses conséquences directes. Au-delà des éléments précédemment présentés, Wisner et al. (2004) ont démontré il y a dix ans déjà, que l'impact d'une catastrophe est directement fonction de la vulnérabilité (telle que définie dans le chapitre précédent). Des travaux plus récents (Peres et al. 2012 ; Ortuño et al. 2013) ont quant à eux souligné le lien qui existe entre la résilience (telle que définie également au chapitre précédent) et l'impact d'une catastrophe.

Cette double dépendance est reprise par (UNESCAP, 2008) et peut être synthétisée par la formule suivante :

$$\text{Impact d'une catastrophe} = f(\text{vulnérabilité, résilience}) \quad (1)$$

Dès lors, il convient de s'intéresser en détail aux facteurs qui sont susceptibles de qualifier et quantifier les concepts de vulnérabilité d'une part et de résilience d'autre part, pour les événements récurrents ou cycliques. L'analyse des travaux de (Hardy, 2001 : UNDP, 2004 : D'Ercole et al. 2009 : Alinovi et al. 2009 : D'Ercole et al. 2012 : Metzger et al. 2013) ont permis de mettre au jour des dizaines de facteurs de caractérisation de ces notions. Les deux tableaux suivants reprennent chacun de ces facteurs. Ils précisent aussi, lorsque c'est possible, les sources (bases de données) auprès desquelles il est possible de récupérer les données quantitatives. Selon les facteurs, ces bases sont internationales ou nationales. Dans ce dernier cas, les bases citées ici sont les bases péruviennes utilisées pour l'étude de cas du dernier chapitre.

Ce travail bibliographique montre que la vulnérabilité peut théoriquement revêtir plusieurs dimensions complémentaires. Il y a notamment les données relatives au contexte et aux activités économiques du territoire concerné. On doit également considérer les données environnementales, démographiques et sanitaires. Le niveau d'éducation est également prépondérant, au même titre que les capacités de développement humain. L'accès aux ressources et la qualité intrinsèque des infrastructures constituent également une catégorie à retenir. Enfin, la vulnérabilité de la réponse humanitaire doit être considérée.

Catégorie	Facteurs de caractérisation de la vulnérabilité	Base de données exploitable
Contexte économique	Produit Intérieur Brut (PIB) par habitant	WB <sup>1</sup>
	Indice de pauvreté humaine (IPH)	UNDP <sup>2</sup>
	Balance commerciale (% des exportations de biens et services)	WB
	Inflation annuelle (%)	WB
	Taux de chômage (% de la population active totale)	ILO <sup>3</sup>
Type des activités économiques	Surfaces agricoles (en milliers d'hectares)	FAO <sup>4</sup>
	Ratio de terres arables	FAO
	Taux d'urbanisation	UNPOP <sup>5</sup>
	Part de l'agriculture dans le Produit National Brut (PNB)	WB / FAO
Démographie	Taux de croissance démographique	UNDESA
	Taux de croissance des centres urbains	GRID <sup>6</sup>
	Densité de population	GRID <sup>7</sup>
	L'âge de dépendance	WB

Table III-1 : Les facteurs de caractérisation de la vulnérabilité - Partie 1.

D'après UNDP (2004), Metzger et al. (2013), D'Ercole et al (2009) et Hardy (2001).

<sup>1</sup> WB: La Banque Mondiale (*The World Bank*).

<sup>2</sup> UNDP : Le Programme de Développement des Nations Unies (*The United Nations Development Program*).

<sup>3</sup> ILO : Le Bureau International du Travail (*The International Labor Office*).

<sup>4</sup> FAO : L'Organisation pour l'Alimentation et l'Agriculture (*The Food and Agriculture Organization*).

<sup>5</sup> UNPOP : La Division des Nations Unies pour la population (*The United Nations Population Division*).

<sup>6</sup> Le calcul a été fait utilisant les données du Département des Affaires Sociales et Economiques des Nations Unies

<sup>7</sup> Le calcul a été fait avec la modélisation spatiale de l'UNEP/GRID avec les données de la population en CIESIN.

Santé et hygiène	Nombre de médecins (pour 1000 habitants)	WB
	Nombre de lits dans les hôpitaux (pour 1000 habitants)	WB
	Espérance de vie à la naissance	UNDESA <sup>1</sup>
	Taux de mortalité infantile (moins de 5 ans)	UNDESA
Qualité de l'environnement	Surfaces forestières (en % de la superficie du territoire)	FAO
	Indice GLASOD (niveau de dégradation des sols par l'homme)	FAO/UNEP <sup>2</sup>
Capacité d'alerte	Nombre de radios ou équivalents (pour 1000 habitants)	WB
Éducation	Taux d'alphabétisation	WB
Développement	Indice de Développement Humain (IDH)	UNDP
Exposition	Exposition aux risques géologiques (séismes, inondations, etc.)	IGP <sup>3</sup>
Accès aux ressources	Accessibilité des ressources en situation d'urgence	INDECI
Vulnérabilité structurelle et fonctionnelle	Vulnérabilité des infrastructures (ponts, chemins de fers, échangeurs, aéroports, etc.) vis-à-vis du fonctionnement nominal des ressources de première nécessité (hôpitaux, pompiers, etc.).	N/A
Accès aux ressources et connaissance de la vulnérabilité de la population	L'accès aux ressources	INEI
	Pyramide des âges	INEI
	Part de la population ayant accès aux services sanitaires élémentaires	INEI
	Part de la population n'ayant pas accès aux services sanitaires élémentaires	INEI
	Densité familiale moyenne (nombre de personnes par foyer)	INEI
	Taux de raccordement à l'eau potable	INEI
	Taux de raccordement aux systèmes d'évacuation des eaux usés	INEI
	Taux de raccordement aux réseaux électriques	INEI
	Taux d'habitation précaire (terre battue, murs fragiles, etc.)	INEI
	Taux de population déplacée	INEI
	Taux d'immigration	INEI
	Taux de couverture des réseaux de transport public	INEI
	Taux d'éducation	INEI
Vulnérabilité de la réponse humanitaire	Niveau de coordination nationale des crises	INDECI
	Qualité des secours	INDECI
	Maturité et expertise des organisations de secours	INDECI
	Disponibilité des ressources dédiées aux réponses humanitaires	INDECI
	Disponibilité des centres d'approvisionnement d'eau	INDECI
	Disponibilité des centres d'approvisionnement de nourriture	INDECI
	Disponibilité des centres de santé	INDECI
	Disponibilité des réseaux de télécommunication et d'énergie	INDECI
	Disponibilité des moyens de transport et des routes	INDECI
	Disponibilité d'abris et établissement de refuge	INDECI
	Disponibilité de ressources de stockage et traitement des déchets	INDECI
	Assise financière des acteurs de la réponse	INDECI
	Niveau de visibilité (risques, vulnérabilité, etc.) du territoire	INDECI

Table III-2 : Les facteurs de caractérisation de la vulnérabilité - Partie 2.

D'après UNDP (2004), Metzger et al. (2013), D'Ercole et al (2009) et Hardy (2001).

<sup>1</sup> UNDESA : Le Département des Affaires Sociales et Economiques des Nations Unies/Division Population.

<sup>2</sup> UNEP : L'Equipe pour les Perspectives d'Environnement Mondial (*The Global Environment Outlook Team*).

<sup>3</sup> IGP : Institut de géophysique du Pérou (*The Geophysical Institute of Peru*).

De son côté, la résilience se concrétise au travers du niveau de revenu et de la facilité d'accès aux produits de première nécessité, les services publics et services sociaux sont également à considérer. La résilience fait aussi référence à l'ampleur des actifs présents sur un territoire. Enfin, la capacité d'adaptation et la stabilité sociétale terminent de caractériser cette notion.

Catégorie	Facteurs de caractérisation de la résilience	Base de données exploitable
Revenu et accès à l'alimentation	Revenu moyen par personne par jour	INEI <sup>1</sup>
	Dépense moyenne par personne et par jour	INEI
	Taux d'insécurité alimentaire	N/A
	Taux de qualité globale alimentaire	N/A
	Consommation moyenne alimentaire (kcal par personne et par jour)	N/A
Accès aux services de base	Proximité des services de santé	MINSA <sup>2</sup>
	Niveau de qualité des services de santé	MINSA
	Niveau de qualité du système éducatif	MINEDU <sup>3</sup>
	Taux d'insécurité	N/A
	Couverture et efficacité des transports publics	MTC <sup>4</sup>
	Couverture et efficacité des réseaux d'eau, d'électricité et de téléphone	MTC
Services sociaux	Moyens engagés dans le cadre des services sociaux (ratio pour 1000 habitants)	MTPE <sup>5</sup>
	Taux de qualité de l'aide apportée	MTPE
	Taux de la population active suivie dans le cadre de l'aide à l'emploi	MTPE
	Durée moyenne d'un accompagnement	MTPE
	Efficacité perçue des services sociaux	MTPE
Actifs	Surface moyenne des logements	INEI
	Taux d'équipement moyen (TV, voitures, etc.)	INEI
	L'Unité Bétail Tropical (UBT) équivalent à 250 kg	N/A
	Taux de propriétaires (vs locataires)	MINAG <sup>6</sup>
Capacité d'adaptation	Nombre de sources de revenus par foyer (moyenne)	INEI
	Niveau d'études moyen du foyer	INEI
	Taux d'emploi moyen du foyer	INEI
	Ratio de consommation alimentaire (dépenses alimentaires/dép. totales)	INEI
Stabilité	Taux de chômage par foyer	INEI
	Evolution des revenus (%)	INEI
	Evolution des dépenses (%)	INEI
	Taux d'épargne moyen	INEI
	Reste à vivre moyen	INEI
	Stabilité du système éducatif	INEI

Table III-3 : Les facteurs de caractérisation de la résilience.

D'après Alinovi et al (2009).

<sup>1</sup> INEI: L'Institut National de la Statistique et de l'Informatique au Pérou (*The National Institute of Statistics and Informatics*).

<sup>2</sup> MINSA : Le Ministère Péruvien de la Santé (*The Peruvian Ministry of Health*).

<sup>3</sup> MINEDU: Le Ministère Péruvien de l'Education (*The Peruvian Ministry of Education*).

<sup>4</sup> MTC: Le Ministère Péruvien des Transports et Communications (*The Ministry of Transport and Communications of Peru*).

<sup>5</sup> MTPE : Le Ministère Péruvien du Travail (*The Peruvian Ministry of Labor*).

<sup>6</sup> MINAG: Le Ministère Péruvien de l'Agriculture (*The Peruvian Ministry of Agriculture*).

A l'évidence, tous ces facteurs ne sont pas indépendants et ne sont pas nécessairement aussi sensibles sur les conséquences potentielles de toutes les catastrophes. En pratique, le fait de s'intéresser à des crises de type inondation en Amérique du Sud ou de type cyclones en Asie du Sud-Est invitera à considérer tel ou tel facteur de façon privilégiée. La section suivante s'intéresse à cette question en proposant une méthode permettant de sélectionner les facteurs à prendre en compte pour une étude bien spécifique.

#### 1.2.1.2. Etape 2 : Sélection des facteurs discriminants.

Considérant l'ensemble des facteurs que nous venons d'identifier, il est désormais clair que l'établissement d'un modèle de prévision d'impact d'une catastrophe est un système exogène (au contraire de l'estimation d'occurrence qui pouvait être considérée comme endogène). Pour autant, le nombre de facteurs est sans doute à ce stade beaucoup trop important pour être exploitable. En outre, l'interdépendance de ces facteurs n'est nullement garantie, pas plus que leur intérêt à décrire telle ou telle situation spécifique.

La question qui se pose alors est de savoir comment identifier et comprendre, dans le cadre de notre étude, les facteurs remarquables qui permettront d'établir un modèle de prévision des conséquences de catastrophes qui soit pertinent (Pang, 2011 ; Contreras et al. 2013). Pour ce faire, nous proposons d'utiliser des outils d'analyse de données. Les méthodes d'analyse de données permettent en effet une étude globale des individus et des variables les caractérisant en utilisant généralement des représentations graphiques. Dans le contexte de transition actuel où les données qui étaient en quantité limitée sont désormais fréquemment en profusion, mais à l'état brut, les objectifs de telles méthodes sont multiples : maîtriser la qualité de l'information, la traduire sous une forme exploitable, en extraire des connaissances, réaliser un diagnostic (Lauras, 2004).

Plusieurs méthodes classiques d'analyse de données sont à notre disposition. Ces méthodes peuvent être regroupées au sein de deux grandes familles (Lauras, 2004). D'abord, les méthodes de *statistique exploratoire* ou descriptive qui synthétisent et résument l'information contenue dans les données. Leur rôle est de mettre en évidence des propriétés de la population observée et de suggérer des hypothèses. On distinguera dans les statistiques descriptives :

- Les méthodes de classification (former des groupes d'individus) ;
- Les méthodes factorielles (réduire le nombre de caractéristiques) ;
  - ❖ Analyse en composantes principales (quantitatif) ;
  - ❖ Analyse des correspondances (qualitatif) ;
  - ❖ Analyse canonique (lien entre caractères).

Ensuite, les méthodes dites de *statistique inférentielle* ou explicative qui partent (à l'inverse d'un raisonnement déductif classique) de plusieurs réalisations d'un phénomène pour essayer de tirer des conclusions quant aux lois de probabilité qui le régissent. En d'autres termes, la statistique inférentielle étend les propriétés constatées sur un échantillon à la population tout entière, et confirme ou infirme les hypothèses formulées a priori. Plusieurs techniques sont associées à ces méthodes :

- Les techniques d'estimation recherchent les lois de comportement qui décrivent au mieux les mesures dans le cadre d'une incertitude supposée ;
- Les techniques de vérification d'hypothèses (tests) exploitent la propriété mathématique supposée d'une variable (loi de distribution de probabilité) pour éprouver un échantillon.

Dans nos travaux, nous cherchons à identifier quels sont les facteurs discriminants de façon à établir dans un second temps des modèles de corrélation adaptés à la situation étudiée. Notre étude se situe donc clairement dans le cadre d'une politique de statistique exploratoire et nécessite plus particulièrement l'utilisation de méthodes factorielles. L'art consiste alors à savoir maîtriser un rapport de forces entre un modèle de représentation – basé sur la recherche de corrélations linéaires entre paires de caractères – des données et les valeurs des données elles-mêmes. L'Analyse en Composantes Principales (ACP) et l'Analyse Factorielle des Correspondances (AFC) sont les outils qui répondent à cette problématique. L'ACP traite le domaine quantitatif alors que l'AFC traite le domaine qualitatif. Dans notre étude, les données disponibles sont d'ordre exclusivement quantitatif. L'ACP semble donc être une approche satisfaisante pour notre problème, à l'instar de ce que les travaux de Bernanke et Boivin (2003), Boivin et Ng (2005, 2006) ou Bai et Ng (2006 : 2008) ont pu démontrer.

D'après (Saporta, 1990 : Escofier et Pagès, 1997 : Bouroche et Saporta, 2002 : Lauras, 2004), l'ACP a pour objet la description de données contenues dans un tableau individus-caractères numériques par une représentation approchée du nuage des individus dans un sous espace de caractères de dimension plus faible. L'objectif est de permettre, par des interprétations graphiques, de déterminer des groupes d'individus qui se distinguent des autres et des caractères discriminants. Il s'agit de la méthode de base de l'analyse de données (Saporta, 1990).

Considérons un exemple issu de (Bouroche et Saporta, 2002) avec seulement deux caractères  $x^1$  et  $x^2$ . Il est alors possible de représenter sur un plan l'ensemble des données. Chaque individu est alors représenté par un point de coordonnées  $x_i^1$  et  $x_i^2$  et la simple analyse de l'allure du graphique permet de déterminer l'intensité de la liaison (cf. figure suivante). Trois options sont alors possibles : absence de liaison entre les individus, forte liaison (corrélation) entre les individus, mise en évidence de groupes homogènes disposant de caractéristiques semblables relativement à  $x^1$  et  $x^2$ .

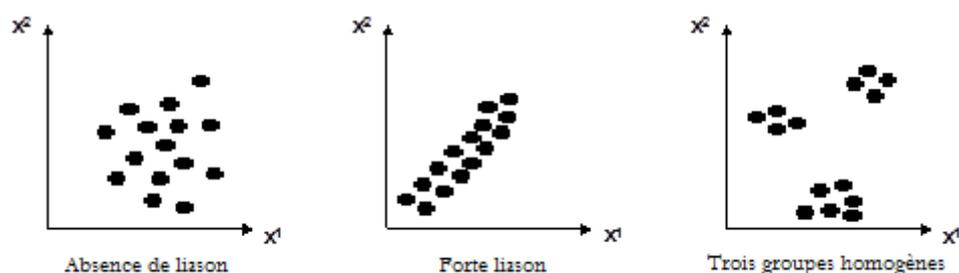


Figure III-5 : Liaisons possibles entre individus.

D'après Bouroche et Saporta (2002).

Avec trois caractères, il est encore envisageable de conduire l'étude de façon visuelle au moyen de la géométrie dans l'espace. Au-delà, il n'est plus possible de faire une représentation graphique des liaisons. Le positionnement d'individus relativement à une étude portant sur  $n$  caractères formera en effet un nuage dans un espace à  $n$  dimensions, puisqu'il y aura  $n$  coordonnées. Pour pouvoir tout de même disposer d'une représentation graphique, il faut projeter les points qui représentent les individus sur un plan (cf. figure suivante). Une telle démarche conduit mécaniquement à une représentation déformée de la configuration exacte (toutes les distances projetées ne pouvant, par construction, être égales aux distances originelles).

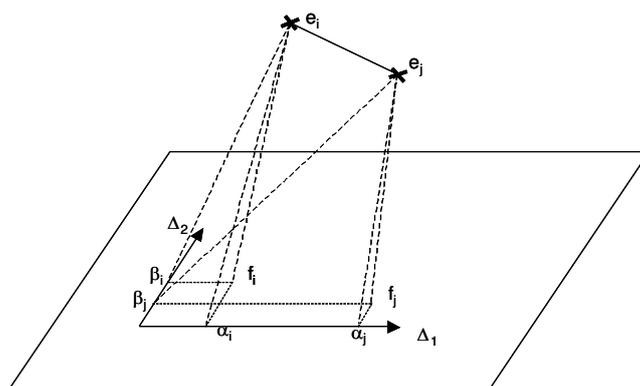


Figure III-6 : Exemple de projection des individus sur un plan.

Source : Lauras (2014).

L'objectif est bien sûr de déterminer le plan de projection pour lequel les distances seront les mieux conservées. Puisque l'opération de projection raccourcit (ou maintient dans le meilleur des cas) les distances entre les points ( $d(f_i, f_j) \leq d(e_i, e_j)$ ), on prendra comme critère la maximisation de la moyenne des carrés des distances entre les projections. Le plan sera alors défini par deux droites  $\Delta_1$  et  $\Delta_2$  perpendiculaires telles que :  $d^2(f_i, f_j) = d^2(\alpha_i, \alpha_j) + d^2(\beta_i, \beta_j)$  où les  $\alpha_i$  et les  $\beta_i$  sont les projections des individus  $e_i$  (et donc des  $f_i$ ) sur  $\Delta_1$  et  $\Delta_2$ . La méthode consiste alors à chercher tout d'abord  $\Delta_1$  en identifiant la valeur maximale des  $d^2(\alpha_i, \alpha_j)$  puis  $\Delta_2$  perpendiculaire à  $\Delta_1$ , qui rend maximale les  $d^2(\beta_i, \beta_j)$ . En prolongeant la méthode au-delà du plan, on trouvera les  $\Delta_3, \Delta_4, \Delta_5, \dots, \Delta_p$  perpendiculaires entre elles. Les  $\Delta_i$  sont appelées les axes principaux du nuage.

En projetant les individus qui avaient pour coordonnées initiales  $(x_i^1, x_i^2, \dots, x_i^n)$  sur les axes principaux, on obtient de nouvelles coordonnées  $(c_i^1, c_i^2, \dots, c_i^n)$ . On construit ainsi de nouveaux caractères appelés composantes principales  $(c^1, c^2, \dots, c^n)$ . Chaque composante se définit alors comme la liste des coordonnées des  $n$  individus sur l'axe  $\Delta_k$  et est une composante linéaire des caractères initiaux :  $c^k = u_1^k x^1 + u_2^k x^2 + \dots + u_p^k x^n$  où les  $u_i^k$  forment le  $k$ -ième facteur principal  $u^k$ . Le principe mathématique consiste donc à effectuer des changements de repère, dans l'espace des caractères, et à calculer le nombre minimal d'axes qui conservent les distances initiales entre les points, après projection.

Finalement, l'ACP est une technique permettant de diagnostiquer le degré de dépendance linéaire entre les caractères d'un système complexe (ce qui est notre objectif, ndlr). L'ACP est donc une méthode de réduction du nombre de caractères permettant des représentations géométriques des

individus et des caractères. Cette réduction n'est possible que si les caractères initiaux ne sont pas indépendants et ont donc des coefficients de corrélation non nuls. L'ACP est une méthode factorielle, car la réduction du nombre des caractères ne se fait pas par une sélection de certains d'entre eux, mais par la construction de nouveaux caractères synthétiques obtenus en combinant les caractères initiaux au moyen des « facteurs ». Les nouveaux caractères sont des combinaisons linéaires des anciens. Ce qui fait la spécificité de l'ACP, c'est qu'elle traite exclusivement de caractères numériques jouant tous le même rôle.

L'utilisation d'une méthode ACP sur les données numériques associées aux facteurs issus de l'étape 1 de notre méthodologie permettra d'obtenir une liste de facteurs discriminants pour le cas étudié, tant pour caractériser la vulnérabilité que pour caractériser la résilience. Evidemment, s'agissant de statistiques exploratoires, le résultat sera directement dépendant des données de base collectées et analysées. Plus les données seront nombreuses, plus le résultat obtenu à l'issue de cette deuxième étape sera pertinent.

### 1.2.1.3. Etape 3 : Etablissement du modèle de prévision.

L'objet de cette troisième étape est désormais de constituer concrètement le modèle de prévision correspondant au cas étudié. Compte tenu de l'hypothèse retenue selon laquelle les occurrences futures sont similaires aux occurrences passées, cette étape cherchera dans un premier temps à expliquer au travers des facteurs discriminants issus de l'étape 2, la corrélation existante entre les catastrophes passées et leurs conséquences. L'approche employée consiste alors en une simple analyse de régression à variables multiples basée sur la formule suivante :

$$\text{Impact constaté d'une catastrophe} = f(v_1, v_2, \dots, v_m; r_1, r_2, \dots, r_n) \quad (2)$$

Dans laquelle :

$\{V_1, \dots, V_m\}$  sont les facteurs discriminants (identifiés lors de l'étape 2) caractérisant la vulnérabilité.

$\{R_1, \dots, R_n\}$  sont les facteurs discriminants (identifiés lors de l'étape 2) caractérisant la résilience.

Sans essayer de procéder aux démonstrations, il convient de s'arrêter un instant sur les mécanismes de fonctionnement de la corrélation linéaire multiple. L'objectif est de déterminer la valeur d'une variable numérique aléatoire, dite « expliquée », par les valeurs de plusieurs autres variables, dites « explicatives », numériques ou binaires. Dans notre cas, la variable expliquée est le nombre de victimes impacté par la catastrophe étudiée alors que les variables explicatives sont celles issues de l'étape 2 de notre démarche. La mise en œuvre d'une telle approche suppose donc qu'un modèle décrit parfaitement la réalité et l'on cherche à s'en approcher :

$$y_i = a_1x_{1i} + a_2x_{2i} + \dots + a_kx_{ki} + b + \epsilon_i \quad (3)$$

Néanmoins, il faut rester réaliste et le modèle ne sera pas parfait. Pour chaque observation ou presque, il existera une différence entre la valeur mesurée dans la réalité et la valeur calculée par le modèle. C'est ce que l'on nomme les résidus. (<http://jybaudot.fr/>). Saporta (2006) et Baudot (<http://jybaudot.fr/>) décrivent le mécanisme de régression linéaire multiple comme suit : les

paramètres de la régression, estimateurs des vrais paramètres inconnus, sont des variables aléatoires dont on peut estimer les variances. Soit  $X$  la matrice rectangulaire des valeurs des  $k$  variables explicatives pour les  $n$  observations, plus une colonne de 1 (pour la constante).  $Y$  est la matrice colonne des valeurs de la variable à expliquer. Soit  $X'$  la transposée de  $X$ . La matrice des coefficients est égale à  $(X'X)^{-1}X'Y$ . Cette matrice fournit les espérances des vrais paramètres. Mais il faut également connaître leurs dispersions pour juger de leur plus ou moins bonne « qualité ». On calcule alors l'erreur quadratique moyenne  $\sigma^2$  (somme des carrés des résidus divisée par  $n - k - 1$ ) qui permet d'établir leur matrice des variances-covariances (matrice diagonale avec les variances sur la diagonale et les covariances en élément extra-diagonal) :  $\sigma^2(X'X)^{-1}$ . Une fois l'analyse des événements passés réalisée, il est possible d'établir un modèle de prévision capable de prévoir l'impact (en nombre de victimes) de la survenue probable à la période  $t$  de telle ou telle catastrophe. Ce modèle de prévision est représenté dans le cadre de la formule suivante :

$$\text{Impact attendu d'une catastrophe future} = f_{t=1,\dots,T}(x_1, \dots, x_k, \dots, x_{mxn}) \quad (4)$$

Dans laquelle :

$x_k$  est une variable indépendante composée de  $\{m \times n\}$  valeurs enregistrées dans une période donnée  $t$ , à partir d'une base de données qui comprend  $T$  périodes totales.

L'ACP a été appliquée comme outil complémentaire aux outils de prévision économétriques. (Pang, 2011) estime que la prévision fondée sur l'utilisation d'ACP améliore la précision de la prédiction. Il donne comme exemple les cas de prévision macroéconomiques. D'autres cas de prévisions similaires ont été fait par (Bernanke et Boivin, 2003 ; Bovin et Ng, 2005 ; 2006 ; Bai et Ng, 2006 ; 2008). Dans l'humanitaire notamment, les travaux applicatifs de (Frazier et al., 2013) et les travaux connexes de (Alsahehri et al., 2013) sont remarquables, sur la caractérisation des territoires et de la population face aux crises.

#### 1.2.1.4. Etape 4 : Validation du modèle de prévision.

La qualité globale d'une régression linéaire multiple peut être évaluée par le coefficient de détermination (appelé aussi coefficient de corrélation) ou par l'Akaike Information Criterion (souvent jugé plus fiable). Ces tests sont généralement utilisés lorsque les données ne sont pas suffisamment nombreuses pour disposer d'un échantillon d'apprentissage et d'un échantillon de test.

Dans notre cas, les catastrophes récurrentes sont par définition nombreuses et il est donc possible de disposer d'un échantillon suffisamment vaste pour tester la qualité globale du modèle obtenu. En d'autres termes, il s'agit d'essayer de reconstruire certains événements passés à partir du modèle obtenu et de mesurer l'écart entre le résultat théorique et la réalité des faits. Ce critère est généralement appelé ratio de déviation ou résidu. Nous avons considéré les bornes suivantes pour évaluer la qualité des modèles de prévision obtenus :

Si  $75 \% < Ratio < 100 \%$  alors le modèle est considéré comme « bon ».

Si  $50 \% < Ratio < 75 \%$  alors le modèle est considéré comme « douteux ».

Si  $Ratio < 50 \%$  alors le modèle est considéré comme « mauvais ».

### 1.2.2. Application numérique

Pour faciliter la compréhension de l'outil, nous présentons une application numérique simple. On s'intéresse à la prévision de la demande d'aide humanitaire en réponse à des catastrophes de type tremblement de terre ou inondation, dans une région d'un pays donnée. Ce cas exemple est inspiré de données issues de la région de Cusco au Pérou.

#### Étape 1 : Identification des facteurs d'influence.

Parmi tous les facteurs mentionnés précédemment, nous ne retenons que les 12 facteurs présentés dans le tableau ci-dessous. Ces 12 facteurs sont ceux qui, au dire des experts locaux de la gestion de crise, sont a priori les plus significatifs eu égard aux types de catastrophes que le territoire concerné a l'habitude de subir.

Variables	Facteurs					
<b>Vulnérabilité</b>	Densité de population (quantité ; en millions)	Indice de Développement Humain, IDH (ratio)	Taux de raccordement aux réseaux d'eau potable (%)	Taux de raccordement aux réseaux électriques (%)	Taux d'habitation précaire (%)	Espérance de vie à la naissance (/ an et pour mille enfants)
<b>Résilience</b>	Taux d'alphabétisation (%)	Taux de scolarisation dans le secondaire (%)	Taux de réussite éducatif aux examens du secondaire (%)	Revenu moyen par personne par jour ( <i>Nuevos Soles</i> )	Taux d'insécurité (nb de plaintes / mille habitants)	Taux de qualité de l'aide apportée (montant des aides sociales / millions d'habitants)

Table III-4 : Facteurs caractérisant l'état de vulnérabilité et de résilience.

Le tableau ci-dessous fournit les données numériques associées pour les années 1993, 2000 et 2007.

Facteurs	Années		
	1993	2000	2007
<b>Vulnérabilité</b>			
Densité de population (mille)	983,7	1067,3	1063,5
Indice de Développement Humain, IDH	0,53	0,58	0,60
Taux de raccordement à l'eau potable	49 %	61 %	72 %
Taux de raccordement aux réseaux électriques	45 %	65 %	73 %
Taux d'habitation précaire (ou vulnérabilité de la construction)	69 %	68 %	59 %
Espérance de vie à la naissance	66,8	67,9	72,3
<b>Résilience</b>			
Taux d'alphabétisation	78 %	81 %	88 %
Niveau de qualité du système éducatif (secondary registration)	65 %	75 %	86 %
Niveau de qualité du système éducatif (réussite éducatif)	74 %	79 %	87 %
Revenu moyen <sup>1</sup>	87,9	307,27	320,8
Taux d'insécurité <sup>1</sup> (ou nombre de crimes et délits)	7,88	8,20	4,50
Taux de qualité de l'aide apportée <sup>1</sup>	0,0	0,0	0,4

<sup>1</sup> Voir l'explication dans le tableau précédent.

Table III-5 : Données pour le cas exemple. Cas d'Ancash au Pérou.

Etape 2 : Sélection des facteurs discriminants.

Conformément à la méthodologie présentée, nous réalisons une Analyse en Composantes Principales (ACP) sur ces données. Le logiciel « Le Sphinx Plus<sup>2</sup>-V5 » a été utilisé pour ce faire. Les résultats sont présentés ci-dessous :

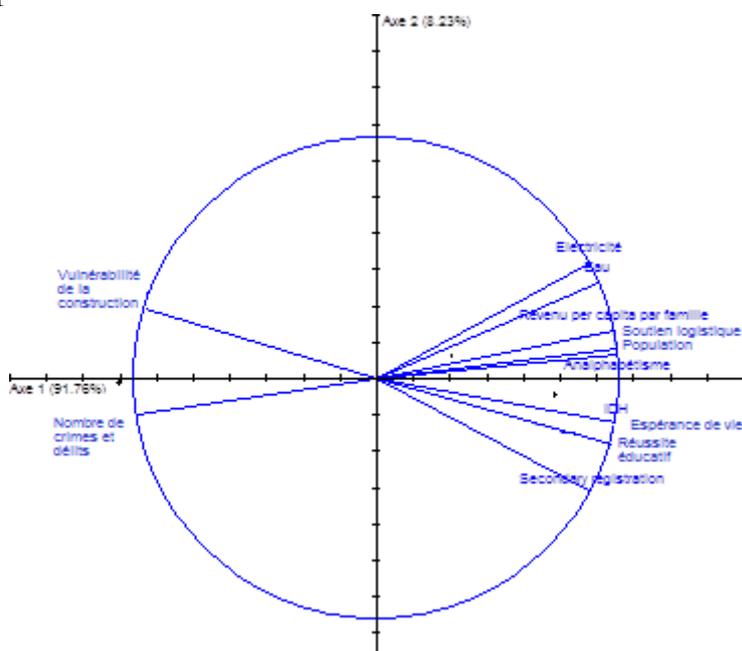


Figure III-7 : Carte d'Analyse en Composantes Principales sur le cas exemple.

Cette carte donne les positions des 12 critères et les coordonnées des 3 observations (1993, 2000 et 2007). Les deux axes représentés expliquent 99,99 % de la variance, ce qui est très satisfaisant. Chaque observation est représentée par un point.

Facteurs	Population	IDH	(EV)	(AN)	(SR)	(RE)	(RP)	(NCD)	(SL)	Eau	(EL)	(VDC)
Population	1,00											
IDH	0,96	1,00										
Espérance de vie (EV)	0,93	1,00	1,00									
Analphabétisme (AN)	1,00	0,98	0,96	1,00								
Secondary registration (SR)	0,82	0,95	0,98	0,88	1,00							
Réussite éducatif (RE)	0,93	1,00	1,00	0,96	0,98	1,00						
Revenu per capita (RP)	0,99	0,92	0,88	0,98	0,76	0,88	1,00					
Nombre crimes et délits (NCD)	-1,00	-0,94	-0,91	-0,99	-0,80	-0,91	-1,00	1,00				
Soutien logistique (SL)	1,00	0,95	0,92	0,99	0,82	0,92	1,00	-1,00	1,00			
Eau	0,95	0,83	0,77	0,92	0,82	0,77	0,98	-0,97	0,96	1,00		
Électricité (EL)	0,92	0,77	0,71	0,88	0,54	0,71	0,96	-0,94	0,93	1,00	1,00	
Vulnérabilité de const. (VDC)	-0,92	-0,99	-1,00	-0,95	-0,98	-1,00	-0,87	0,90	-0,91	-0,76	-0,69	1,00

Table III-6 : Corrélations entre les facteurs de la vulnérabilité et résilience.

On observe alors qu'il y a une forte corrélation ( $R > 0,85$ ) entre la plupart des facteurs étudiés, à l'exception de ceux relatifs à la couverture des réseaux d'électricité (EL) et d'eau (E). Ces deux derniers facteurs sont par ailleurs très fortement corrélés dans cet exemple. Il est alors possible de conclure que, pour cet exemple, les facteurs IDH et E sont les facteurs discriminants, qui expliquent à eux seuls le niveau de résilience et de vulnérabilité de ce territoire.

Etape 3 : Etablissement du modèle de prévision.

Pour déterminer l'équation liant le nombre de victimes avec ces trois facteurs, nous avons utilisé les valeurs de quatre années : 1993, 1995, 2000 et 2007 telles que présentées dans le tableau ci-après. Arrêtons-nous un instant sur les données présentes dans ce tableau. On y retrouve évidemment les années et les trois facteurs jugés discriminants dans l'étape précédente vis-à-vis

de l'impact potentiel d'une catastrophe. Mais on y trouve également le « nombre de victimes enregistrées pour une catastrophe ». Cette dernière information présente les valeurs  $y$  de l'équation que nous cherchons à obtenir. Il faut bien comprendre ici que nous cherchons à disposer du nombre de victimes associé à une et une seule catastrophe pour pouvoir établir notre modèle de prévisions. Hors dans certains cas, cette information ne sera pas disponible sous cette forme (vision agrégée de l'ensemble des victimes de l'ensemble des catastrophes pour l'année par exemple). Ainsi, un travail préalable à cette étape devra consister à permettre l'obtention de cette information. Dans les cas les plus difficiles, il sera toujours possible d'intégrer à l'équation de régression à la fois les critères de vulnérabilité et résilience et le nombre d'occurrences associées (par exemple 10 tremblements de terre pour l'année considérée représentant un total de 3200 victimes). Une autre alternative consistera à considérer non pas le nombre victimes enregistrées pour 1 catastrophe mais le nombre total de victimes pour une année donnée. Il faut aussi noter que dans certains cas, les bases de données peuvent mélanger catastrophes récurrentes et catastrophes exceptionnelles. Dans ce cas il faudra être en mesure de faire le distinguo entre les catastrophes récurrentes et exceptionnelles. Ici encore si l'accès à ces données spécifiques n'est pas possible, une solution dégradée consistera à appliquer la méthode à l'ensemble des données et à assouplir les critères de validation du modèle.

Revenons maintenant à la mise en œuvre pratique de l'analyse de régression. Elle a été ici réalisée avec le logiciel Excel sur la base des données suivantes :

Année	IDH	NCD	VDC	Nombre des victimes
1993	0,53	7 884	69,0 %	1 440
1995	0,56	6 207	67,3 %	830
2000	0,58	8 200	68,0 %	2 406
2007	0,60	4 503	59,0 %	491
2012	0,62	6 862	55,0 %	2 566 (calculé) / 2 193 (constaté)

Table III-7 : Données et résultats obtenus sur le cas exemple Ancash.

Sur la base des données de 1993, 1995, 2000 et 2007, nous obtenons l'équation de régression multiple (MV) suivante :

$$Victimes = + 16826,995 * IDH + 0,686 * NCD - 2581,551 * VCS - 11168,97$$

Les deux variables expliquent 100,0 % de la variance des « Victimes ». Le coefficient de corrélation multiple :  $R = 1,00$ . Il est donc exceptionnellement bon (ce qui est logique compte tenu de l'extrême simplicité de ce cas). Afin d'évaluer la pertinence de notre modèle, nous avons

cherché à établir les prévisions de victimes<sup>1</sup> pour l'année 2012. Le calcul donne un nombre de victimes potentiel de 2 566 pour une catastrophe dans cette région qui aurait pu survenir en 2012.

#### Etape 4: Validation du modèle de prévision.

Pour vérifier l'ordre de grandeur de cette prévision, nous comparons simplement le résultat du calcul obtenu sur l'année 2012 avec la réalité des faits enregistrés sur cette année. Dans cet exemple, nous avons constaté que la survenue des catastrophes dans cette région avait conduit à 2 193 victimes. Le ratio de déviation par rapport aux 2 566 victimes calculés est alors de l'ordre de 17 %, ce qui semble correct et permet dans ce cas de valider le modèle avec les résultats du reste de régions du pays.

---

<sup>1</sup> Notons que l'impact va s'exprimer au travers de la quantité de victimes et non pas au travers de la quantité de personnes affectées :

- « victime » est définie par l'INDECI (2006) comme une personne touchée partiellement ou totalement par une situation d'urgence ou de catastrophe, qui a subi des blessures, des dommages sur sa santé ou sur sa propriété. Dans ce dernier cas, elle est généralement sans logement de façon permanente ou temporaire et nécessite de l'aide humanitaire et un abri temporaire.
- « affecté » qualifie une personne, un animal, un territoire ou des infrastructures qui ont souffert des perturbations de leur environnement sous l'effet d'un phénomène d'origine humaine ou naturelle. Une personne affectée peut nécessiter un soutien immédiat pour éliminer ou réduire les causes de la perturbation et poursuivre son activité normale.

Par exemple dans l'application numérique qui nous intéresse ici, il est possible de remarquer des différences significatives entre « victimes » et « affectés ». A cause des chutes de grêle sur la région de Cusco en 2009 notamment, le nombre de personnes affectés s'est élevé à 84 089 (affectés en raison d'impact sur les cultures, la santé, économie, etc.), alors que seules 19 personnes ont été concernées par une intervention humanitaire cette année-là. Précisions que les personnes « affectées » seront normalement prises en charge par les pouvoirs publics concernés alors que les « victimes », elles, seront prises en charge directement par l'aide humanitaire.

## 2. Méthode de conception de scénarios humanitaires réalistes

Dans les deux premiers chapitres de ce manuscrit, nous avons expliqué l'intérêt, les enjeux et l'engouement existant depuis une vingtaine d'années dans le domaine de la recherche scientifique vis-à-vis de la gestion de crises, notamment humanitaires. Pour autant, force est de constater qu'il existe un écart très significatif entre les propositions scientifiques faites et leur mise en œuvre concrète sur le terrain. Les professionnels de la gestion des secours sont souvent très réservés sur la pertinence des solutions proposées par les scientifiques, notamment lorsqu'il s'agit d'approches quantitatives. Une des causes majeures de ce manque de confiance résulte du fait que les chercheurs utilisent pour concevoir et valider leurs propositions, des scénarios jugés trop simplistes ou pas suffisamment réalistes. C'est par exemple le cas de ces nombreux modèles de conception de réseaux logistiques qui émettent implicitement dans leurs scénarios que l'ensemble des routes et infrastructures demeure utilisable, même après la survenue d'une catastrophe. Hors l'expérience montre combien cette hypothèse est rarement recevable en situation de crise.

Un des enjeux actuels est donc de pouvoir bénéficier de scénarios plus fidèles à la réalité et plus à même de garantir une bonne représentativité des problèmes du terrain. Cette section s'intéresse à ce sujet en proposant d'abord un rapide état de l'art sur la notion de scénario en général, et dans le monde de la gestion de crise en particulier. Puis nous proposerons une méthode structurée d'élaboration de scénarios réalistes.

### 2.1. Revue de littérature.

#### 2.1.1. *Incertitude et scénario.*

Comes et al. (2014) indiquent que plusieurs approches existent dans la littérature vis-à-vis de la notion de scénario. On parle de scénario pour :

- Définir « une vision consistante et personnelle de ce que pourrait être le futur ». Il ne s'agit pas d'une prévision mais d'un état possible pour l'avenir. (Porter, 1985) ;
- Analyser et structurer un ensemble d'idées dans l'optique de supporter une prise de décision stratégique. (Moschandreas et Karuchit, 2002) ;
- Combiner un ensemble de paramètres. (Groves et Lempert, 2007) ;
- Examiner ce qui pourrait arriver si une stratégie donnée était mise en œuvre. (Van Notten et al. 2003).

Dans la suite, nous ne retiendrons que la première acception.

Frigg et Hartmann (2012) indiquent que la notion de scénario telle qu'utilisée dans un cadre scientifique revêt deux approches fondamentalement différentes. Un scénario peut d'abord être une représentation d'un sous-ensemble du monde (système cible). C'est ce qu'il est convenu d'appeler « modèle de données ». Un scénario peut également représenter une théorie ou un système d'aide à la décision capable d'interpréter des lois et/ou axiomes concernés. Ces scénarios sont généralement appelés « modèles de théorie ».

Les propositions relatives à des modèles de théorie sont nombreuses (voir Chapitre IV pour plus de détail sur ces sujets). La nature des travaux développe autant des approches qualitatives (études de cas, sociologie, anthropologie, etc.) que quantitatives (recherche opérationnelle, simulation de flux, etc.) (Altay et Green, 2006 : Galindo et Batta, 2013). Notre étude s'intéresse plus spécifiquement aux propositions quantitatives du type modèle analytique. Le pan de littérature concerné montre que de nombreux outils, méthodes et techniques ont été proposés ces dernières années. Bien qu'historiquement plutôt orientés sur des approches déterministes (Charles, 2010), les modèles qui sont aujourd'hui développés privilégient des approches stochastiques, robustes ou floues (Peres et al. 2012 : Galindo et Batta, 2013) afin de mieux appréhender l'incertitude inhérente à ce domaine d'application. Ces approches nécessitent toutes de disposer d'un modèle de données pour pouvoir fonctionner.

Malheureusement (Peres et al. 2012 : Galindo et Batta, 2013) démontrent que la quasi-totalité des contributions existantes ont été développées sur la base de modèles de données irréalistes et souvent très simplistes. Aussi, bien que théoriquement très pertinent, la plupart de ces contributions sont jugées inconsistantes par les professionnels du secteur et restent bien souvent lettre morte.

Pour éviter ce problème, l'idée consiste à établir un modèle de données qui soit plus robuste. Un modèle de données est une version corrigée, rectifiée, régentée, voire idéalisée ou simplifiée des données brutes que l'on est susceptible de collecter (Frigg et Hartmann, 2012). Evidemment, le problème prend une tout autre dimension dans le cadre humanitaire tant l'accès aux données est délicate et tant l'incertitude sur la donnée elle-même est élevée. Concernant l'incertitude, Bronner (1997) en distingue deux acceptions complémentaires :

- L'incertitude de résultats qui caractérise le manque de connaissance sur la survenue d'un événement futur sur lequel il est impossible d'avoir une influence. Le cas le plus typique est la survenue potentielle d'une catastrophe naturelle (tremblement de terre par exemple). Il est possible de savoir qu'une région sera touchée sans pour autant savoir exactement quand ou avec quelle ampleur ;
- L'incertitude d'état qui caractérise le risque qu'une partie du système considéré risque d'être impactée par la survenue d'un événement perturbateur - naturel ou anthropique (PAHO, 2000a) et donc de ne plus être en capacité de fonctionner de façon nominale. Le cas le plus typique ici concerne l'indisponibilité de telle ou telle infrastructure suite à la survenue d'une catastrophe naturelle.

Dans le cas des crises de nature humanitaire, notamment les crises d'origine naturelle, les deux niveaux d'incertitude sont évidemment à considérer (Huang et Shi, 2002). En effet, au-delà de l'incertitude de résultat portant sur l'occurrence d'une catastrophe et de son impact humain (voir section 1 du présent chapitre pour plus d'information), toutes les situations de gestion de crise comporte une incertitude d'état lié à la disponibilité des ressources et des infrastructures utiles aux opérations de secours. Comme le démontre (Long et Wood, 1995) ou (Cassidy, 2003), la gestion de crise s'effectue généralement dans le cadre d'une perturbation majeure des ressources et infrastructures disponibles : routes coupées, entrepôt inopérant, alimentation électrique

coupée, etc. En conséquence, la notion de scénario attachée au cas des opérations humanitaires se doit de couvrir ces deux types d'incertitude pour être pertinent.

### 2.1.2. Méthodes de définition de scénarios

Comes et al. (2014) rappellent qu'il existe trois grandes familles de méthodes de conception de scénarios : analytiques, discursives ou formelles.

#### 2.1.2.1. Méthodes analytiques

Il existe différentes catégories d'approches analytiques utilisant la simulation, les statistiques, etc. Parmi toutes ces catégories, les techniques qui adressent les cas stochastiques sont probablement les plus complètes et les plus pertinentes à l'égard à notre sujet d'étude. Nous allons nous intéresser un instant à ces approches. Pour bien comprendre la suite, nous considérons la formulation générique suivante d'un problème de conception en environnement stochastique :

$$\min_{x \in C} \int_{\Xi} f(\xi, x) P(d\xi) \quad (4)$$

Dans laquelle les paramètres aléatoires de dimension  $n$  sont définies sur un espace de probabilité  $(\Xi, \mathcal{F}, P)$ , où  $\mathcal{F}$  est le  $\sigma$ -champ de Borel sur  $\Xi$  et  $P$  représente la mesure de probabilité représentant l'incertitude du problème, et  $C \subseteq \mathbb{R}^n$  est l'ensemble (déterministe) des décisions possibles. Dans la pratique, l'ensemble des résultats possibles correspondant à  $P$  est souvent approchée par un nombre fini  $\xi_k$  ( $k = 1, \dots, K$ ) de scénarios de probabilités  $w^k$ , ce qui permet une approximation de l'intégrale exprimée dans l'équation (4) par une somme telle que définie ci-après :

$$\min_{x \in C} \int_{\Xi} f(\xi, x) P(d\xi) \approx \min_{x \in C} \sum_{k=1}^K w^k f(\xi^k, x) \quad (5)$$

Le problème de génération de scénarios dans un environnement stochastique vise donc à alimenter correctement l'ensemble des paramètres de l'équation (5). Dupačová et al. (2000), Kaut et Wallace (2003), Legrain (2011) et Chen et al. (2012) proposent des états de l'art sur les différentes méthodes existantes pour résoudre ce problème. Un premier ensemble de méthodes est basé sur le tirage des scénarios selon une loi de distribution (Legrain, 2011) :

- La plus connue est sans doute la méthode dite de Monte Carlo (MC). Cette méthode considère un ensemble de nombres pseudo-aléatoires (vecteurs) comme scénarios et un ensemble de poids uniformes  $w_1 = \dots = w_K = 1/K$  ;
- Une autre méthode, développée dans (Pennanen et Koivu, 2002), est celle dite de Quasi Monte Carlo (QMC), dans laquelle des scénarios de poids uniformes sont générés en utilisant des séquences telles que la séquence de Faure, la séquence de Sobol ou la séquence de Niederreiter ;
- Certains auteurs comme Loretan (1997) se servent d'une Analyse en Composantes Principales (ACP) pour réduire les paramètres décrivant la corrélation entre les états du processus. Puis ils utilisent ces vecteurs pour générer des arbres de scénarios corrélés pour ce processus stochastique ;

- D'autres auteurs comme Cario et Nelson (1997) utilisent les mêmes méthodes, mais au lieu de corrélérer les scénarios avec les composantes principales, ils prennent en compte directement la distribution marginale ou la matrice de corrélation ;
- Un nombre important d'auteurs développe des techniques comparables aux précédentes. Il se trouve qu'il est très souvent difficile d'obtenir les distributions marginales. Il est beaucoup plus facile d'utiliser les moments de la distribution. Vale et Maurelli (1983) se servent des quatre premiers moments pour générer des arbres simulant le score de tests psychologiques. Smith (1993), Høyland et al. (2003) ou plus récemment Mehrotra et Papp (2013) illustrent d'autres méthodes et exemples utilisant les moments pour choisir les scénarios ;
- Dupačová et al. (2000) enfin montrent comment générer des arbres de scénarios à partir d'un ensemble de scénarios, des chemins donnés.

Un deuxième point sensible de la génération de scénarios est la réduction de la taille des ensembles de scénarios, des arbres de scénarios. Cet aspect est aussi très important, car la majorité des modèles d'optimisation stochastique ne peut être résolue qu'avec un nombre restreint de scénarios (Legrain, 2011). En effet, le temps de résolution sera beaucoup trop grand si la taille n'est pas correctement dimensionnée. Selon (Consigli et al. 2000 : Heitsch et Römis, 2007), le principe consiste généralement à générer aléatoirement un grand nombre de scénarios et à sélectionner un « sous-ensemble limitée de scénarios satisfaisants » parmi eux. Pour ce faire, différentes heuristiques et techniques d'exploration de données existent. Dempster et Thompson (1999) utilisent par exemple une heuristique basée sur la valeur dite de l'information parfaite. Dupačová et al. (2003) de leur côté utilisent une approche dite de discrétisation optimale. Pflug (2001) propose quant à lui de choisir les scénarios en minimisant une fonction erreur. Enfin, Casey et Sen (2005) développent une analyse de sensibilité sur une approche de type programmation linéaire pour guider la génération de scénarios.

Plusieurs auteurs se sont intéressés à la mesure de la qualité de chacune de ces méthodes. Ainsi (King et Wets, 1991 : Donohue, 1996) ont étudié l'épi-convergence de la méthode de MC alors que (Pennanen et Koivu, 2005 : Pennanen, 2005) se sont intéressés à celle de la méthode de QMC. Kaut et Wallace (2003) quant à eux ont comparé différentes méthodes deux à deux. Loin d'aboutir à des conclusions absolues, ces études montrent que selon les cas et les données manipulées, telle ou telle approche est plus ou moins satisfaisante.

#### 2.1.2.2. Méthodes discursives.

Les approches discursives, au contraire des précédentes, sont adaptées aux situations fortement incertaines, pour lesquelles peu d'information formalisée et structurée est disponible (Comes et al. 2014). Les approches discursives s'appuient en grande partie sur la connaissance d'experts. Comes et al. (2014) expliquent que dans ces écoles de pensée, les scénarios sont typiquement créés à partir de rien en appliquant des techniques créatives. Ceci peut être un problème quand le temps disponible est court ou quand la disponibilité des experts est limitée. De plus, le manque de formalisation limite la comparabilité des scénarios et rend complexe et coûteuse les mises à jour.

### 2.1.2.3. Méthodes formelles.

Les méthodes formelles se placent à l'intersection des deux familles précédentes. Le principe consiste à établir pour chaque variable considérée, une liste de valeurs et à générer des scénarios de façon aléatoire par combinaison entre toutes ces valeurs. Des méthodes formelles telles que La Prospective (Godet, 2000) ou une combinaison de la méthode Delphi et de la méthode d'analyse d'impact croisé (Bañuls et Turrof, 2011) sont des exemples de méthode formelle. Le risque inhérent à ce genre d'approche est sans nul doute la génération de scénarios inconsistants qui viendraient « noyer » l'utilisateur (Comes et al. 2014). Des traitements ex-post peuvent alors être mis en œuvre pour limiter cet écueil.

### 2.1.3. Définition de scénarios dans le monde humanitaire.

Si des méthodes analytiques existent bel et bien pour définir des scénarios, elles ne doivent pas être considérées comme des outils miracles. Leur qualité dépend avant toute chose, de la pertinence et de l'exhaustivité des données initiales qui sont considérées. Aussi, dans un contexte humanitaire où les données sont généralement peu abondantes et/ou peu fiables (Charles, 2010), la mise en œuvre de telles méthodes peut apparaître comme relativement délicate. De fait, une grande majorité des travaux existants s'est appuyée sur des méthodes discursives ou des approches analytiques limitées en termes de périmètre et de combinatoire.

En outre, force est de constater que très peu de travaux adressant la question humanitaire proposent aujourd'hui des approches basées sur la génération d'ensemble de scénarios réalistes. La plupart du temps, les sujets s'appuient sur des (Charles, 2010 : Peres et al. 2012 : Galindo et Batta, 2013) :

- Cas d'étude (souvent unique) de crises passées, souvent établies avec une approche essentiellement qualitative ;
- Jeux de tests numériques simplistes (limités en taille, incluant de nombreuses hypothèses simplificatrices, etc.) permettant de valider les principes généraux et les éventuels mécanismes de calcul de la proposition mais en aucun cas de les confronter à la complexité du terrain.

### 2.1.4. Bilan.

L'approche par scénarios pose deux problèmes (Snyder, 2006). Le premier est de définir les scénarios et d'assigner des probabilités fiables. Le second est d'identifier un nombre suffisamment restreint de scénarios pour des raisons de calcul. Nous retenons de cet état de l'art que la génération de scénarios s'exécute selon deux temps complémentaires :

- La génération selon une loi de distribution d'un large ensemble de scénarios répondant aux caractéristiques recherchées dans le cadre du problème étudiée (dans notre cas, les deux types d'incertitudes évoquées dans la section précédente devront être adressés) ;
- La sélection d'un sous-ensemble représentatif et couvrant parmi tous ces scénarios afin de répondre à la question tout en maintenant des temps de calcul raisonnable.

Par ailleurs, la plupart des travaux de recherche actuels (en logistique humanitaire, et plus généralement en aide à la décision pour l'humanitaire) s'appuient sur des scénarios qualitatifs

(approche discursive), simplistes ou irréalistes (approches analytiques ou formelles). Les modèles et solutions qui en résultent sont donc soumis aux doutes des utilisateurs finaux qui de facto, ne les utilisent pas. Pour permettre de développer des solutions plus concrètes et plus robustes, il convient de développer des scénarios qui soient plus représentatifs de la réalité des faits d'une part, et qui soient compatibles avec les exigences des modèles numériques d'autre part.

## 2.2. Contribution n°2 : méthode d'élaboration de scénarios réalistes.

### 2.2.1. Principes généraux.

En cohérence avec l'état de l'art que nous venons de développer, nous proposons ici de structurer une démarche de génération de scénarios dédiée à l'alimentation de modèles d'aide à la décision de nature quantitative tels que programmation linéaire. En particulier notre approche concerne les modèles de nature stochastique capables d'appréhender véritablement la dimension incertaine du secteur humanitaire. Le cadre de la proposition se limite ici encore aux catastrophes de type récurrent.

Un des enjeux de notre approche est de pouvoir adresser la question de l'incertitude selon ses deux dimensions complémentaires « résultat » et « état ». Dans notre contexte, il est possible de considérer que l'incertitude de résultat s'intéresse à la notion de victimes engendrées par la catastrophe (et donc de matériel nécessaire aux secours), alors que l'incertitude d'état s'intéresse à la notion de capacité résiduelle du réseau logistique susceptible d'être employé pour répondre à la crise. Un autre enjeu concerne la capacité à générer un ensemble complet et réaliste de scénarios. Cet ensemble doit à la fois être couvrant vis-à-vis des problématiques étudiées, et suffisamment restreints en taille pour être traité par des approches analytiques. Une question centrale porte notamment sur la probabilité d'occurrence que l'on doit associer à tel ou tel scénario pour rester représentatif et crédible au regard des professionnels du secteur.

De ces éléments, nous avons tiré une démarche articulée autour de 5 étapes majeures (voir figure suivante) :

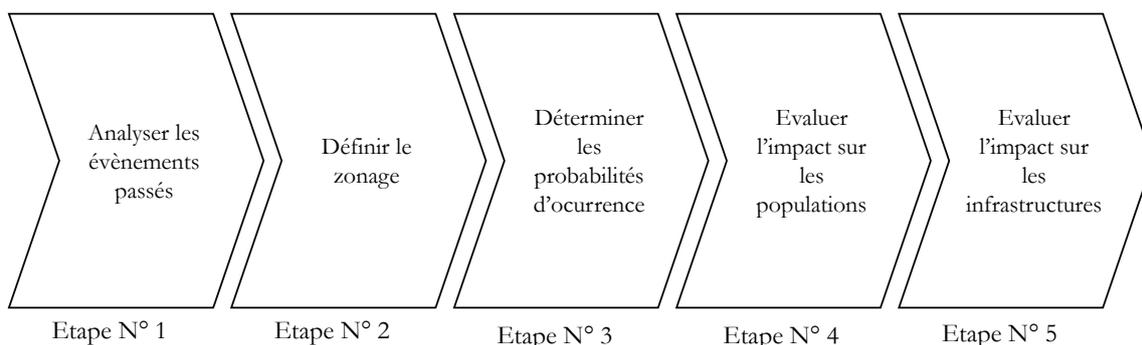


Figure III-8 : Méthodologie pour définir des scénarios réalistes.

#### 2.2.1.1. Etape n°1 : Analyser les événements passés

Cette étape consiste à analyser les caractéristiques des catastrophes (localisation et intensité) afin d'en comprendre les principaux mécanismes.

L'objectif de cette étape est d'objectiver la situation étudiée. Il s'agit de collecter un maximum d'informations tangibles (i.e. quantitatives et fiables) sur les événements passés :

- Date ;
- Localisation ;
- Typologie de phénomène enregistré ;
- Intensité et durée du phénomène.

Compte tenu du caractère récurrent des catastrophes étudiées, on prendra garde à considérer un intervalle de temps suffisamment grand pour être représentatif. Cet intervalle comprend généralement un minimum de dix ans, étant entendu que plus l'intervalle sera grand, plus la génération de scénarios qui en résultera sera pertinente. On pourra utiliser des bases de données propres au pays concerné ou bien exploiter des bases plus généralistes telles que OFDA/ CRED (EM-DAT, <http://www.emdat.be/database>).

#### 2.2.1.2. Etape n°2 : Définir le zonage

L'objectif de cette étape consiste à proposer un découpage géographique du territoire concerné qui soit cohérent avec les phénomènes naturels constatés. Pour ce faire, il faut utiliser les informations sur les caractéristiques sociodémographiques, géomorphologiques ou climatologiques des territoires. Un pays régulièrement soumis à des tremblements de terre par exemple, le sera avec des fréquences d'apparition différentes selon la localisation géographique considérée. Les futurs scénarios à définir devront donc être organisés selon ce découpage.

Au-delà de ce zonage, cette étape doit permettre de caractériser les lois de propagation des phénomènes naturels qui existent entre les zones identifiées. Une éruption volcanique par exemple (voir la figure suivante) aura tendance à se propager dans des directions bien déterminées correspondant aux caractéristiques géologiques de l'environnement. L'analyse des données collectées lors de l'étape 1 doit permettre de relever les liens de cause à effet qui existe entre les différentes zones. Ces liens sont plus ou moins complexes selon les phénomènes naturels. Dans les cas les plus simples, une matrice de corrélation pourra être construite. Sur les cas plus complexes, des fonctions spécifiques seront à établir avec les experts métiers concernés.

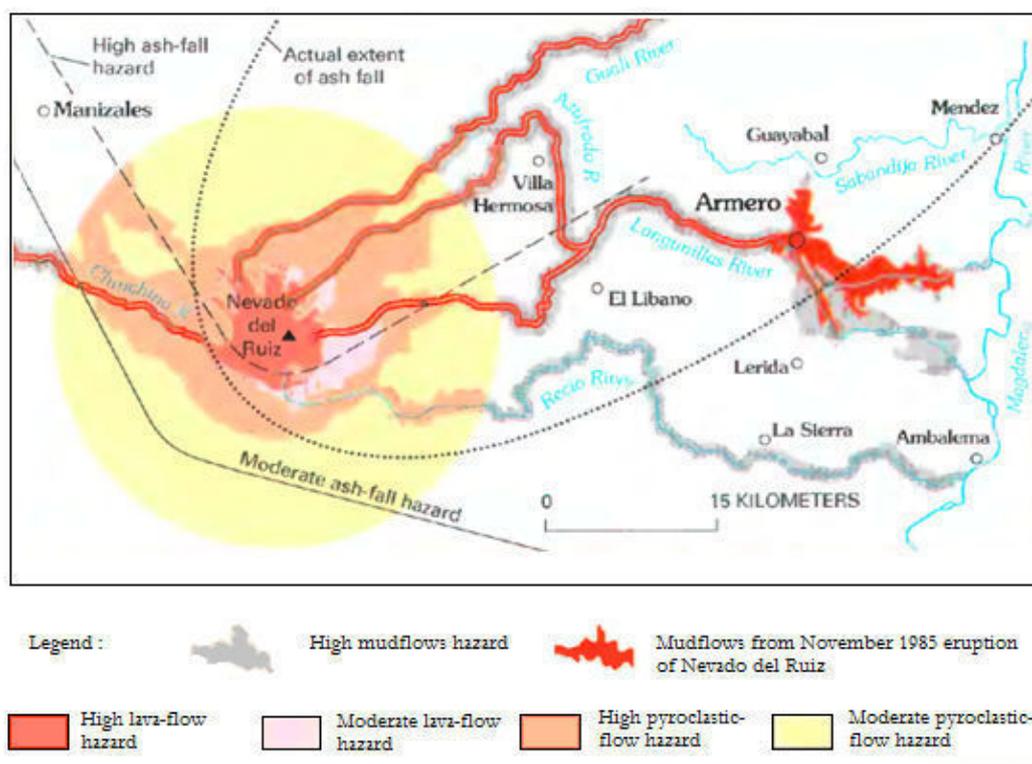


Figure III-9 : Exemple de propagation d'une éruption volcanique.

Source : <http://thewatchers.adorraeli.com/2011/09/13/unrest-at-nevado-del-ruiz-volcano-in-colombia/>.

### 2.2.1.3. Etape n°3 : Déterminer les probabilités d'occurrence

L'objectif de cette étape est de construire une liste exhaustive de scénarios plausibles, et d'associer pour chacun d'entre eux une probabilité de survenu. Cette étape s'appuie sur l'hypothèse qu'il y a une valeur quasi-périodique de l'occurrence de la catastrophe par périodes fixes de temps. Cette hypothèse n'est valide que parce que les crises concernées sont de nature récurrente (au sens de ce qui a été défini au Chapitre II).

En pratique, nous utilisons les informations collectées lors de l'étape 1. Pour déterminer où l'épicentre de la catastrophe sera localisé, nous étudions (sur un intervalle de temps donné) le nombre d'occurrences qu'une région a enregistré. Puis nous affectons à chaque région du territoire étudié un poids relatif (pourcentage) dans la survenue du phénomène.

Le phénomène déclencheur d'une catastrophe étant défini à la fois par son occurrence et par son intensité, il convient donc d'ajouter une information caractérisant l'intensité moyenne de la survenue de catastrophe dans telle ou telle région. Pour garantir une bonne représentativité des scénarios résultant, les cas extrêmes doivent être retirés du panel étudié. Ainsi, les événements passés d'intensité trop faible ou trop forte ne seront pas considérés dans le calcul de la moyenne. Ces limites sont à définir par des experts des phénomènes concernés. On notera également que des approches un peu plus riches peuvent être envisagées à ce stade comme par exemple le fait

d'associer à chaque scénario une loi triangulaire comprenant les valeurs d'intensité moyenne, minimale et maximale. Ce cas de figure pourra par exemple être très adapté si les modèles utilisés par la suite sont des modèles de type flou. Pour permettre une utilisation efficace des scénarios, les valeurs manipulées à ce niveau pourront être traitées par intervalle. Ainsi, pour les tremblements de terre par exemple, les experts découpent généralement les événements en 5 classes distinctes : magnitude en dessous de 5,5 : entre 5,5 et 6,0 : entre 6,0 et 7,0 : entre 7,0 et 8,0 : et au-dessus de 8,0.

Au final, le nombre de scénarios généré pour une même région peut varier de 0 à n, n étant égal au nombre d'intervalles d'intensité retenu. C'est cette approche combinée d'un découpage géographique raisonné (étape 2) et d'une définition de scénarios par intervalle (étape 3) qui permettra de limiter le nombre de scénarios au strict nécessaire tout en garantissant la représentativité et l'exhaustivité. Nous suggérons ici de vérifier que l'ensemble des scénarios représente un minimum 75 % des données enregistrées lors de l'étape 1.

#### 2.2.1.4. Etape n°4 : Evaluer l'impact sur les populations

Cette étape se concentre sur l'incertitude de résultat et met en œuvre la contribution n°1 sur le « modèle de prévision de la demande pour des catastrophes récurrentes » présentée plutôt dans ce chapitre. De cette étape, résulte une estimation de la quantité de victimes (et donc des produits et services nécessaires pour leur venir en aide) associées à chaque scénario. L'estimation est faite pour un événement donné et une région donnée.

#### Etape n°5 : Evaluer l'impact sur les infrastructures

Cette dernière étape adresse le sujet de l'incertitude d'état en proposant une estimation des conséquences sur les infrastructures liées à la survenue d'une catastrophe. Pour ce faire cette étape exécute trois temps successifs.

D'abord, il s'agit de collecter l'ensemble des informations relatives aux infrastructures disponibles. Cette information doit comprendre à minima le relevé des capacités des liaisons transport existantes entre chaque région. De la même façon, un recensement des capacités d'entreposage disponibles dans chaque région pourra être réalisé.

Ensuite, il faut déterminer comment un événement arrivant dans une région donnée se répercute sur les autres régions. On peut par exemple, utiliser un arbre de propagation décrivant les liens de cause à effet entre plusieurs régions. Dans la figure suivante, on considère un tremblement de terre de magnitude 7,5 dans la région R1 qui aura pour conséquence de diminuer les capacités d'entreposage de 60 % dans la région R2 si elle est en zone sensible (ZS, i.e. à forte propagation des ondes) ou 40 % si elle est en zone non sensible (ZNS, i.e. à faible propagation des ondes). En suivant le même raisonnement, la diminution sera de 5 % ou 15 % dans la région R3 selon que la région sera ZNS ou ZS. Un raisonnement similaire pourra être conduit sur les capacités routières.

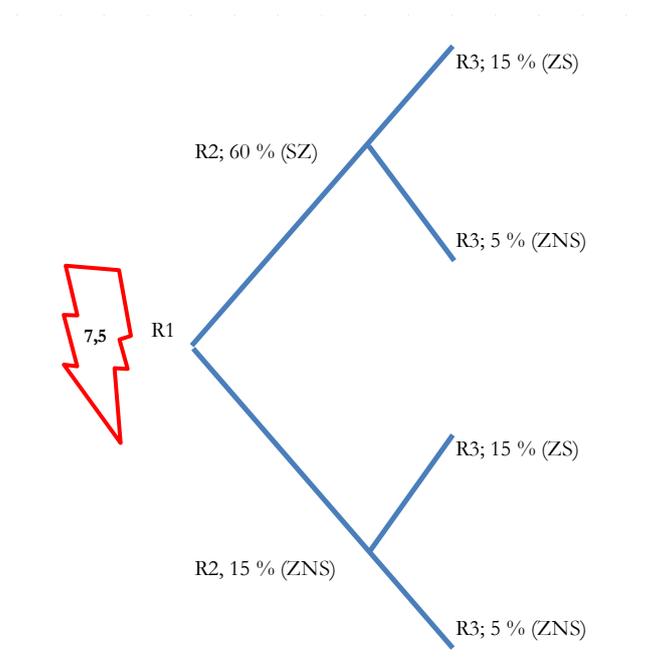


Figure III-10 : Exemple d'arbre de propagation.

L'instanciation des différents scénarios issus des étapes précédentes sur cet arbre de propagation constitue l'ultime temps de cette cinquième étape.

On remarquera que l'arbre de propagation inclut toutes les options, y compris celles qui n'ont pas de réalité physique. Dans cet exemple, une région est forcément sensible ou non sensible mais ne peut-être les deux à la fois. Pour autant, l'approche consiste à envisager l'ensemble des branches de façon à ne pas construire les scénarios « manuellement », ce qui pourrait potentiellement être très fastidieux. A l'évidence, l'exploitation des scénarios résultant fera que telle ou telle branche ne sera jamais parcouru et n'aura donc pas d'incidence sur le résultat final. Néanmoins, d'autres approches auraient pu être envisagées ici pour ne construire dès cette étape que les branches ayant un sens physique établi.

A l'issue de cette étape, un ensemble complet de scénarios est disponible présentant des garanties en termes de :

- Réalisme : basée sur une analyse factuelle des phénomènes passés ;
- Couverture : estimation des impacts humains et matériels ;
- Complétude : génération d'un ensemble de scénarios couvrant au moins les trois-quarts des phénomènes attendus.