



La constitution d'une base de données
"Témoignages sur événements"
et un exemple d'exploitation : l'affaire AZF

Conférence présentée devant
L'Institut de Recherche Criminelle de la Gendarmerie Nationale
(IRCGN)

par la

Société de Calcul Mathématique SA

2 juin 2010

en application de l'accord cadre de collaboration IRCGN/SCM du 15/02/2010

Résumé

Dans une première partie, nous présentons les idées fondamentales qui permettent la construction d'une base de données "témoignages", de manière générique.

Une telle base de données est destinée à être construite et alimentée en toutes circonstances où un certain nombre de témoignages sont recueillis : attentat, explosion, phénomène naturel, etc.

Elle repose sur deux principes fondamentaux :

- Elle doit être constituée indépendamment de toute tentative d'explication des phénomènes ;
- Elle doit être constituée indépendamment de toute autre source d'information (enregistrements par des appareils, mesures physiques, etc.).

Une fois la BD constituée, il y aura donc un travail de recoupement éventuel avec les autres sources d'information, mais il ne faut pas les mélanger au départ.

Disposer de principes généraux permet en particulier de "noter" une BD existante : les enregistrements faits sont-ils pertinents ? Sont-ils de bonne qualité ? Permettent-ils une conclusion ? Cela permet aussi de préparer le travail des enquêteurs dans l'avenir : leur fournir une liste de questions à poser, de formulaires à remplir.

Notre approche représente ici un travail préliminaire à la définition de "bonnes pratiques normalisées" pour la constitution d'une BD "témoignages".

D'un point de vue pratique, la construction de la BD se fait en deux temps :

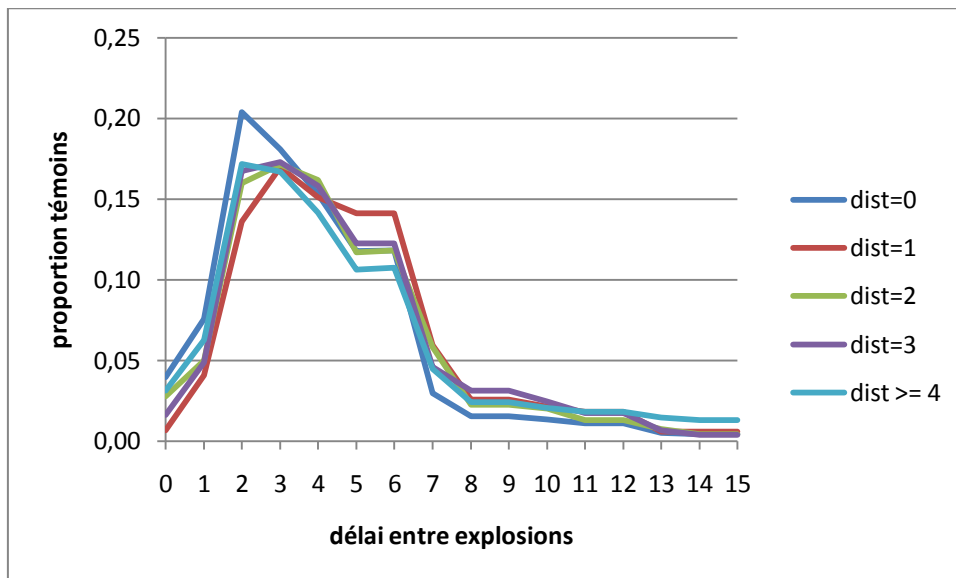
- Une première BD permet de coder les informations en langage humain simplifié, qui est un simple résumé de chaque témoignage ;
- Une seconde BD réalise un codage informatisé : pour chacun des cinq sens, on caractérise un "épisode" par son intensité (graduée de 0 à 3) et par sa durée (également de 0 à 3) et on tient compte également des éventuels laps de temps qui s'écoulent entre épisodes.

Dans une seconde partie, nous construisons cette BD dans le cas particulier de l'explosion AZF.

Dans une troisième partie, nous exploitons cette BD, en particulier en ce qui concerne les témoignages acoustiques : 456 témoins sur 807 rapportent avoir entendu deux explosions distinctes.

Nous introduisons des méthodes probabilistes robustes qui rendent compte de l'imprécision associée à chaque témoignage. Plus précisément, si un témoin annonce une durée d entre deux événements, nous admettons que cela peut signifier n'importe quoi entre $d/2$ et $3d/2$ (d'autres variantes sont possibles, que nous examinons aussi).

Notre conclusion est très claire : la proportion de témoins qui annoncent une durée fixée est indépendante de la distance. Cette conclusion subsiste pour toutes les variantes que nous avons prises quant aux hypothèses ; elle subsiste aussi si on retire les témoins ayant perçu une vibration (bruit sismique).



La seule explication possible est qu'il y a eu deux explosions distinctes. En effet, une explosion unique produit effectivement un bruit sismique, mais l'onde sonore et l'onde sismique ont des vitesses très différentes, et le délai entre les deux variera nécessairement avec la distance. Ce n'est pas le cas ici, et l'on peut conclure que l'on est en présence de deux explosions distinctes, séparées par environ quatre secondes.

Première Partie : Principes généraux pour la constitution d'une base de données "témoignages"

I. Une approche scientifique

L'idée dont nous partons est que la base de données que nous allons constituer doit être purement factuelle : elle doit recenser les informations apportées par les témoins. Celles-ci doivent être présentées "in abstracto", et non pas (comme on le voit souvent) à la lumière d'une explication prédéfinie, qu'il s'agit de valider.

Non seulement la BD doit être constituée sans idée préliminaire, mais encore elle ne doit incorporer aucun fait extérieur : elle se limite aux informations liées aux témoignages. Les autres informations recueillies (par exemple par des appareils) sont insérées dans une autre BD et un traitement ultérieur tient compte des deux.

Nous aurons dans un deuxième temps une approche probabiliste : nous considérons en effet que les déclarations des témoins ne peuvent être considérées comme absolument précises. Si un témoin dit "j'ai vu telle chose au bout de 10 secondes", cela peut être 5 secondes, cela peut être 15. Le témoin n'a pas un chronomètre en main ; sa perception est imprécise, tout comme sa mémoire.

II. Préliminaires sur les témoignages

A. *Un accident*

Supposons un événement quelconque. Il ne faut pas se figurer nécessairement une explosion, qui est ponctuelle, quasi-instantanée : ce peut être un événement ayant une certaine durée, comme par exemple la rupture d'un barrage, la dérive d'un navire. Un certain nombre de gens ont vu la scène, ou, plus exactement, ont eu une perception de la scène, pas nécessairement visuelle. Il s'agit de retranscrire leurs témoignages.

Nous parlerons de manière générique d' "accident", pour décrire l'événement, dans le sens commun "Événement imprévu et soudain qui entraîne des dégâts, des dangers" (dictionnaire). Ce peut parfaitement être un attentat : nous ne recherchons pas les causes, naturelles ou artificielles.

B. *Témoins et observateurs*

Comme nous voulons constituer une BD "témoignages", il faut commencer par se demander ce qu'est un témoin : "Personne qui assiste à un événement, un fait, et le perçoit" (dictionnaire). Mais nous ferons une différence essentielle entre "témoin" et "observateur" :

- Pour nous, un témoin n'a pas été prévenu qu'il devrait observer quelque chose ;
- Un observateur, au contraire, a la consigne d'observer et il est formé pour cela.

Par exemple, un contrôleur de la navigation aérienne observe l'espace aérien : il est formé pour cela. De même, un pompier qui surveille une région et détecte les incendies à partir d'une plate-forme.

La fiabilité que l'on peut accorder aux témoignages, dans chaque cas, n'est pas la même, et ceci se traduira par des lois de probabilité différentes : plus concentrées pour les observateurs, plus diffuses pour les témoins ordinaires.

Chaque témoignage étant nécessairement imprécis, imparfait, notre idée fondamentale est qu'il est inutile d'essayer de faire préciser quoi que ce soit aux témoins. Leur demander "l'événement B s'est-il produit 5 secondes ou 6 secondes après l'événement A, 3 minutes ou 4 minutes avant l'événement C ?" n'a pas de sens. Il faut s'accommoder de cette imprécision : elle sera incorporée sous forme d'une loi de probabilité ; ceci sera décrit plus loin.

C. La notion d' "épisode"

L'être humain ne disposant que de cinq sens, il s'agit simplement de caractériser les informations relatives à chacun d'eux. Une situation d'information s'appellera un "épisode", dont on cherchera à caractériser le moment, l'intensité et la durée. Si par exemple un témoin dit : j'ai vu un éclair intense, puis une minute après une lueur diffuse, nous dirons que nous sommes en présence de deux "épisodes visuels", le premier intense, le second faible, séparés par une minute. Le premier épisode avait une durée quasi-instantanée, et le témoin n'a rien dit quant à la durée du second.

Si le témoin dit : j'ai senti une odeur, puis dix minutes après j'ai vu un éclair, nous avons un épisode olfactif, suivi d'un épisode visuel.

III. Construction de la base de données témoignages

A. Organisation de la BD

La base de données sera donc présentée sous la forme d'un fichier Excel organisé comme suit :

1. En première colonne, la liste des témoins

Il est préférable que cette liste soit anonymisée (les noms sont remplacés par des numéros) ; cela évite toute difficulté avec la loi "Informatique et Libertés" et toute difficulté avec les personnes elles-mêmes.

Un problème se pose immédiatement : celui de la crédibilité d'un témoin. On est tenté de juger différemment, selon l'âge, la profession, les antécédents, etc. Mais rien ne dit, en vérité, que quelqu'un de 20 ans soit plus crédible que quelqu'un de 40, ou l'inverse. Il se peut, évidemment, qu'un témoin soit myope et refuse de le dire : si on s'en aperçoit, il faut supprimer son témoignage, et sinon, cela rentre dans les incertitudes générales.

Nous mettrons dans la base de données un certain nombre de critères explicites (distance, par exemple) qui seront détaillés plus loin. La crédibilité d'un témoin dépendra évidemment de ces facteurs, mais uniquement de ceux-là. Nous faisons l'hypothèse explicite que, si tous les paramètres figurant dans la BD ont la même valeur, les témoins ont la même crédibilité.

2. Seconde colonne : la distance à l'événement

S'il s'agit d'un événement unique, distance à cet événement ; si l'événement se déplace, distance à la situation considérée comme initiale.

La distance n'étant généralement pas connue avec précision, il suffit généralement de classes, du type 0 – 100 m, 100 – 200, etc.

3. Troisième colonne : position par rapport à l'événement

L'événement étant pris comme origine des axes, il faut positionner les témoins sur un repère (x, y) (et quelquefois z) dans ce repère. Il n'est pas clair que les témoins sachent précisément où ils étaient ; quelquefois ils peuvent le dire par rapport au paysage, aux routes, etc., et cela suffit à le reconstituer de manière grossière. Le positionnement précis étant généralement inutile, nous nous contenterons de quatre zones dans l'exemple traité plus bas : NO, NE, SO, SE.

4. Quatrième colonne : situation (intérieur ou extérieur) par rapport à l'événement

Dans le cas d'une explosion, la perception du témoin (et l'intérêt de son témoignage) seront différents selon qu'il est situé à l'extérieur ou bien à l'intérieur de bâtiments. Dans un cadre plus général, il faut mettre ici toute indication permettant de savoir si le témoin avait ou non un accès direct à l'événement. Un témoin peut se trouver à proximité immédiate d'un événement, être tourné vers l'événement, et ne rien ressentir, tout simplement parce qu'un mur épais l'en empêche.

5. Cinquième colonne : orientation par rapport à l'événement

L'être humain n'a pas des sens isotropes ; nous entendons dans toutes les directions (mieux vers l'avant que vers l'arrière) et nous ne voyons que vers l'avant. Il est donc nécessaire de demander au témoin son orientation par rapport à l'événement. Nous nous contenterons de deux classes : 1 (vers l'événement), 0 (sens opposé). Cela n'a d'intérêt, en général, que pour les témoins à l'extérieur, puisque les témoins à l'intérieur n'ont pas une vue de l'événement.

6. Sixième colonne : situation d'attention du témoin

Elle sera caractérisée de manière binaire : 0 ou 1. "0" signifie que le témoin n'est pas en situation d'attention (il dort, il écoute de la musique, etc.) ; "1" signifie que le témoin est en situation normale d'attention.

Il ne faudrait pas croire que les seuls témoins intéressants sont ceux qui ont vu ou entendu quelque chose. Pour nous, un témoin qui n'a rien vu et rien entendu est également tout à fait intéressant, s'il était en situation de voir et d'entendre, parce que son témoignage tend à prouver que rien ne s'est produit. Les témoins négatifs ont exactement la même valeur que les témoins positifs.

A ce stade, nous avons enregistré de manière grossière l'ensemble des caractéristiques du témoin. Bien entendu, il est toujours possible de rajouter des colonnes, si des éléments particuliers sont souhaités et disponibles.

7. Septième colonne et au-delà

Les séquences d'épisodes, organisées selon chacun des cinq sens.

Il y a deux manières de procéder :

- Ou bien on collecte l'ensemble des événements perçus par l'ensemble des témoins, et on les organise selon les cinq sens, par ordre chronologique.

Mais ceci n'est pas possible en général, parce que les témoins ne sont pas unanimes. Pour les uns, A a précédé B, pour les autres c'est l'inverse. Certains ont perçu A, d'autres non, etc.

- Ou bien on travaille témoin par témoin, chacun étant indépendant des autres ; c'est ce que nous ferons.

On aura ainsi par exemple :

Témoin no 1 : épisode visuel 1, puis épisode olfactif, puis épisode visuel 2, etc.

Témoin no 2 : épisode auditif, puis épisode visuel 1, puis épisode visuel 2, etc.

Dans un tel cas, comme il est fréquent, les deux témoins n'ont pas perçu la même chose. Le second a entendu quelque chose en premier, que le premier témoin n'a pas entendu. Par contre, le premier témoin a senti quelque chose, que le second n'a pas perçu.

Bien entendu, l'exploitation est difficile. Il y aura un raisonnement probabiliste à mettre en œuvre pour dire si, oui ou non, le phénomène qualifié de "éclair" par le témoin no 37, à telle date approximative, est la même chose que celui qualifié de "lueur", par le témoin no 73, à telle autre date. Mais c'est bien ainsi que le problème se pose, et il ne faut pas faire semblant d'avoir diagnostiqué par avance un ensemble clair d'événements.

La base de données sera donc constituée de la façon suivante :

Témoin no 1 ; septième colonne et au-delà :

Episode 1 avec ses caractéristiques,
séparation en temps
puis épisode 2
séparation en temps
puis épisode 3, etc.

Ici, les caractéristiques sont :

l'indication du sens concerné (vue, ouïe, odorat, toucher, goût, sous la forme V, A, O, T, G)

son intensité (faible, moyen, fort, etc.)

sa durée

ses caractéristiques spécifiques (couleur, etc.), sous forme d'un mot-clef.

Puis une séparation en temps, qui est le laps de temps, perçu par le témoin, séparant deux épisodes.

Et ainsi de suite pour le témoin no 2, etc.

Par exemple, la séquence :

V 0 2 éclair ; ? ; A 0 3 explosion

signifie que le témoin a vu quelque chose, qualifié d' "éclair", instantané (durée 0 secondes), assez fort (classement 2 sur une échelle de 0 à 3). Puis, après un laps de temps non spécifié "?", il a entendu quelque chose, qualifié d'explosion, instantané, très forte.

B. Rien vu ou pas d'information

Il est important de faire la différence entre

- "n'a rien vu" ou "n'a rien ressenti" ;
- "ne sait pas" ou "la question n'a pas été posée" ou "l'information a été perdue".

Pour nous,

- Sait qu'il n'y a rien sera représenté par 0 ;
- Ne sait pas, ou pas d'information, sera représenté par ND ou "?".

Il faudra être vigilant lors du traitement informatique, car Excel considère une cellule vide comme ayant la valeur 0.

C. Remédier aux défauts habituellement constatés

La description de la BD que nous présentons ici tente de remédier aux défauts habituellement constatés :

- Souvent, il n'y a pas d'uniformité dans les informations provenant des divers témoignages : certains sont complets, d'autres non, tout simplement parce que les témoins ne présentent pas tous les choses de la même façon.

Ici nous disposons d'une "grille d'interrogation" : on voit quelles sont les questions à poser.

- Il y a souvent confusion entre "n'a rien vu" et "ne pouvait rien voir" : le champ est laissé blanc. Or pour nous, la distinction est essentielle : un témoin qui pouvait voir et qui n'a rien vu apporte une information de toute première importance.

La démarche entreprise ici permet également de noter une base de données existante : a-t-elle été correctement réalisée ? Est-elle exploitable ? La question est d'importance, si ensuite cette BD est utilisée dans le cadre d'une instruction judiciaire.

L'évaluation de la qualité d'une BD devrait être faite systématiquement, en préalable à toute exploitation.

Nous avons réalisé, en 2007, une démarche du même ordre, à la demande de l'Institut de Radioprotection et de Sécurité Nucléaire : il s'agissait d'expertiser les bases de données "post-Tchernobyl" (environnement, épidémiologie, sarcophage), du point de vue de la qualité des données qui y étaient conservées et de faire des recommandations pour la constitution de BD ultérieures.

Seconde Partie : constitution de la BD AZF

I. Données générales concernant l'accident AZF

Nous mettons maintenant en œuvre les principes généraux décrits plus haut dans le cas particulier de l'accident AZF.

Celui-ci s'est produit le 21 septembre 2001 sur le site de Toulouse. Les caractéristiques générales sont :

- Explosion(s) localisée(s) et ponctuelle(s) : on ne sait pas s'il y a eu une ou plusieurs explosions, mais dans tous les cas il ne s'agit pas d'un événement qui dure.
- Au moins l'une des explosions peut être localisée (présence d'un cratère) et datée (interruptions électriques enregistrées, sismographes).

Nous n'utiliserons pas cette datation et cette localisation pour la constitution de notre première BD témoignages, puisque aussi bien la datation que la localisation reposent sur des informations indépendantes des témoignages.

- On ne connaît pas l'origine de l'explosion (accident industriel, attentat, phénomène externe, du type arc électrique, etc.).

Environ 950 témoignages ont été recueillis. Quatre des cinq sens sont utilisés, à savoir :

- La vue V ;
- L'ouïe A ;
- L'odorat O ;
- Le toucher T.

La base de données devra donc comporter des champs correspondants aux épisodes relatifs à ces quatre sens. Tous ne sont pas également mentionnés par les témoins.

II. Difficultés rencontrées lors de l'exploitation des fichiers disponibles

La constitution d'une base de données à partir des fichiers d'information existants se heurte systématiquement à des difficultés (voir par exemple nos commentaires à propos des bases de données Tchernobyl). Il y a trois types de difficultés :

- Certaines données sont manquantes ;
- Certaines données sont erronées ;
- Certaines données sont contradictoires.

Le traitement des données, qui permet de passer des fichiers recueillis à la BD finale, doit faire l'objet d'un soin particulier (même en l'absence de procédure judiciaire).

En particulier, il ne faut pas procéder par copier-coller et, de manière générale, il ne faut pas manipuler les données à la main, car :

- des erreurs peuvent être commises (arrondis différents, erreurs de formules, etc.) ;
- toute manipulation doit laisser des traces, pour qu'on puisse la vérifier.

Notre méthode de travail consiste à écrire des macros VBA sous Excel ; ces macros permettent une vérification systématique et une compréhension de ce qui a été fait. Si le fichier est réutilisé cinq ans après, personne ne se souvient des copiers-collés qui ont été faits, alors que les macros VBA peuvent être relus indéfiniment.

Nous détectons ainsi que 8 témoins n'ont pas de coordonnées géographiques. Nous n'attribuons pas de coordonnées factices à ces témoins ; nous notons simplement que les coordonnées ne sont pas disponibles.

Nous détectons également les distances supérieures à 100 km (a priori, un témoin peut se trouver à une distance plus grande, mais cela exige justification). Nous détectons ainsi qu'un témoin se trouve à 1600 km : il s'agit évidemment d'une erreur de report dans la coordonnée en y .

Nous disposons en outre d'un tableau spécifique, Bruit-2-rev52-2.xls, qui regroupe les témoignages auditifs. Dans ce tableau, les coordonnées figurent, ainsi que les distances au cratère. Nous vérifions que ce tableau est bien issu du tableau général. En fait, nous détectons des anomalies pour 7 témoins et nous les corrigeons, en revenant au tableau général.

Pour la constitution de la BD proprement dite, nous procédons en deux étapes :

Une première BD, appelée SCM_AZF1.xlsx (format Excel 2007) contient les témoignages, avec les mots-clé employés par les témoins.

Ensuite, cette BD est transformée en une seconde BD, notée SCM_AZF2.xlsm (format Excel 2007 avec macros), qui code l'information sous une forme exploitable.

Troisième partie : Exploitation et Résultats

I. Introduction

La base ainsi constituée, comme expliqué dans la seconde partie, comporte 936 témoignages. Elle reçoit un traitement probabiliste, qui est détaillé plus loin. Il s'agit de mettre en œuvre une méthode robuste, permettant d'incorporer l'information provenant de tous les témoignages.

Au sein de la base de données, nous ne traitons que les éléments qui sont de nature statistique ; par exemple, un éclair (vu par plus de cent personnes) est pris en compte, tandis qu'une boule de feu (vue par huit personnes) ne l'est pas.

A priori, tous les témoignages sont pris en considération, mais si un témoin annonce (par exemple) un délai incompatible avec la physique du problème (comme un bruit parvenant une demi-heure après), il est éliminé ; aucun témoin n'est corrigé.

II. Conclusions

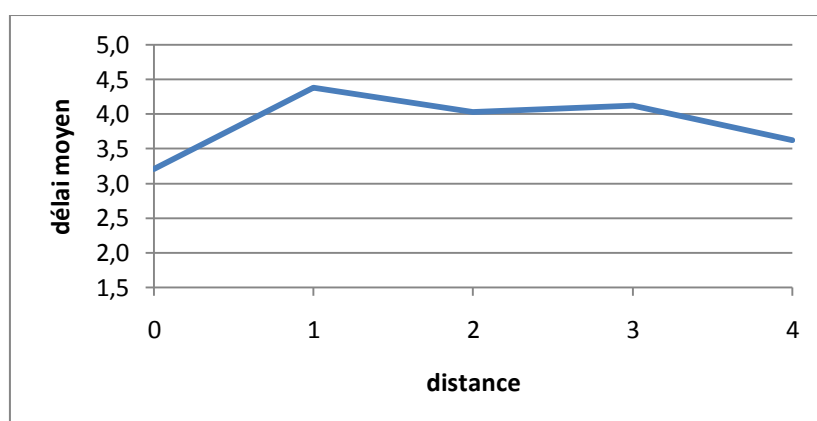
Nos conclusions sont les suivantes :

Il y a eu deux explosions distinctes : ce fait est reporté par 456 témoins sur 807.

La proportion de témoins ayant entendu deux explosions est la même dans les quatre secteurs géographiques autour de AZF (NE, NO, SE, SO). Il n'y a donc pas eu de réflexions parasites de l'onde sonore, pas de différences de propagation anisotropes, etc.

Le laps de temps entre ces deux explosions est le même, quelle que soit la distance à AZF.

Pour démontrer ceci, nous commençons par faire un traitement déterministe simple : la moyenne du laps de temps évalué par les témoins, en fonction de la distance :

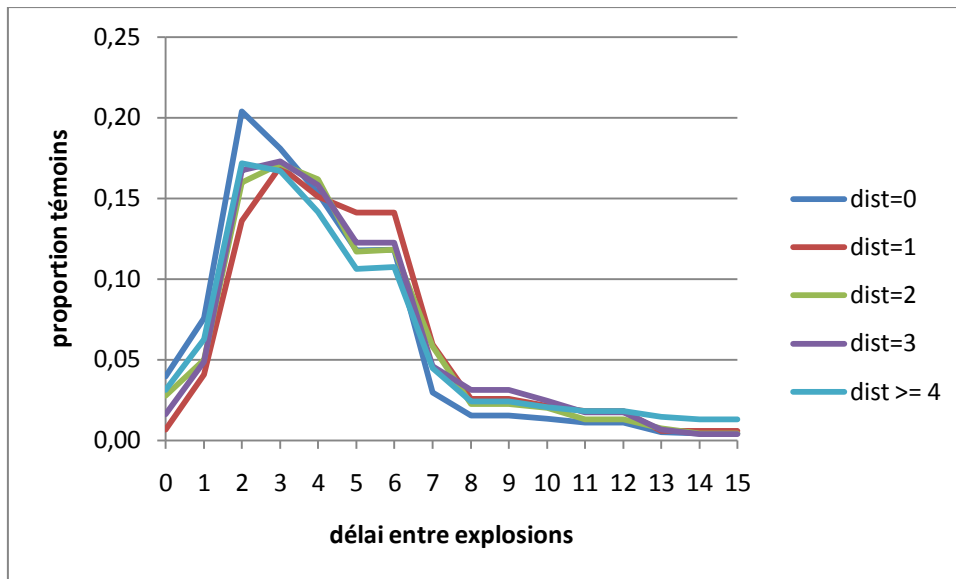


Graphie 1 : le délai moyen entre explosions, en fonction de la distance

Comme on le constate, ce laps de temps n'augmente pas avec la distance.

Ensuite, nous mettons en place une méthode probabiliste, qui permet de déterminer, pour chaque distance au lieu de l'explosion, le pourcentage de témoins qui ont perçu un délai donné.

Voici l'aspect de ces courbes :



Graph 2 : les lois de probabilité de la durée du délai entre explosions, en fonction de la distance

Ici, nous avons cinq courbes, l'une pour la distance 0 - 1000 m, la seconde pour 1000 - 2000, etc., et la dernière pour tous les témoins au-delà de 4 km. Chaque courbe représente la proportion de témoins (en ordonnée) ayant perçu le délai (en abscisse).

Ces courbes sont issues d'un traitement probabiliste : si un témoin assure avoir perçu un délai de x secondes, on considère qu'il a pu vouloir dire n'importe quoi entre $x/2$ et $3x/2$ secondes. Cette méthode, appliquée à l'ensemble des témoignages, permet de tenir compte des imprécisions normales quant à l'appréciation des durées.

Nous réalisons trois variantes : lorsqu'un témoin dit "il s'est écoulé quelques secondes", nous interprétons ceci comme 2 secondes, ou 4, ou 6, et refaisons les calculs sur chacune de ces bases. Les résultats sont les mêmes dans tous les cas : les courbes ont toujours la même forme.

Nous réalisons également deux variantes sur l'intervalle retenu autour de la durée annoncée par le témoin.

Enfin, nous réalisons une variante en éliminant les témoins qui rapportent un bruit sismique (vibration), car ce bruit peut être confondu avec celui d'une explosion.

La conclusion la plus importante ne porte pas tant sur la forme des courbes que sur le fait que toutes sont identiques : les témoins situés à des distances différentes ont eu la même perception.

De ce fait, il est exclu que les deux signaux sonores perçus par les témoins proviennent d'une même et unique source, avec des modes de propagation différents. Il ne peut s'agir, par exemple, d'une onde acoustique et de la résurgence d'une onde sismique, parce que l'onde sismique se propage beaucoup plus vite et que le délai entre les deux serait alors en augmentation avec la distance. Pour fixer les idées, rappelons ceci : l'onde sismique se propage à environ 5 km/s et l'onde sonore à environ 1/3 km/s ; pour un témoin situé à 2 km, la différence de perception serait de 5.5 s environ ; pour un témoin situé à 4 km, elle serait de 11 s environ : le délai augmente proportionnellement à la distance, de manière très visible.

Les deux explosions distinctes qui se sont produites sont séparées par environ 4 secondes.

Ces deux explosions se sont produites au même endroit ou en des endroits proches ; si ce n'était pas le cas, les délais perçus par les témoins devraient être différents dans différentes régions du plan. Or il n'y a pas de différence de perception significative entre des zones différentes. Nous montrons que l'hypothèse selon laquelle la première explosion se serait produite à plus de 340 m de la première ne peut être retenue, mais nous ne pouvons pas être plus précis.

La quasi-totalité des témoins qui ont évalué la force des deux explosions disent que la seconde a été nettement plus forte que la première.

L'onde sismique a bien été perçue par les témoins : 128 en font état, et la proportion augmente avec la distance. Les témoins disent que cette onde sismique a précédé l'explosion, et le délai augmente avec la distance, comme il est normal : délai moyen de 2.9 s pour les témoins situés à moins de 2 000 m et de 3.3 s pour les témoins situés à plus de 2 000 m. Pour nous, l'onde sismique correspond à un phénomène "tactile" (noté par T dans la base de données) ; c'est une vibration, qui ne se confond pas avec un bruit.

Dans le cas de deux explosions, un "observateur idéal" aurait dû percevoir quatre phénomènes : deux ondes sonores et deux ondes sismiques. Mais la première explosion semble avoir été faible, et l'onde sismique s'est atténuée très vite ; pour la seconde explosion, l'onde sismique a été détectée par certains ; elle a pu, pour d'autres, être confondue avec les vibrations provenant de la première explosion.

En résumé, un seul scénario se dégage, au terme de l'exploitation probabiliste de la BD "témoignages" : il y a eu deux explosions, à quelques secondes d'intervalle (environ 4 s), la première étant nettement plus faible que la seconde. Il y a également eu un éclair, confirmé par plus de cent trente témoignages.

III. Exploitation de la base de données

A. Statistiques globales

Nombre de témoignages : 931.

Nous avons utilisé les notations suivantes, pour certains témoignages :

NA : Non Actif. Les témoins étaient non-actifs lors de l'explosion, c'est-à-dire qu'ils ont été blessés, qu'ils se sont évanouis...

NC : Non Concordant. Les différents fichiers de témoignage sont contradictoires sur les faits.

ND : Non Détaillé. Les témoignages ne sont pas assez détaillés pour les coder correctement ; parce qu'ils ne concernent pas le moment même de l'explosion ou parce qu'ils sont insuffisants en informations sur l'incident.

NE : Non Entendu. Les témoins n'ont rien entendu, rien vu ou n'ont pas été assez marqués pour en avoir le souvenir.

NP : Non Présent. Les témoins n'étaient pas présents sur le site et ne peuvent donc pas nous aider quant au déroulement des faits.

NT : Non Trouvé. Nous n'avons pas trouvé les fichiers de témoignage correspondants.

D : décédé ;

Ces renseignements sont portés dans la colonne I de la BD.

Cinq personnes ont vécu les événements à bord d'un véhicule en mouvement ; nous avons choisi de mettre la position de début de la perception (100 km/h = 27 m/s ; cette vitesse est négligeable devant celle du son, a fortiori devant celle des ondes sismiques).

Voici la répartition pour chacun des codes ci-dessus :

NA	16
NC	4
ND	22
NE	17
NP	7
NT	6
D	21
vides	0
Valides	838
total	931

Tableau 3 : les différents codes relatifs aux témoignages

Il reste donc 838 témoignages utilisables.

Tous les témoins, sauf un, ont des coordonnées en X et en Y ; celles-ci ont fait l'objet de vérifications ; les modifications, portant sur une trentaine de témoins, ont été incorporées dans la BD.

A partir des coordonnées, on calcule la distance à AZF, qui est donc disponible pour tous les témoins sauf un ; on calcule également le secteur de localisation (NE, NO, SE, SO).

Voici la répartition par secteurs :

NE	168
NO	228
SE	200
SO	334
total	930

Tableau 4 : répartition par secteurs géographiques

Certains témoins mentionnent qu'ils se trouvaient à l'extérieur des bâtiments, d'autres qu'ils se trouvaient à l'intérieur. Voici la répartition :

extérieur	197
intérieur	657
Inconnu	77
total	931

Tableau 5 : répartition intérieur/extérieur

Certains témoins déclarent qu'ils étaient tournés vers AZF (qu'ils soient à l'intérieur, au travers d'une vitre, ou à l'extérieur, avec vue directe). Ces témoins sont au nombre de 41.

Parmi les témoins, la plupart étaient en situation d'attention :

Attention	823
Pas Attention	15
Autres	93
total	931

Tableau 6 : situation d'attention

B. Exploitation des cinq sens

Les témoins déclarent avoir perçu des phénomènes auditifs (bruits, sifflements, explosions), notés A, des phénomènes visuels (flash, éclair), notés V, des phénomènes tactiles (souffle ou vibrations), notés T. Certains indiquent avoir perçu une odeur (avant ou après le phénomène), notée O. Le goût n'est pas utilisé ici. Nous avons le tableau suivant :

Auditifs	815
Visuels	165
Tactiles	190
Odorat	7

Tableau 7 : répartition selon les cinq sens

Le nombre de témoins ayant perçu une odeur est trop faible pour permettre une exploitation statistique.

C. Exploitation des phénomènes auditifs

C'est évidemment la première source d'information, par le nombre des témoignages. En outre, ces témoignages concernent aussi bien des gens situés à l'intérieur qu'à l'extérieur, à faible ou à grande distance. L'ouïe est un sens plus disponible que la vue : plus permanent, moins directionnel.

1. Localisation

La plupart des témoins (57%) ont perçu deux explosions distinctes ; voici la répartition en fonction de la localisation :

	NE	NO	SE	SO	?	total
une seule explosion	62	79	83	126	1	351
deux ou plus	81	117	98	168	0	464
prop. 2 explosions %	56,64	59,69	54,14	57,14	0	56,93

Tableau 8 : perceptions auditives selon la zone

La plupart des témoins ayant perçu deux explosions disent que la seconde était plus forte que la première ; voici la répartition :

nb témoins ayant perçu deux explosions	464
nb témoins ayant perçu 2ème plus forte	322
nb témoins ayant perçu 2ème égale 1ère	47
nb témoins ayant perçu 2ème plus faible	11

Tableau 9 : comparaison des deux explosions

Voici la comparaison de la force des explosions en fonction de la localisation

	tous	NO	NE	SO	SE
nb témoins ayant perçu deux explosions	464	117	81	168	98
nb témoins ayant perçu 2ème plus forte	322	86	56	117	63
nb témoins ayant perçu 2ème égale 1ère	47	15	9	17	6
nb témoins ayant perçu 2ème plus faible	11	3	1	4	3
ne savent pas	84	13	15	30	26

Tableau 10 : comparaison des deux explosions en fonction du secteur

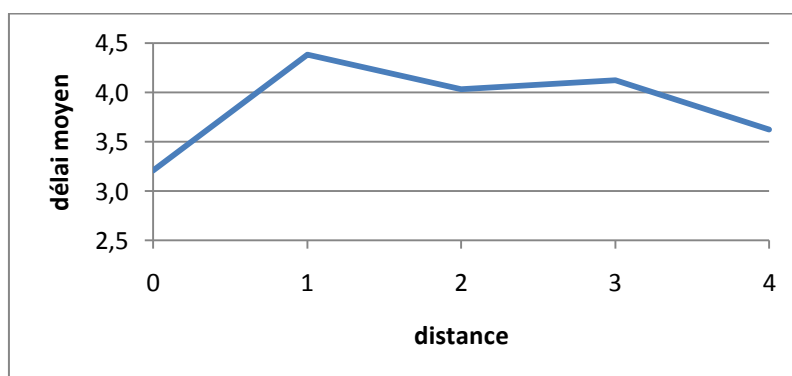
2. Répartition en fonction de la distance

Nous faisons d'abord un traitement simple, en ne tenant compte que des témoins qui annoncent effectivement une durée, entre les deux explosions, sans aucun phénomène entre ces deux explosions : ainsi, il n'y a pas de confusion possible sur l'interprétation de cette durée.

distance	nombre témoins	délai moyen
0	93	3,21
1	55	4,38
2	32	4,03
3	25	4,12
4	12	3,63

Tableau 11 : répartition des témoins explicites

Voici le graphique associé :



Graphique 12 : le délai en fonction de la distance

Ce délai n'augmente pas avec la distance. L'écart entre les deux explosions, calculé par moyenne sur tous les témoins, est de 3,7 secondes.

Nous introduisons des lois de probabilité, en considérant que si le témoin annonce un délai d entre deux explosions, ce peut être n'importe quoi (avec probabilité uniforme) entre $d/2$ et $3d/2$. Plus précisément, comme il s'agit de valeurs entières, voici le tableau que nous retenons :

d	d-d0	d+d0	largeur
0	0	0	1
1	0	2	3
2	1	3	3
3	2	4	3
4	2	6	5
5	3	7	5
6	3	9	7
7	4	10	7
8	4	12	9

Tableau 13 : le délai minimum et le délai maximum retenus, selon ce que dit le témoin

Ceci se comprend de la manière suivante : si le témoin dit d , ce peut être entre $d - d_0$ et $d + d_0$, avec probabilité uniforme entre ces bornes.

Nous commençons par prendre en compte uniquement les témoins qui chiffrent précisément le délai entre les deux explosions qu'ils ont entendues.

On obtient le tableau suivant :

temps	proba,dist=0	proba,dist=1	proba,dist=2	proba,dist=3	proba,dist=4
0	0,08	0,02	0,07	0,04	0,04
1	0,15	0,08	0,14	0,07	0,09
2	0,22	0,12	0,23	0,17	0,15
3	0,18	0,18	0,18	0,17	0,15
4	0,12	0,13	0,12	0,15	0,11
5	0,05	0,13	0,05	0,09	0,07
6	0,05	0,13	0,05	0,09	0,07
7	0,05	0,10	0,05	0,07	0,07
8	0,02	0,03	0,03	0,04	0,04
9	0,02	0,03	0,03	0,04	0,04
10	0,01	0,02	0,03	0,03	0,03
11	0,01	0,01	0,01	0,01	0,03
12	0,01	0,01	0,01	0,01	0,03
13	0,01	0,01	0,01	0,01	0,02
14	0,01	0,01	0,00	0,00	0,02
15	0,01	0,01	0,00	0,00	0,02
16	0,00	0,00	0,00	0,00	0,00
17	0,00	0,00	0,00	0,00	0,00
18	0,00	0,00	0,00	0,00	0,00
19	0,00	0,00	0,00	0,00	0,00
20	0,00	0,00	0,00	0,00	0,00

Tableau 14 : répartition des témoins précis

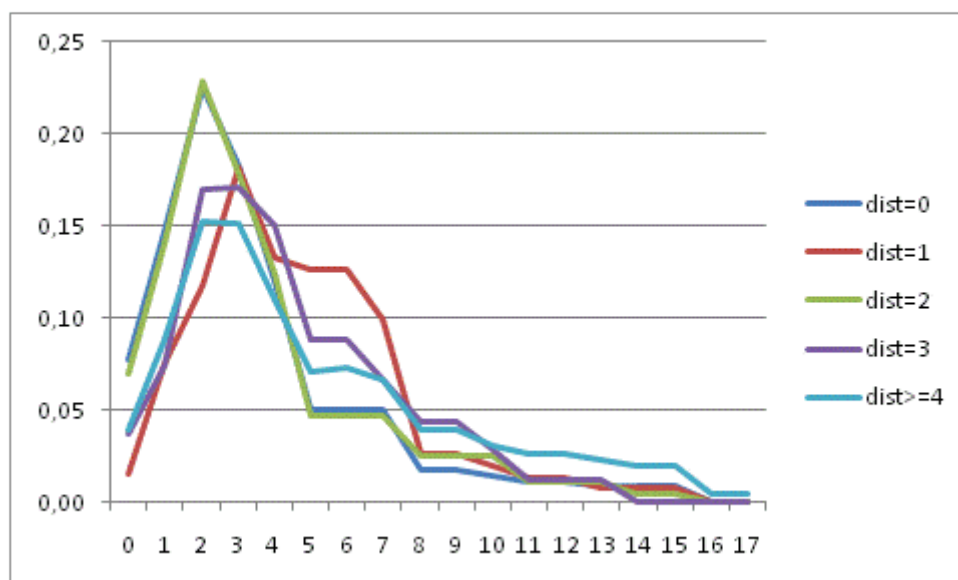


Figure 15 : lois de probabilité associées aux témoins précis

En outre, on peut vouloir prendre en compte les témoins qui disent "quelques secondes", ainsi que l'ensemble de l'épisode entre deux "A", et pas seulement s'ils sont consécutifs. Nous avons supprimé tous les témoins qui annonçaient un délai de plus de trente secondes : un tel délai est impossible. Dans ce cas, il ne faut pas corriger le témoin, mais le supprimer (non digne de foi).

Nous avons fait trois variantes, en remplaçant "quelques secondes" par 2, 4 ou 6 secondes. Les résultats sont identiques et nous ne les reproduisons pas ici.

De même, nous avons fait varier l'amplitude de l' "écart d'indécision" pour chaque témoin, noté d_0 : cela ne modifie pas les résultats.

Le fait que les quatre courbes (cinq, avec les témoins lointains) soient pratiquement identiques montre que les deux épisodes auditifs sont dus à deux explosions distinctes. En effet, une seule explosion produit certes, à elle seule, deux sortes d'effets auditifs : l'onde acoustique principale, et des ondes acoustiques provenant de l'onde sismique (l'explosion produit un séisme, et ce séisme à son tour produit un bruit). Mais l'onde sismique se propage beaucoup plus vite que l'onde acoustique, et, dans ces conditions l'écart de temps entre les deux bruits devrait dépendre de la distance : on voit que ce n'est pas le cas. La seule explication à un écart de temps indépendant de la distance est l'existence de deux explosions, perçues indépendamment par les témoins. Comme les ondes des deux explosions se propagent à la même vitesse, les écarts perçus par les témoins restent les mêmes, quelle que soit la distance.

Le fait que les témoins lointains aient exactement la même perception que les témoins proches corrobore cette assertion de manière très claire : à distance 4, 5, 6 km et au-delà, la différence de propagation entre ondes de nature différente devient très perceptible ; les délais entre onde sismique et onde acoustique augmentent linéairement avec la distance.

Bien entendu, chaque explosion produit aussi ses ondes sismiques, et donc on devrait percevoir quatre épisodes sonores. Certains témoins mentionnent effectivement des phénomènes sismiques et des grondements.

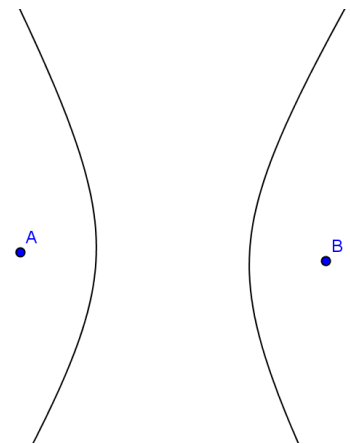
D. Localisation de la première explosion

La question se pose ensuite de déterminer la localisation de la première explosion, la seconde (la plus forte) étant prise pour origine des temps et des axes.

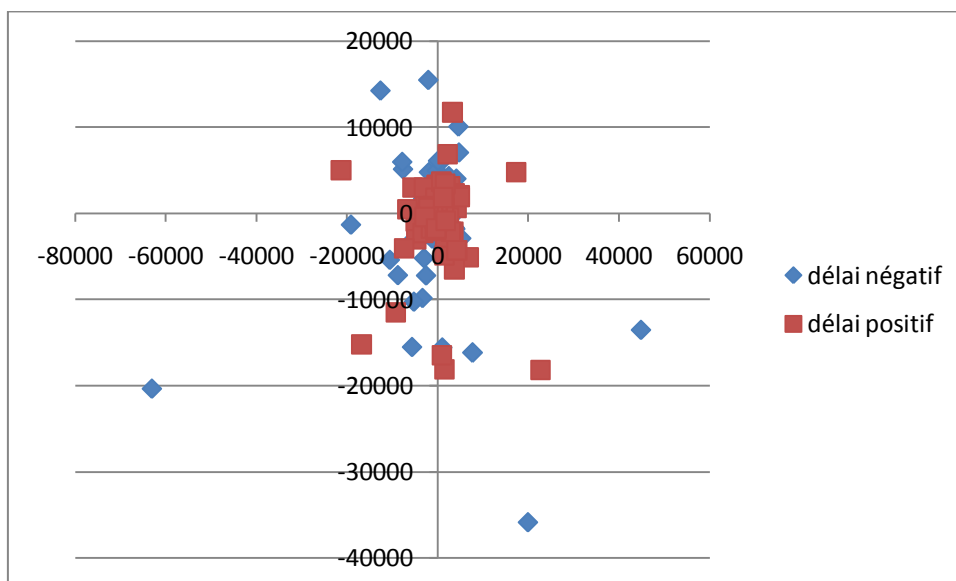
Commençons par ramener les deux explosions à être simultanées, en rajoutant $\tau = 3.7$ s à l'instant annoncé pour la première. Nous avons donc un "délai corrigé", perçu par les témoins : ce délai corrigé tient compte de l'intervalle de temps qui s'est effectivement produit entre les explosions.

Si les deux explosions ne sont pas situées dans la même zone, une part du plan doit avoir un délai corrigé positif, et une autre part un délai corrigé négatif. Pour préciser ceci, imaginons que les deux explosions (simultanées) se soient produites en des points *A* et *B* distants d'un km. L'ensemble des points du plan où les deux explosions sont perçues simultanément est la médiatrice de *AB*, l'ensemble où la première est perçue une seconde (ou davantage) avant la seconde est la portion du plan limitée par un arc d'hyperbole, et de même pour l'ensemble où la seconde est perçue une seconde (ou davantage) après la première : c'est la portion limitée par l'arc opposé.

Si les deux explosions n'étaient pas dans la même zone, il devrait y avoir une région du plan à dominante positive (pour le délai corrigé) et une autre à dominante négative.

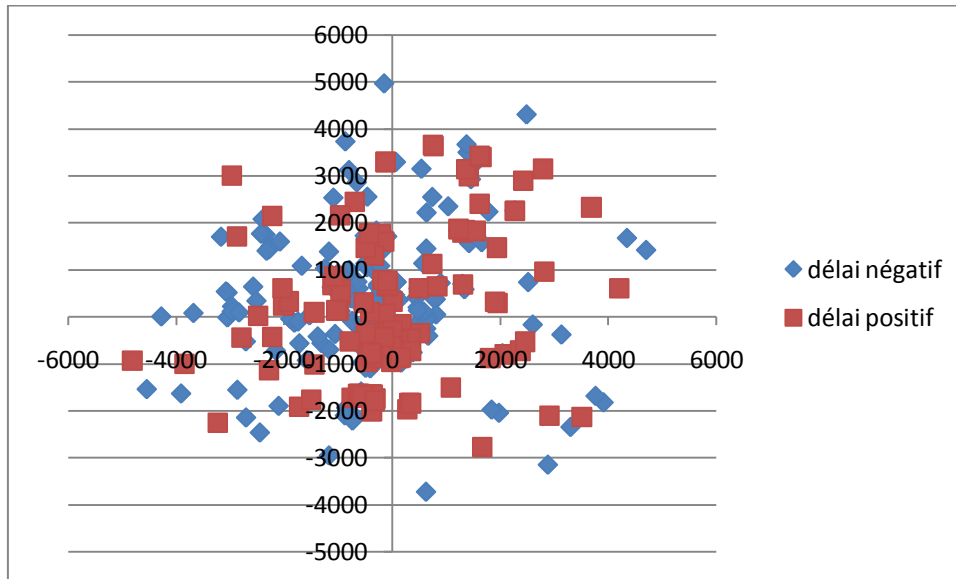


Or, lorsqu'on positionne les témoins sur la carte en tenant compte du délai corrigé, aucune séparation de ce type n'apparaît :



Graphique 16 : position des témoins en fonction du délai corrigé

On peut "zoomer" sur les distances inférieures à 5 km :



Graphique 17 : Focus sur les distances < 5 km

Nous allons rendre ceci quantitatif : nous considérons les 8 points M_k régulièrement espacés sur le cercle de rayon 340 m :

$$x_k = \rho \cos\left(\frac{2k\pi}{8}\right)$$

$$y_k = \rho \sin\left(\frac{2k\pi}{8}\right)$$

avec $k = 0, \dots, 7$.

Faisons l'hypothèse que l'explosion est en M_k et notons, pour chaque témoin, $dist0$ sa distance à l'origine et $dist1$ sa distance au point M_k . Nous avons deux secteurs hyperboliques, à savoir :

$$E_1 = \left\{ dist0 - dist1 > \frac{\rho}{2} \right\}$$

$$E_2 = \left\{ dist1 - dist0 > \frac{\rho}{2} \right\}$$

Si l'explosion est en M_k , les témoins du premier ensemble doivent percevoir l'explosion en O une demi seconde au moins après l'explosion en M_k et c'est l'inverse pour les témoins du second ensemble. Autrement dit, la divergence d'appréciation entre les deux zones devrait être d'une seconde au moins.

Or voici ce que l'on observe pour les 8 points possibles pour la première explosion :

	0	1	2	3	4	5	6	7
nb1	73	87	112	120	113	180	186	98
délai moyen 1	0,27	-0,21	-0,19	-0,28	-0,30	-0,11	-0,25	-0,43
nb2	173	244	224	183	99	122	150	146
délai moyen 2	-0,50	-0,41	-0,37	-0,21	0,01	-0,04	-0,41	-0,42

Tableau 18 : nombre de témoins et délai moyen par zone hyperbolique

Dans ce tableau, nb1 et nb2 désignent respectivement le nombre de témoins dans la zone 1 et la zone 2 : ces nombres sont importants et les résultats sont significatifs.

La différence de délai moyen entre zone 1 et zone 2 est faible, partout inférieure à la seconde. Il n'est donc pas légitime de retenir l'hypothèse selon laquelle l'explosion no 1 se serait produite en l'un quelconque des points M_k .

Nous en déduisons que l'hypothèse la plus vraisemblable est que les deux explosions sont situées dans la même zone ; nous ne sommes pas capables de chiffrer plus précisément l'écart possible entre les deux points.

Déterminer un écart en position à partir d'un délai de propagation n'est possible que si le délai est correctement mesuré ; les témoins sont trop imprécis pour que nous puissions en déduire une localisation de la première explosion.

Bien entendu, cette méthode n'apporte aucune information sur l'altitude éventuelle de la première explosion, puisque nous utilisons uniquement les coordonnées des témoins.

E. Sifflements

Un petit nombre de témoins (8) ont perçu ce qu'ils rapportent comme des "sifflements". Ceci est indépendant de la distance et de la localisation.

F. Exploitation des phénomènes visuels

1. Localisation

On trouve 163 témoins qui disent avoir vu quelque chose et pour lesquels les coordonnées sont disponibles. La répartition par secteur est la suivante :

NE	33
NO	47
SE	39
SO	44
total	163

Tableau 19 : localisation des phénomènes visuels

Le nombre de témoins visuels décroît évidemment avec la distance :

distance	nb témoins
0	94
1	33
2	15
3	13
4	4
5	1
6	1
7	0
8	0
9	1
10	0

11	0
12	1
total	163

Tableau 20 : décroissance des témoins visuels avec la distance

La majorité des témoins ont perçu un seul phénomène visuel :

un seul V	149
deux V	13
trois V	1
total	163

Tableau 21 : répartition par nombre

et parmi eux 130 reportent un éclair, généralement bleu ou blanc.

Huit témoins parlent de boule de feu.

2. Chronologie des événements

Il est intéressant de savoir dans quel ordre les témoins perçoivent les événements :

V avant A	138
A avant V	30

Pour 138 d'entre eux, le phénomène visuel a précédé le phénomène acoustique. Pour 30, il y a eu un phénomène acoustique avant un phénomène visuel (les deux ne sont pas incompatibles, s'il y a eu deux phénomènes visuels : VAV, ou deux phénomènes acoustiques : AVA).

3. Durée moyenne après éclair

Voici la répartition des intervalles de temps, observés par les témoins, en fonction de la distance :

distance (km)	nb témoins	durée moyenne après éclair
0	37	1,15
1	14	3,36
2	8	1,94
3	5	2,60
4	0	

Tableau 22 : répartition selon la distance

La durée dont il s'agit est celle qui sépare l'éclair du phénomène suivant, quel qu'il soit. La statistique ne porte que sur les témoins qui ont effectivement déclaré des durées.

On s'attend normalement à ce que la durée augmente avec la distance (la propagation de l'éclair est instantanée et non celle du son) ; c'est le cas de 0 à 1 km (plages 0 – 1 et 1 – 2), mais ce n'est plus le cas ensuite ; le nombre de témoins est trop faible pour que le résultat soit significatif. On ne sait pas non plus si l'éclair est lié à la première ou à la seconde explosion.

Le fait qu'un éclair se soit produit est clairement établi : 130 témoins au minimum l'ont vu, d'autres emploient un langage différent.

G. Phénomènes tactiles

On compte 190 témoins concernés par les phénomènes tactiles, à savoir souffle et vibrations.

Parmi eux, 128 rapportent des vibrations (phénomènes sismiques).

Répartition par zones :

NE	32
NO	33
SE	21
SO	42
total	128

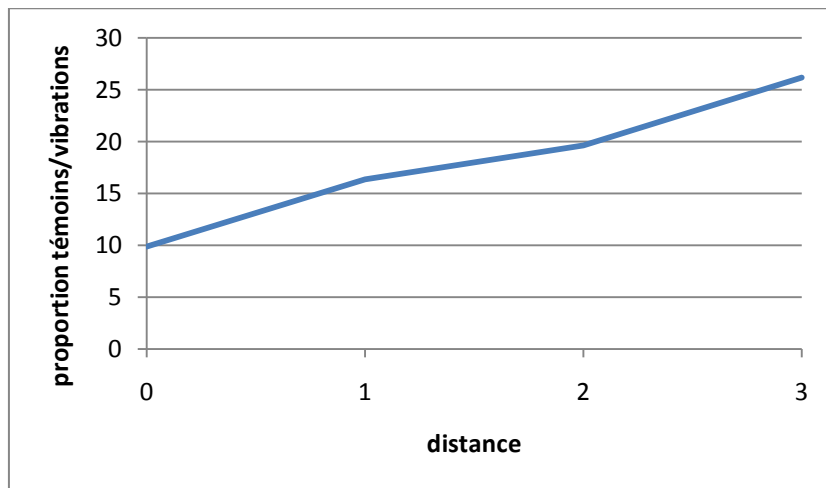
Tableau 23 : répartition par zones

En fonction de la distance :

distance	nb témoins/vibrations	nb total témoins	proportion
0	44	446	9,87
1	23	141	16,31
2	20	102	19,61
3	17	65	26,15
4	8	38	21,05
5	4	22	18,18
6	1	15	6,67
7	0	11	0,00
8	2	13	15,38
9	1	12	8,33
10	2	6	33,33
11	0	8	0,00
12	1	6	16,67

Tableau 24 : répartition en fonction de la distance

Seules les quatre premières lignes (distance 0-1, 1-2, 2-3, 3-4) sont significatives. Voici une représentation graphique :



Graphique 25 : proportion de témoins en fonction de la distance

Elle indique un accroissement de la proportion de témoins ayant perçu des vibrations, avec la distance. Ceci nous paraît normal : les différentes ondes se propagent à des vitesses différentes, et deviennent plus discernables les unes des autres, lorsqu'on s'éloigne.

La très grande majorité des témoins n'a ressenti qu'une seule onde sismique :

une seule vibration	125
deux vibrations	3
trois vibrations	0

Tableau 26 : nombre de vibrations ressenties

Chronologie :

distance (km)	nb témoins	durée moyenne entre vibration et explosion
0	35	1,47
1	13	2,69
2	15	1,77
3	13	3,00
4	6	0,33

Tableau 27 : durée moyenne en fonction de la distance

La durée est calculée pour les témoins qui mentionnent une vibration et une forte explosion (intensité 3) ; elle est le cumul des durées entre épisodes séparant cette vibration et cette explosion. On ne décompte pas les durées où le témoin dit "?", ni celles > 15, manifestement aberrantes.

On s'attendrait normalement à une augmentation de ce laps de temps avec la distance (l'onde sismique se propage plus vite que l'onde acoustique). L'imprécision des témoignages ne permet pas de conforter ceci.

H. Effets de souffle

On compte 70 témoins concernés par l'effet de souffle.

La répartition par secteurs est la suivante :

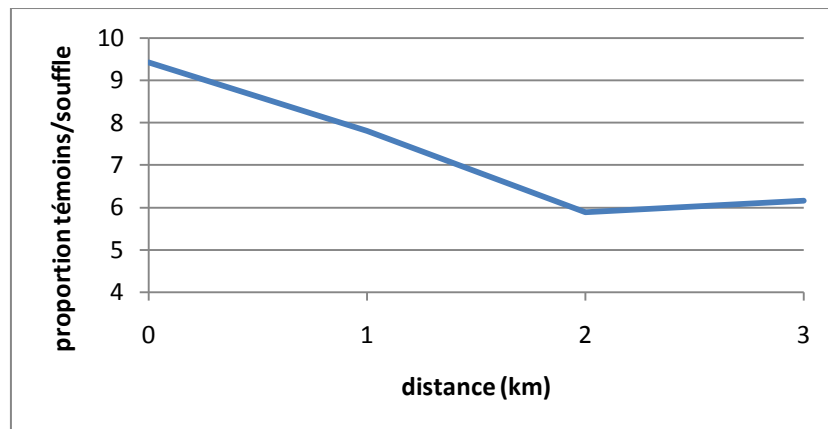
NE	13
NO	18
SE	20
SO	19
total	70

Tableau 28 : répartition par secteurs

La répartition en fonction de la distance est la suivante :

distance	nb témoins/souffle	nb total témoins	proportion
0	42	446	9,42
1	11	141	7,80
2	6	102	5,88
3	4	65	6,15
4	2	38	5,26
5	2	22	9,09
6	1	15	6,67
7	0	11	0,00
8	0	13	0,00
9	0	12	0,00
10	0	6	0,00
11	0	8	0,00
12	0	6	0,00

Tableau 29 : répartition en fonction de la distance



Graphique 30 : perception du souffle en fonction de la distance

La perception du souffle décroît avec la distance, comme on s'y attend.

Sommaire

Résumé	2
Première Partie : Principes généraux pour la constitution d'une base de données "témoignages"	4
I. Une approche scientifique	4
II. Préliminaires sur les témoignages	4
A. Un accident	4
B. Témoins et observateurs	4
C. La notion d' "épisode"	5
III. Construction de la base de données témoignages	5
A. Organisation de la BD	5
1. En première colonne, la liste des témoins	5
2. Seconde colonne : la distance à l'événement	6
3. Troisième colonne : position par rapport à l'événement	6
4. Quatrième colonne : situation (intérieur ou extérieur) par rapport à l'événement	6
5. Cinquième colonne : orientation par rapport à l'événement	6
6. Sixième colonne : situation d'attention du témoin	6
7. Septième colonne et au-delà	7
B. Rien vu ou pas d'information	8
C. Remédier aux défauts habituellement constatés	8
Seconde Partie : constitution de la BD AZF	9
I. Données générales concernant l'accident AZF	9
II. Difficultés rencontrées lors de l'exploitation des fichiers disponibles	9
Troisième partie : Exploitation et Résultats	11
I. Introduction	11
II. Conclusions	11
III. Exploitation de la base de données	13
A. Statistiques globales	13
B. Exploitation des cinq sens	15
C. Exploitation des phénomènes auditifs	15
1. Localisation	15
2. Répartition en fonction de la distance	16
D. Localisation de la première explosion	19

E.	Sifflements.....	21
F.	Exploitation des phénomènes visuels.....	21
1.	Localisation	21
2.	Chronologie des événements.....	22
3.	Durée moyenne après éclair.....	22
G.	Phénomènes tactiles.....	23
H.	Effets de souffle	24