

Comment chercher (et trouver) une aiguille dans une botte de foin

MARC-ANDRÉ DÉSAUTELS, MATHÉMATIQUES
CÉGEP SAINT-JEAN-SUR-RICHELIEU
marc-andre.desautels@cstjean.qc.ca
<http://www.cstjean.qc.ca/accueil>

Résumé

Le 17 janvier 1966, suite à un accident en vol, l'armée américaine perd une bombe à hydrogène au large des côtes espagnoles, à proximité de Palomares. Le 22 mai 1968, le sous-marin nucléaire USS Scorpion est perdu en mer avec 99 marins à bord. Le 12 septembre 1857, le navire SS Central America sombre au fond de l'océan en emportant avec lui une cargaison de près de 11 tonnes d'or. Dans cet article, nous reviendrons brièvement sur l'historique de ces trois situations et sur les mathématiques qui ont permis aux personnes concernées de retrouver la bombe, le sous-marin et le trésor.

Mots clés : Théorème de Bayes, Probabilités, Statistiques, Recherche Bayésienne

1 Introduction

Les événements brièvement relatés dans cet article forment les balbutiements de la recherche bayésienne (voir [8] et [9]). Nous introduirons de façon succincte les événements historiques qui ont contribué à faire connaître cette technique et nous en présenterons les éléments essentiels. En raison de la longueur de l'article, nous ne discutons que de la partie théorique portant sur l'utilisation du théorème de Bayes pour effectuer des recherches. Nous manquons d'espace pour présenter toutes les parties de simulation faisant appel aux méthodes de Monte-Carlo (voir [10]).

2 L'accident nucléaire de Palomares

Une description beaucoup plus détaillée de cet accident peut être trouvée en [5].

2.1 Contexte historique

L'opération Chrome Dome est l'un des nombreux programmes militaires de l'United States Air Force pendant la Guerre froide. Il impliquait des Boeing B-52 en vol, armés d'armes thermonucléaires, assignées à des objectifs en URSS. Jusqu'à une douzaine de bombardiers étaient en vol en permanence, avec des missions de 20 à 30 heures au-dessus du cercle arctique et de la mer Méditerranée. Démarrée initialement en 1958 sous le nom de code *Head Start* puis *Round Robin*, l'opération Chrome Dome a commencé en 1962 et a pris fin le 22 janvier 1968, à la suite de l'accident de Thulé.

2.2 L'accident

Aux alentours de 10h00 le matin, le 17 janvier 1966, deux B-52G basés en Caroline du Sud approchent deux KC-135 Stratotanker au-dessus de la côte espagnole au sud-ouest de Carthagène. Lorsqu'un des B-52 approche l'avion ravitailleur trop rapidement, celui-ci n'a pas le temps de réagir et les avions se percutent à 9 450 mètres d'altitude. Le tuyau de ravitaillement accroche l'un des longerons et l'aile gauche du bombardier se casse. Trois des sept membres d'équipage du bombardier perdent la vie lors de l'écrasement. Les quatre occupants du ravitailleur meurent dans l'explosion lorsque le carburant se consume.

Le bombardier s'écrase près de Palomares avec à son bord quatre bombes à hydrogène. En moins de 24 heures, une équipe de l'armée américaine arrive sur les lieux. Les équipes de recherche trouvent trois bombes en moins d'une journée. Une des bombes se trouve sur une pente douce. Les explosifs dans les deux autres ont explosé lors de l'impact en créant un cratère d'une centaine de pieds et dispersant de la terre, du plutonium, de l'uranium et du tritium. Durant plusieurs semaines, des équipes de l'armée américaine et des équipes de la police espagnole tentent de retrouver la quatrième bombe, sans succès.

La recherche pour la quatrième bombe fit appel à la méthode de recherche bayésienne, supervisée par John Craven. Cette méthode assigne des probabilités aux endroits où la bombe peut se trouver. Elle met ensuite à jour ces probabilités au fur et à mesure de la recherche. Les probabilités initiales ont été trouvées après qu'un pêcheur local, Francisco Simó Orts, ait vu la bombe tomber à l'eau à un certain endroit. La bombe fut ensuite retrouvée grâce aux informations données par Orts, voir figure 1.



FIGURE 1 – La bombe nucléaire B28, retrouvée à 870 m de fond, sur le pont de l’USS Petrel (ASR-14). Par U.S. Navy, Courtoisie de Natural Resources Defense Council [domaine public], via Wikimedia Commons

3 L’USS Scorpion (SSN-589)

Pour en savoir davantage sur le naufrage du USS Scorpion, nous invitons le lecteur à consulter [3] et [4].

3.1 Historique

La quille de l’USS Scorpion est posée le 20 août 1958 au chantier naval Electric Boat. Il est lancé le 19 décembre 1959. Le Scorpion est commissionné le 29 juillet 1960, et son commandement est confié à Norman B. Bessac.

3.2 La mission

Le Scorpion quitte la Méditerranée le 16 mai 1968. Ce jour là deux sous-marins d’attaque soviétique sont à proximité. Le Scorpion est ensuite envoyé pour observer les activités navales soviétiques dans l’Atlantique à proximité des Açores. Ayant observé et écouté les unités soviétiques, le Scorpion s’apprête à retourner à la base navale de Norfolk.

Commençant peu avant minuit le 20 mai et se terminant après minuit le 21 mai, le Scorpion tente d’envoyer des messages radio à la base navale de Rota en Espagne sans succès. Il ne parvient à joindre qu’une station de communication de la Navy, en Grèce, qui a transféré les

messages du Scorpion au ComSubLant¹.

Le lieutenant John Roberts reçoit le dernier message du commandant Slattery, indiquant qu'il se rapproche d'un groupe naval soviétique à une vitesse constante de 15 noeuds (28 km/h) et à une profondeur de 350 pieds (107 m) pour «commencer la surveillance des Soviétiques».

La Navy redoute une possible perte du bâtiment et lance une opération de recherche. L'USS Scorpion et son équipage sont déclarés «présumés perdus» le 5 juin. Son nom est retiré du Naval Vessel Register² le 30 juin. Les recherches se poursuivent avec une équipe de mathématiciens conduite par le Dr John Pinà Craven, le même homme responsable d'avoir retrouvé la bombe perdue à Palomares.

4 Mathématiques

Nous introduirons maintenant la théorie de base qui sous-tend les recherches bayésiennes (voir [8] et [9]). Nous reviendrons ensuite au naufrage du USS Scorpion.

4.1 Les probabilités conditionnelles

La notion de probabilité conditionnelle permet de tenir compte d'une information complémentaire, lors du calcul de certaines probabilités. Par exemple, si je tire au hasard une carte d'un jeu, j'estime à une chance sur quatre la probabilité d'obtenir un coeur. Si j'aperçois un reflet rouge sur la table, je corrige mon estimation à une chance sur deux. Cette seconde estimation correspond à la probabilité d'obtenir un coeur sachant que la carte est rouge.

La probabilité conditionnelle d'un événement A , sachant qu'un autre événement B de probabilité non nulle s'est réalisé est le nombre $P(A | B)$:

$$P(A | B) = \frac{P(A \cap B)}{P(B)}$$

Le théorème de Bayes peut être écrit sous la forme suivante :

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

où A et B sont des événements et $P(B) \neq 0$. De plus,

1. Commander, Submarine Force Atlantic (COMSUBLANT) est le commandant de la force sous-marine.

2. L'inventaire officiel de tous les navires et bâtiments sous le commandement de la marine de guerre des États-Unis d'Amérique.

- $P(A)$ et $P(B)$ sont les probabilités d'observer les événements A et B respectivement.
- $P(A | B)$ est la probabilité conditionnelle d'observer l'événement A étant donné que l'événement B s'est réalisé.
- $P(B | A)$ est la probabilité conditionnelle d'observer l'événement B étant donné que l'événement A s'est réalisé.

4.2 La zone de recherche

Avant d'utiliser le théorème de Bayes (voir [7]) pour effectuer des recherches, nous allons mathématiser la zone de recherche. Cette zone peut être séparée en une grille formée de plusieurs cellules.

$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$
$i = 6$	$i = 7$	\dots		

Nous définissons maintenant quelques variables :

i = indice des cellules

Y_i = position du sous-marin dans la cellule i (1 signifie qu'il est là et 0 signifie qu'il n'y est pas)

π_i = probabilité d'occurrence = $P(Y_i = 1)$

Une composante essentielle de la recherche bayésienne est que les chercheurs peuvent ne pas trouver le sous-marin, même si celui-ci est dans la cellule recherchée. En raison des limitations de la technologie pour observer les fonds marins, il existe une certaine probabilité de détection du sous-marin. Nous définissons donc :

Z_i = résultat de recherche dans la cellule i (1 signifie une détection et 0 signifie pas de détection)

p_i = probabilité de détection = $P(Z_i = 1 | Y_i = 1)$

4.3 Lois de probabilités

Les probabilités p_i et π_i sont des valeurs *a priori* obtenues à partir de l'avis d'experts avant la recherche. Comme indiqué en ouverture de cet article, nous ne discuterons pas de cet aspect dans cet article. Nous pouvons ensuite utiliser le théorème de Bayes pour obtenir des équations permettant de mettre à jour ces probabilités au fur et à mesure que de nouvelles informations

sont obtenues. Nous allons supposer que la présence du sous-marin dans chaque cellule ainsi que la détection de sa présence sont indépendantes l'une de l'autre. Nous supposons également que la détection et la présence du sous-marin suivent la distribution de Bernoulli.

Nous avons donc :

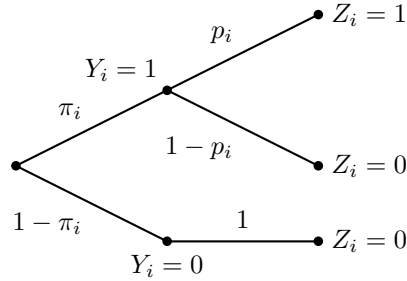
- Le résultat de la recherche dans la cellule i étant donné la présence du sous-marin dans cette cellule est donné par :

$$Z_i | Y_i \sim \text{Ber}(Y_i p_i) = (Y_i p_i)^z (1 - Y_i p_i)^{1-z}, \quad i = 1, \dots, n$$

- La présence du sous-marin dans la cellule i est donnée par :

$$Y_i \sim \text{Ber}(\pi_i) = (\pi_i)^y (1 - \pi_i)^{1-y}, \quad i = 1, \dots, n$$

Nous sommes particulièrement intéressé au cas où la cellule i contient le sous-marin ($Y_i = 1$ avec une probabilité π_i) mais que nous ne le détectons pas ($Z_i = 0$ avec une probabilité $1 - p_i$), c'est-à-dire $P(Y_i = 1 | Z_i = 0)$. Nous pouvons utiliser l'arbre suivant pour calculer cette probabilité.



La probabilité $P(Y_i = 1 \cap Z_i = 0)$ est obtenue en suivant la première branche vers le haut et la seconde branche vers le bas. Nous multiplions les probabilités rencontrées pour obtenir $\pi_i \cdot (1 - p_i)$. Pour obtenir la probabilité $P(Z_i = 0)$, nous suivons les deux branches qui nous permettent d'atteindre $Z_i = 0$. En multipliant les probabilités rencontrées, nous obtenons $\pi_i(1 - p_i) + 1 \cdot (1 - \pi_i)$.

La probabilité est donnée par :

$$\begin{aligned} P(Y_i = 1 | Z_i = 0) &= \frac{P(Y_i = 1 \cap Z_i = 0)}{P(Z_i = 0)} \\ &= \frac{\pi_i(1 - p_i)}{\pi_i(1 - p_i) + 1 \cdot (1 - \pi_i)} \\ &= \frac{\pi_i(1 - p_i)}{1 - p_i\pi_i} \end{aligned}$$

Au départ, nous avons des valeurs initiales (*a priori*) pour les probabilités π_i et p_i . La probabilité p_i est considérée comme constante pour ce problème, car nous supposons que nos appareils de détection restent les mêmes tout au long de notre recherche.

La probabilité π_i quant à elle va changer au fur et à mesure de nos recherches. Nous appellerons donc la probabilité π_i *a priori*, $\pi_{i,\text{ancien}}$. Lorsque nous utiliserons la formule ci-dessus pour calculer une *nouvelle* valeur de π_i (*a posteriori*), nous le noterons comme $\pi_{i,\text{nouveau}}$.

La probabilité *a posteriori* est donc donnée par :

$$\pi_{i,\text{nouveau}} = \frac{\pi_{i,\text{ancien}}(1 - p_i)}{1 - p_i\pi_{i,\text{ancien}}}$$

Remarquons que la probabilité $\pi_{i,\text{nouveau}}$ est **toujours** plus petite que $\pi_{i,\text{ancien}}$. En effet :

$$\begin{aligned} p_i &\geq p_i\pi_i && (\text{car } 0 \leq \pi_i \leq 1) \\ p_i - 1 &\geq p_i\pi_i - 1 \\ 1 - p_i &\leq 1 - p_i\pi_i \\ \frac{1 - p_i}{1 - p_i\pi_i} &\leq 1 \end{aligned}$$

Si le sous-marin n'est pas détecté à la cellule i , nous pouvons donc croire qu'il se trouve dans une autre cellule. La probabilité de la présence du sous-marin dans une cellule $j \neq i$ devrait donc augmenter. Nous voulons trouver la probabilité que la cellule j contienne le sous-marin ($Y_j = 1$) mais que nous ne le détectons pas à la cellule i ($Z_i = 0$), c'est-à-dire $P(Y_j = 1 \mid Z_i = 0)$.

La probabilité est donnée par (en utilisant l'arbre et la démarche précédente) :

$$\begin{aligned} P(Y_j = 1 \mid Z_i = 0) &= \frac{P(Y_j = 1 \cap Z_i = 0)}{P(Z_i = 0)} \\ &= \frac{P(Z_i = 0 \cap Y_j = 1)}{P(Z_i = 0)} \\ &= \frac{P(Z_i = 0 \mid Y_j = 1)P(Y_j = 1)}{P(Z_i = 0)} \\ &= \frac{1 \cdot \pi_j}{1 - \pi_i p_i} \quad \text{car } P(Z_i = 0 \mid Y_j = 1) = 1 \end{aligned}$$

Le résultat $P(Z_i = 0 \mid Y_j = 1)$ représente la probabilité de ne pas détecter le sous-marin à la cellule i étant donné que le sous-marin se trouve à la cellule j . Cette probabilité est nécessairement égale à 1.

La probabilité à *a posteriori* est donc donnée par :

$$\pi_{j,\text{nouveau}} = \frac{\pi_{j,\text{ancien}}}{1 - p_i \pi_{i,\text{ancien}}}$$

La probabilité $\pi_{j,\text{nouveau}}$ est **toujours** plus grande que $\pi_{j,\text{ancien}}$. En effet :

$$\begin{aligned} 0 &\leq p_i \\ 0 &\leq p_i \pi_i \\ -1 &\leq p_i \pi_i - 1 \\ 1 &\geq 1 - p_i \pi_i \\ \frac{1}{1 - p_i \pi_i} &\geq 1 \end{aligned}$$

5 Retour sur le naufrage du USS Scorpion

En utilisant les deux formules trouvées précédemment, voici la façon de trouver le sous-marin :

1. Tant que le sous-marin n'est pas trouvé nous itérons m fois la recherche.
2. À chaque étape, une cellule est cherchée. Cette cellule est choisie comme étant celle avec la probabilité la plus élevée.
3. Nous effectuons la recherche à l'aide d'une épreuve de Bernoulli.
4. Si le sous-marin n'est pas trouvé, nous mettons à jour les probabilités de toutes les cellules.
5. Si le sous-marin est trouvé, nous arrêtons la recherche.

Les probabilités suivantes ont été trouvées à l'aide d'avis d'experts et de simulation de type Monte-Carlo. Vous pouvez voir la grille de recherche à la figure 2.

À la figure 2, les probabilités *a priori*, c'est-à-dire $\pi_{i,\text{ancien}}$, sont présentées. Après l'utilisation des formules, nous obtenons les probabilités *a posteriori*, c'est-à-dire $\pi_{i,\text{nouveau}}$. Nous obtenons donc une nouvelle grille de probabilités. Nous cherchons à l'endroit où la probabilité est la plus élevée et si nous ne trouvons pas le sous-marin, nous recalculons les probabilités *a posteriori* et nous recommençons jusqu'à la découverte du sous-marin.

Nous pouvons visualiser plus facilement la grille en coloriant celle-ci, comme présenté à la figure 3.

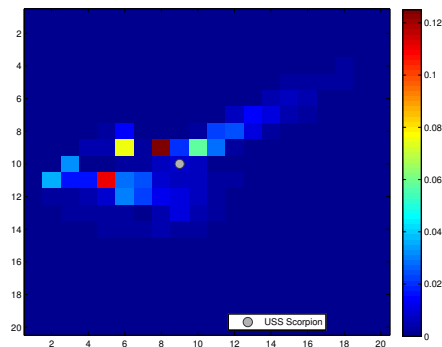


FIGURE 3 – La grille recherche du Scorpion obtenue à partir de l’avis d’experts et de simulation Monte-Carlo. Les probabilités en ordre croissant passent du bleu au rouge.



FIGURE 4 – Une photo de la poupe du USS Scorpion. Transféré à wikipédia common en provenance de www.history.navy.mil.

6.1 Historique

Le Steamship (SS) George Law, construit par William H. Webb à New York et appartenant à la United States Mail Steamship Company, est lancé le 28 octobre 1852 et prend pour la première fois la mer le 20 octobre 1853. C’est un trois-mât à vapeur de 85 m pour 2 141 tonnes. En 1857, il est rebaptisé le Central America. Vous pouvez voir une gravure du navire à la figure 5.

Le 12 septembre 1857, le navire sombre par 2 300 mètres de fond lors d’un ouragan avec ses 477 passagers, ses 101 membres d’équipage et sa cargaison, près de 11 tonnes d’or (30 000 livres),

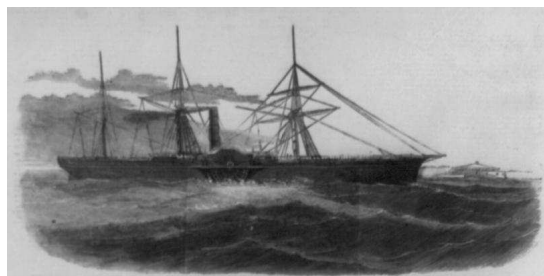


FIGURE 5 – Gravure représentant le Central America. Source : Library of Congress et Auteur : Frank Leslie's Illustrated Newspaper

qui lui valurent son surnom (Ship of Gold), sa célébrité et contribua à la panique de 1857³.

En septembre 1988, Tom Thompson convainc plus de cent investisseurs de financer une expédition destinée à localiser l'épave et à remonter le trésor englouti avec elle. Le navire est localisé le 11 septembre 1988, une grande partie de l'or est remontée à la surface. Suite à la remontée du trésor, 39 compagnies d'assurance poursuivent, en alléguant qu'elles ont payé des dommages au 19ème siècle et que le trésor leur revient. L'équipe responsable de la découverte du navire affirme pour sa part que le trésor avait été abandonné. Après la poursuite juridique, 92% de l'or est remis à l'expédition.

Pour tenter de retrouver le navire, les recherches utilisèrent les informations suivantes :

- Historique : La position de navires, les estimations de la vitesse du vent et sa direction durant l'ouragan, la dérive des survivants, etc.
- Statistique : La distribution statistique des vents et des courants dans la région.
- Analytique : Estimation de l'incertitude dans la navigation céleste, estimation de l'effet du vent sur la dérive du navire, estimation du courant produit par le vent.
- Subjectif : Pondérations représentant la qualité de l'information utilisée pour estimer la position du navire.

La référence [1] raconte en grand détail les recherches faites et l'utilisation de toutes les informations pour créer les probabilités initiales de recherche.

3. Un des premiers krachs de Wall Street.

7 Conclusion

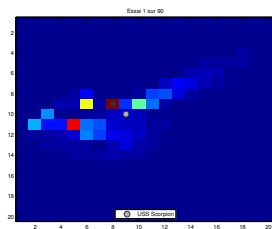
Les trois histoires précédentes sont plus grandes que nature. Elles permettent d'éveiller l'intérêt des interlocuteurs avant d'introduire les notions mathématiques, somme toute assez simples. Elles pourraient être utilisées comme éléments déclencheurs dans un cours de statistiques où l'on discute du théorème de Bayes.

Pour obtenir une idée plus précise de l'utilisation de ces techniques dans des circonstances plus modernes, nous invitons le lecteur à lire la référence [6]. L'article raconte en détail les méthodes utilisées pour retrouver le vol AF 447 d'Air France qui avait été perdu en mer le 1^{er} juin 2009. Le vol reliait Rio et Paris et s'est abîmé dans l'océan Atlantique entraînant la mort des 228 personnes à bord.

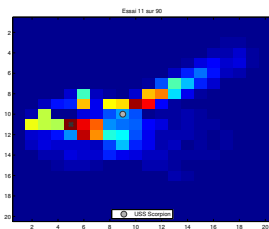
Références

- [1] Lawrence D. Stone (1992). Search for SS «Central America » : Mathematical Treasure Hunting, *Interfaces*, 22, (no. 1), pages 32-54.
- [2] Lawrence D. Stone (2010). Revisiting the SS Central America Search *Information Fusion (FUSION)*, 2010 13th Conference on. Edinburgh, UK : IEEE.
- [3] Sontag, Sherry and Drew, Christopher (1998). *Blind Man's Bluff : The Untold Story Of American Submarine Espionage*. New York, USA : PublicAffairs.
- [4] Richardson, Henry R. and Stone, Lawrence D. (1971). Operations analysis during the underwater search for Scorpions, *Naval Research Logistics Quarterly*, 18, (no. 2), pages 141-157.
- [5] Moran, Barbara (2009). *The Day We Lost the H-Bomb : Cold War, Hot Nukes, and the Worst Nuclear Weapons Disaster in History*. New York, USA : Presidio Press.
- [6] Stone, Lawrence D. and Keller, Colleen M. and Kratzke, Thomas M. and Strumpfer, Johan P. (2014). Search for the Wreckage of Air France Flight AF 447, *Statist. Sci.*, 29, (no. 1), pages 69-80.
- [7] McGrayne, Sharon Bertsch(2012). *The Theory That Would Not Die : How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*. Yale University Press.
- [8] Koopman, B. O. (1956). The theory of search. I. Kinematic bases., *Operational Research*, (no. 4), pages 324-346.

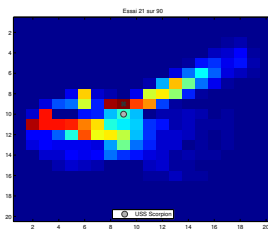
- [9] Koopman, B. O. (1956). The theory of search. II. Target Detection., *Operational Research*, (no. 4), pages 503-531.
- [10] Hammersley, J. M. et Handscomb, D. C. (1964). *Monte Carlo Methods*. Chapman and Hall, London et New York, 1964.



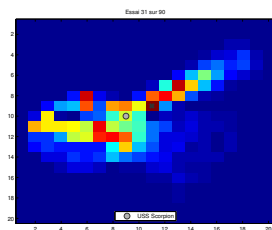
(a) Animation 1 sur 90



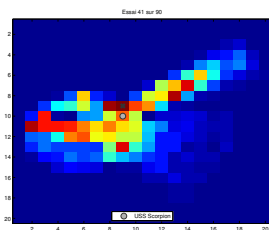
(b) Animation 11 sur 90



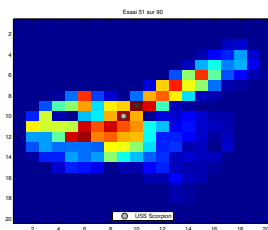
(c) Animation 21 sur 90



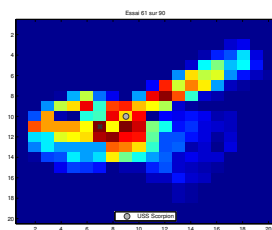
(d) Animation 31 sur 90



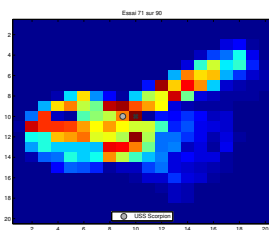
(e) Animation 41 sur 90



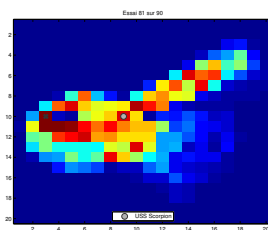
(f) Animation 51 sur 90



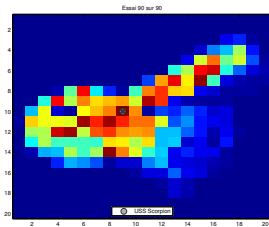
(g) Animation 61 sur 90



(h) Animation 71 sur 90



(i) Animation 81 sur 90



(j) Animation 90 sur 90

FIGURE 6 – Le symbole \times représente la cellule cherchée à cette itération et le point gris représente le sous-marin USS Scorpion.