



Wave Detector

Elément d'un tableau de bord à destination des organisations d'enquêtes ufologiques

Jean-Marc Wattecamps, Responsable du réseau d'enquêteurs du COBEPS

Introduction

Le COBEPS surveille de manière systématique l'activité ufologique en Belgique francophone et germanophone depuis 2010. Il réalise et met en ligne un observatoire disponible à l'adresse www.cobeps.org/fr/notifications.html. Il s'agit d'un rapport cumulatif et synthétique des données ufologiques disponibles : signalements, enquêtes et classement des observations. Il est publié trimestriellement a posteriori. Le responsable des enquêtes, qui est en contact régulier avec les notifications qui parviennent au COBEPS via le formulaire en ligne, peut constater de façon empirique l'évolution de celles-ci et éventuellement donner l'alerte en cas d'augmentation des signalements. Toutefois, d'une part l'observatoire est tardif et d'autre part, l'évaluation empirique reste subjective.

Le but de l'élaboration d'un « Wave Detector », à l'instar du projet du CISU et du travail d'Abraïni (Abraïni, 2005), est de proposer un indicateur mensuel d'activité ufologique et un système d'alerte automatique en cas d'activité ufologique anormale. Ce système est basé sur les données collectées pour l'observatoire et leur analyse par les techniques statistiques des séries temporelles et des distributions de probabilité. Nous invitons le lecteur à se documenter sur les séries temporelles avec les syllabi disponibles en ligne par exemple pour une première approche Nicolau (2006) - <http://www.nicolau.nom.fr/publications.html> ou pour une connaissance plus approfondie Roustant (2008) - http://www.emse.fr/~roustant/Documents/polycopie_series_temporelles_2008_2009.pdf.

1. Les données

1.1. Collecte et choix du paramètre et de la période

Nous disposons de la base de données qui sert à l'élaboration de l'observatoire ufologique du COBEPS. Ces données reprennent l'ensemble des signalements (ou notifications) de Phénomènes Aérospatiaux Non-identifiés (PAN)¹ recensés par le COBEPS, ses membres et ses partenaires. La méthode de collecte a été la plus systématique et stable que possible durant la période concernée. Il s'agit de données brutes contenant des observations parfaitement identifiables et d'autres qui le sont moins ou pas du tout. Au 31/12/2015, la base de données contenait 597 lignes qui sont considérées comme autant d'observations indépendantes, même si formellement tel n'est pas nécessairement le cas². Le lecteur curieux est invité à consulter la page de l'observatoire du COBEPS sur le web (2015 - cfr. adresse ci-dessus).

- 1 Les PAN ne concernent que des phénomènes qui ont une composante visuelle et qui se trouvent à un moment dans l'espace aérien. Cela exclut donc les « entités » vues seule et au sol, les crops circles non accompagnés de phénomènes aériens, les bruits sans cause. Il y a cependant quelques cas de « photos surprises ».
- 2 Sans disposer de chiffre, l'observation de la base de données montre très peu de cas ou deux témoins signalent indépendamment le même phénomène.

Les signalements sont effectués généralement très rapidement après l'observation : 45,86 % des notifications arrivent dans les 24h et 53,97 % dans les 48h (Figure 1 : les deux premières colonnes). Le délai de notification se réduit de plus en plus, grâce à la rapidité d'Internet. La figure 1 comprend 580 données, c'est-à-dire 580 lignes ou les informations relatives au moment de l'observation et simultanément à celui du signalement sont disponibles. Ceci est un excellent indicateur de la qualité des données collectées.

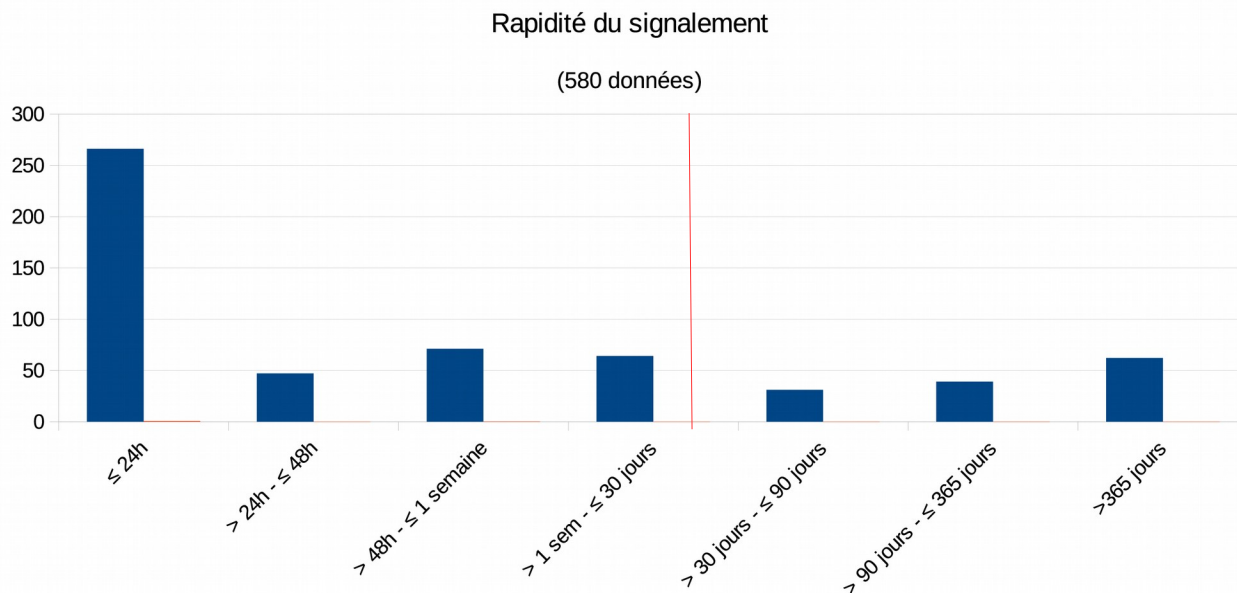


Figure 1 : rapidité du signalement d'une observation par un des témoins

Nous retiendrons comme variable pour notre Wave Detector : le nombre de cas datés signalés mensuellement sans aucun filtre lié à un traitement quelconque (variable NbS). La répartition temporelle des observations dans la base de données montre en effet, que, pour obtenir des lots quantitativement suffisant, nous devons considérer l'intervalle mensuel (période $t \pm 30$ jours). Il s'agit du meilleur compromis entre pertinence et efficacité pour notre Wave Detector. C'est par ailleurs le choix qui a également été fait par Abraini (2005). Nous ne retiendrons donc, dans cette étude, que les signalements qui sont parvenus dans les 30 jours après l'observation (tous ceux situés à gauche du trait rouge sur la figure 1). D'autre part, nous ne considérerons pour cette étude que les données mensuelles parvenues au COBEPS entre 2011 et 2015. En effet, dans le courant de l'année 2010, le relevé systématique des observations n'était pas assez efficace. Nous obtenons de la sorte une série de 424 données mensuelles pour un nombre de périodes $t = 60$ (Tableau 1). Ces mêmes données peuvent être présentées sous la forme d'un graphique de série temporelle dans la figure 2.

Tableau 1 : la série de données étudiée (variable NbS)

| Années (j) | Mois (i) | | | | | | | | | | | | Total |
|--------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | |
| 2011 | 0 | 4 | 6 | 9 | 5 | 5 | 7 | 9 | 7 | 7 | 9 | 10 | 78 |
| 2012 | 7 | 3 | 7 | 3 | 11 | 8 | 7 | 23 | 15 | 3 | 10 | 6 | 103 |
| 2013 | 13 | 5 | 1 | 8 | 3 | 2 | 9 | 15 | 15 | 4 | 4 | 4 | 83 |
| 2014 | 8 | 4 | 7 | 2 | 6 | 8 | 9 | 8 | 4 | 5 | 4 | 8 | 73 |
| 2015 | 6 | 0 | 5 | 4 | 4 | 11 | 8 | 17 | 12 | 6 | 4 | 10 | 87 |
| Total | 34 | 16 | 26 | 26 | 29 | 34 | 40 | 72 | 53 | 25 | 31 | 38 | 424 |

Le total des PAN D, soit les cas qui restent non identifiés à ce jour³, dans cette population, est de 9 ce qui est totalement insuffisant pour une analyse spécifique.

En conséquence notre jeu de donnée correspond à :

- series 1: All UFO and IFO reports, uda de Abraini (2005)

Cette série correspond en fait parfaitement aux données brutes qui arrivent au COBEPS, données sur lesquelles le Wave Detector devra se montrer utile.

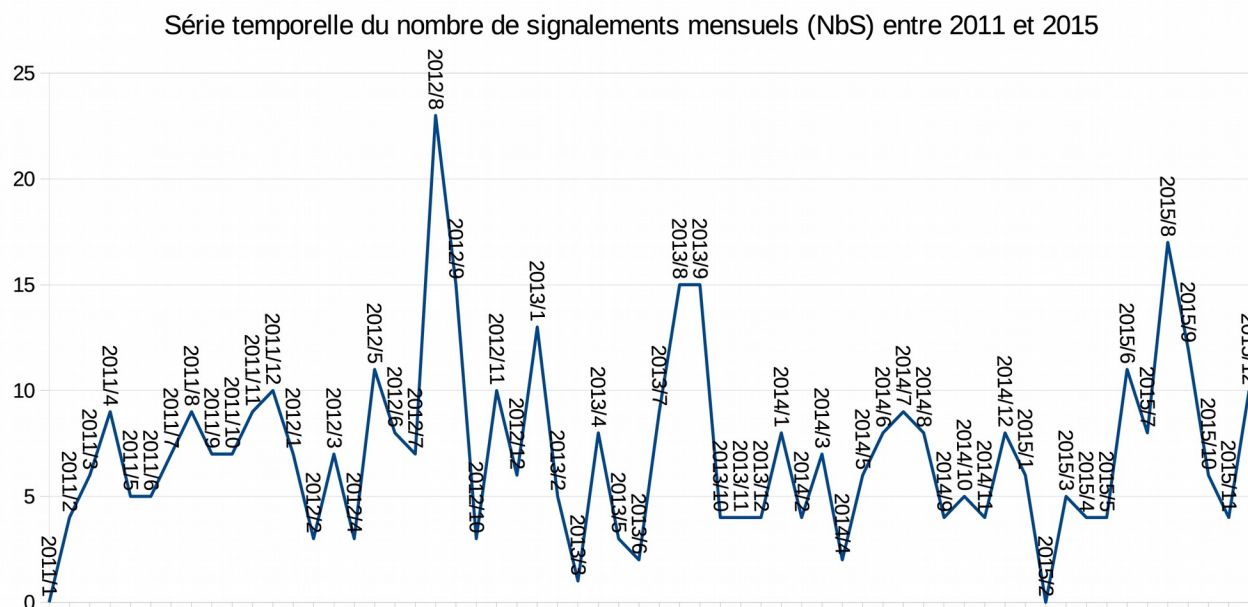


Figure 2 : représentation graphique de la série temporelle étudiée

Le graphique de la figure 2 ne fait pas ressortir d'évidence une tendance générale. Il montre effectivement des pics très nets en août et septembre, des creux moins marqués que les pics en février, mars ou avril selon les années (soit un pic en été et un creux au début du printemps). Notons que cette composante saisonnière ne s'exprime pratiquement pas en 2011 et peu en 2014 au niveau du pic estival.

1.2. Discussion sur la nature de la variable NbS

Il s'agit de données d'une seule variable à temps discret. Nous placerons arbitrairement le point d'encrage des données au 15 de chaque mois. NbS est une variable de débit qui a un caractère quantitatif mais discret (Dehon, Droesbeke & Vermandele, 2015). En effet, il ne s'agit pas de nombres réels mais de nombres entiers. Nous restons dans un cadre non déterministe. Nous utiliserons des méthodes empiriques et aléatoires (Ladiray & Quenneville, 1999). Nous constatons que NbS semble présenter un caractère saisonnier.

Dans NbS se trouve mélangé des observations relatives à des PAN A et B (parfaitement ou probablement identifié), comme non identifiés (PAN D) mais aussi des phénomènes non-identifiables par manque de données (PAN C). Nous postulons aussi que la période considérée (2011 à 2015) est une période ordinaire qui ne présente pas de vague d'OVNI (autre terme employé pour désigner les PAN D) et qui est donc un bon étalon. Selon les données de l'observatoire, le nombre de signalements annuels varie entre 86 et 124 et il n'y a pas de discontinuité particulièrement marquée durant cette période.

Le recul de 5 années de données ne permet cependant pas de tirer des conclusions définitives sur la qualité de notre Wave Detector. Celui-ci devra être réévalué.

3 15/02/2016



2. Méthodologie

L'objectif est d'obtenir un système d'alerte lorsque le NbS dépasse un certain seuil. Nous discuterons et testerons la méthode développée par Abraini en 2005.

Abraini utilise le logiciel X12-ARIMA, développé par l'U.S. Census Bureau (2011)⁴. Ce programme effectue automatiquement une série de procédures pour déterminer tendance générale, saisonnalité, irrégularités, pour aboutir à un ajustement optimum et une prévision de la série temporelle. Il est largement reconnu pour ses qualités et utilisé dans le monde entier par les agences officielles, mais comme tout outil, il doit être utilisé avec certaines précautions. Nous invitons les lecteurs désireux d'en apprendre davantage à consulter le manuel disponible ici : <https://www.census.gov/ts/x12a/v03/x12adocV03.pdf> ou l'excellent manuel de « Office for National Statistics of United Kingdom (2007) disponible à l'adresse www.ons.gov.uk/ons/guide.../time.../guide-to-seasonal-adjustment.pdf .

La principale condition pour pouvoir appliquer X12-ARIMA est qu'il ne peut y avoir de données manquantes sur la série temporelle. Il faut aussi disposer de minimum trois années de relevés pour mettre en évidence une saisonnalité et cinq années pour que les prévisions soient suffisamment validées. Abraini dispose de trois années et neuf mois de données et seule la saisonnalité peut être approchée, ce qui justifie sans doute l'usage partiel qu'il fait du logiciel X12 dans son étude. L'auteur en effet n'utilise que les valeurs de saisonnalité alors qu'il pourrait utiliser les prévisions et leur intervalles de confiance obtenue par régression.

Abraini calcule ensuite le nombre de signalements théorique supérieur au percentile 95 % de la distribution observée (uda). Il ajoute enfin pour chaque mois la composante saisonnière caractéristique déterminée par X12, pour obtenir une valeur seuil indicatrice d'un début de vague. Cette méthode combine des techniques de séries temporelles et une approche plus classique de la statistique.

Nous allons appliquer la **méthode Abraini** (méthode utilisant la saisonnalité établie par X12 et le percentile de la distribution de uda) à nos données en vérifiant pas à pas la validité des hypothèses. Nous verrons s'il est possible d'obtenir prévisions et intervalles de confiance par la **méthode X12** sans traitement a posteriori. Enfin nous testerons une méthode qui part de l'étude des différentes distributions mensuelles observées. Il s'agit de la **méthode COBEPS**. Nous comparerons alors les différentes méthodes pour choisir celle qui apparaîtra la plus efficace et justifiée. La première étape revient donc à appliquer l'analyse X12 à NbS et d'étudier sa distribution globale avant de développer la méthode COBEPS.

2.1. Application de l'analyse X12 à NbS

L'analyse X12 consiste à appliquer un ensemble de procédures automatiques successives à une série temporelle. Nous avons appliqué X12 à travers l'interface Gretl (Cottrell & Lucchetti, 2016). Le logiciel produit en sortie, un fichier qui contient les résultats de l'application de ces procédures (Annexe 1 - sortie pour la variable NbS).

Nous avons d'abord vérifié que notre variable NbS est compatible avec X12. Le tableau 1 et la figure 2 montrent qu'il y a deux zéros dans notre variable NbS. Ce qui peut poser des problèmes dans l'application du modèle multiplicatif. Fort heureusement la série correspond à un modèle additif⁵ plutôt que multiplicatif (cfr Abraini ou les autres ressources en matière de séries temporelles). Le test réalisé par X12 pour vérifier cette conclusion est celui d'Akaike.

4 La version X13 la remplace actuellement.

5 Le modèle additif utilise l'équation suivante : $NbS_{ij} = T_{ij} + S_{ij} + I_{ij}$ avec T : la composante relative à la tendance, S : la composante saisonnière et I : la composante irrégulière ; alors que le modèle multiplicatif s'écrit $NbS_{ij} = T_{ij} \times S_{ij} \times I_{ij}$ mais peut aussi s'écrire $\text{Log}(NbS_{ij}) = \text{Log}(T_{ij}) + \text{Log}(S_{ij}) + \text{Log}(I_{ij})$ ce qui justifie une transformation logarithmique des données.

La première étape est celle d'une régression et d'une auto-corrélation afin d'établir un modèle prévisionnel. ARIMA est choisi automatiquement par la méthode TRAMO (Gomez & Maravall, 2000) incluse dans les procédures X12, les coefficients les mieux adaptés pour nos données sont les suivants : (0 0 0) – (0 1 1). Ce traitement abouti, après 13 itérations et l'évaluation de 27 fonctions d'ajustement, à des tableaux de prévisions et de leurs intervalle de confiance sur 2016 (Annexe 1 : tableaux A et B). Il s'agit d'un modèle encore non ajusté par l'étude de la saisonnalité.

L'étude de la saisonnalité est ensuite réalisée en utilisant le modèle additif. Le logiciel détecte des ruptures dans la saisonnalité des données et les corrige (Annexe 1 : tableaux C6, D8 et E5), il teste une saisonnalité résiduelle dans la composante saisonnière (S) augmentée de la composante irrégulière (I) (Annexe 1 : tableau D8A). X12 montre aussi qu'il n'y a pas d'effet « calendrier » sur la saisonnalité (Annexe 1 : tableau D16). Il aboutit finalement à la composante saisonnière corrigée (Annexe 1 : tableau D10, D10 A) qu'utilise Abraini.

Tableau 2 : la composante saisonnière moyenne après l'ajustement

| Mois | S-moyen |
|------|---------|
| 1 | 0 |
| 2 | -4 |
| 3 | -2 |
| 4 | -2 |
| 5 | -1 |
| 6 | 0 |
| 7 | 1 |
| 8 | 7 |
| 9 | 4 |
| 10 | -2 |
| 11 | -1 |
| 12 | 1 |

Le fichier de l'annexe 1, se poursuit par une série de tableaux (E et F) qui permettent la réalisation d'un diagnostic final sur la qualité des statistiques calculées (Annexe 1 : Tableau F3). Selon Ladiray & Quenneville (1999) : « Ce tableau présente onze statistiques permettant de juger la qualité de l'ajustement saisonnier. ... Ces statistiques varient entre 0 et 3, mais seules les valeurs en dessous de 1 sont jugées acceptables. Un indicateur synthétique de la qualité de la désaisonnalisation est construit comme combinaison linéaire de ces 11 statistiques. » Notre indicateur linéaire porte sur les seules sept statistiques mesurables : « Les statistiques M8 à M11 ne sont calculables que si la série couvre au moins 6 années. Dans le cas contraire, le vecteur de poids utilisé pour calculer la statistique Q est (14, 15, 10, 8, 11, 10, 32, 0, 0, 0, 0). » La valeur globale du diagnostic de NbS atteint la note de 2,18. Six tests sur huit sont en échec. L'ajustement saisonnier est donc de mauvaise qualité.

Il est possible par ailleurs de s'en rendre compte graphiquement à la figure 3. La composante irrégulière, en particulier, présente une très grande variabilité et de même amplitude que la série originale !

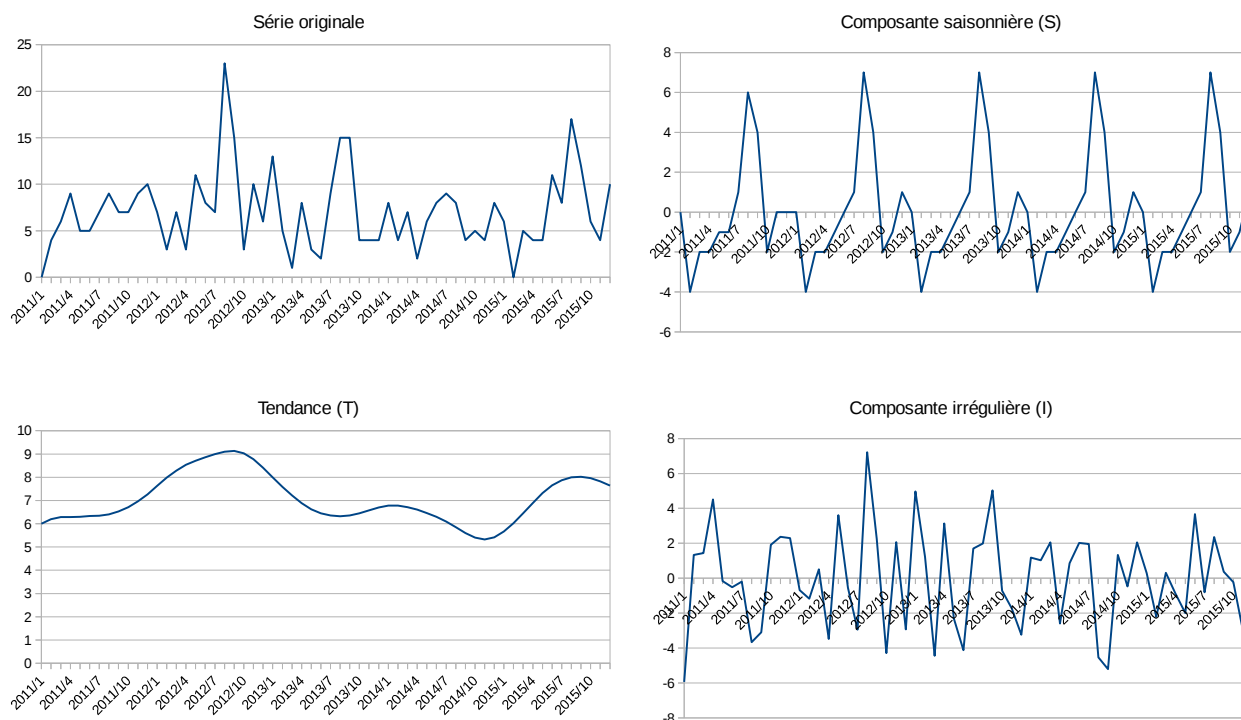


Figure 3 : décomposition de la série NbS en ses composantes par X12-Arima

Ceci ne veut pas dire que les données n'ont pas un caractère saisonnier. Le tableau F2I de l'annexe 1 montre que le F-test pour une saisonnalité stable donne une valeur de 3,236. La probabilité de l'absence d'une saisonnalité est de 0,37%. Cependant, les tests combinés sur la série soustraite de la tendance (NbS-T) semblent écarter un caractère saisonnier (Annexe 1 tableau D8). Les données ont un caractère saisonnier mais leur désaisonnalisation ne donne pas de garantie de qualité suffisante.

En conclusion l'emploi des valeurs mensuelles de S ou de NbS-S (c'est-à-dire les données désaisonnalisées) dans l'établissement de seuils d'alerte n'est pas statistiquement justifié pour les données NbS de la partie francophone de la Belgique.

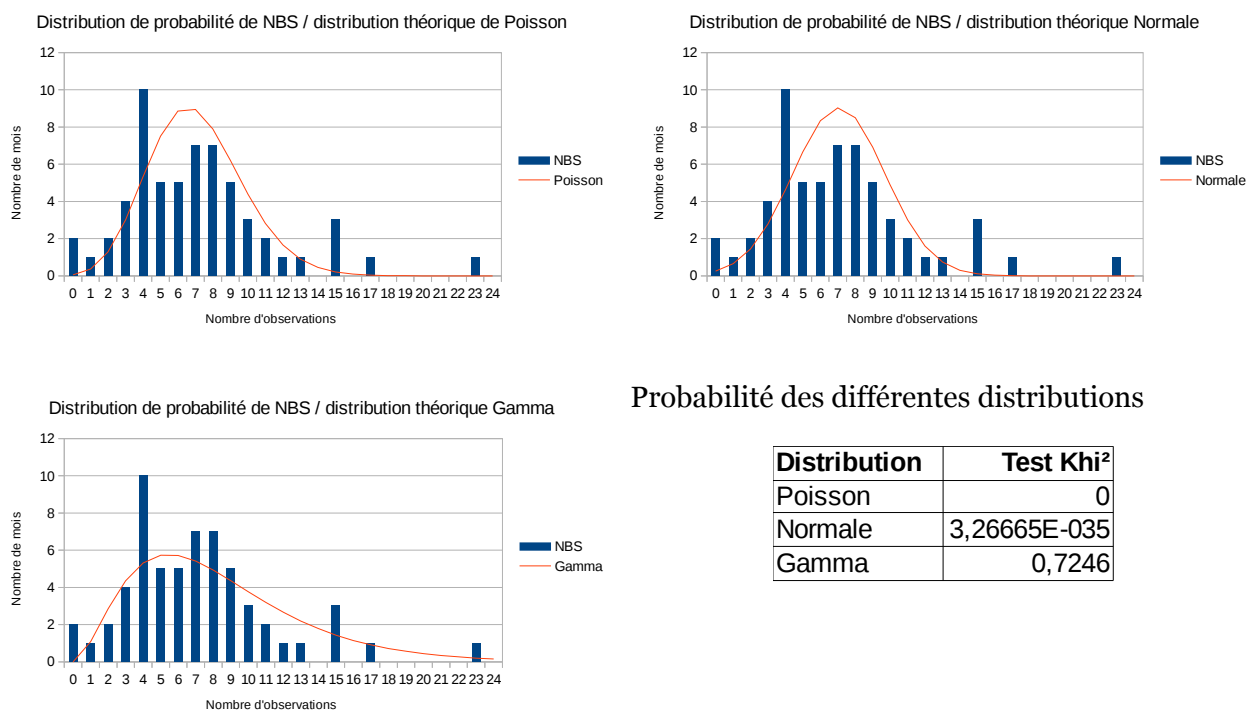
2.2. Loi de probabilité de la distribution de NbS

Il semble donc difficile de se fier à l'ajustement saisonnier par X12 pour NbS. L'autre stratégie est de voir si la distribution des signalements obéit à un ordre précis, à une distribution statistique quelconque ce qui nous permettrait de vérifier si une fréquence mensuelle observée à un temps X s'écarte significativement de la distribution aléatoire.

Berger (1976), corrigé ensuite par Bellester-Olmos (1977), considèrent les fréquences du nombre de signalements journaliers durant la vague de 1954 en Autriche entre le 31 août et le 20 novembre 1954. Finalement, selon ces auteurs, la fréquence du nombre de signalements journaliers ne peut y être ramené à une distribution aléatoire de Poisson.

Abraïni (2005) signale que pour les données italiennes qui couvrent la période de janvier 2002 à octobre 2005, si l'on considère cette fois la distribution de fréquence du nombre de signalements mensuels, on peut considérer aussi bien une loi de distribution normale qu'une distribution de Poisson pour l'ensemble des séries considérées sur la période (dont la série uda). Plus le nombre de signalements est bas, plus il se rapproche de la distribution de Poisson ; plus il est élevé, de la distribution normale.

La figure 4 présente la distribution de fréquence du nombre de signalements mensuels, ajustée à différentes distributions théoriques.



Probabilité des différentes distributions

| Distribution | Test χ^2 |
|--------------|---------------|
| Poisson | 0 |
| Normale | 3,26665E-035 |
| Gamma | 0,7246 |

Figure 4 : distributions de la variable NbS

Le test χ^2 indique des probabilités nulles ou extrêmement faibles pour les distributions de Poissons ou normales mais une probabilité de 0,7246 pour la distribution Gamma. Celle-ci semble la mieux adaptée. Nous avons encore testé la loi lognormale qui a une probabilité de 0,1484.

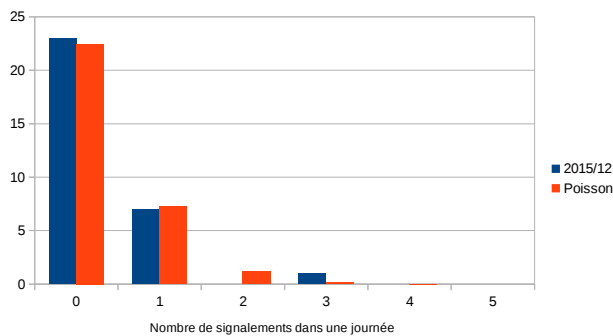
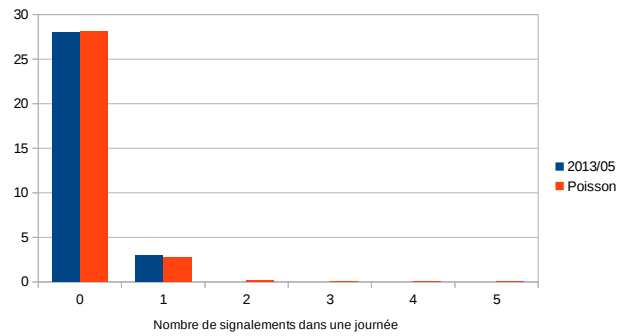
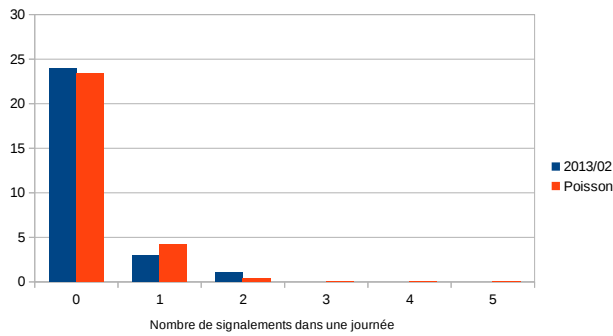
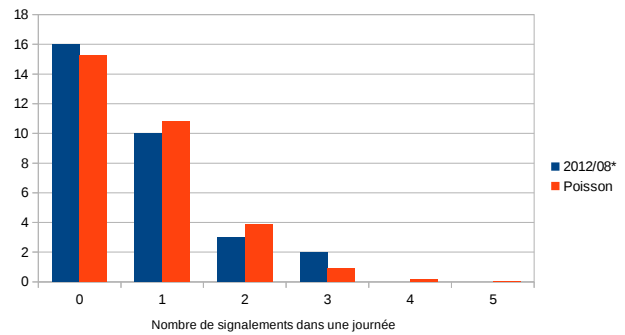
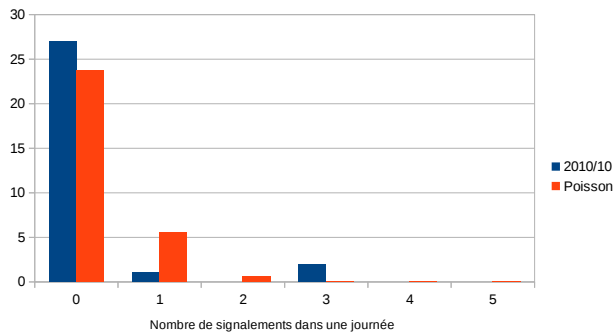
Cependant, si l'on considère le nombre de signalements par mois durant une période de plusieurs années comme Abraini ou le nombre de signalements par jour durant une période plus courte comme Berger, il est possible que les distributions ne soient pas identiques. Abraini signale qu'on tend vers la distribution de Poisson lorsque le nombre de signalements diminue. En principe dans le cas d'un phénomène purement aléatoire de type tirage de Bernouilli quelle que soit l'échelle de temps considérée la distribution devrait être identique.

Nous avons donc testé la distribution de Poisson de nos données belges concernant les fréquences du nombre de signalements par jour pour cinq mois choisis parmi les 60 de la période. Nous avons choisi le mois d'août 2012 qui comprend le nombre le plus élevé de signalements et quatre autres mois prélevés aléatoirement. Voici le tableau et les distributions de fréquences journalière de chacun de ces mois et la figure qui compare leur distribution avec la distribution de Poisson.

Tableau 3 : répartition des fréquences de 0 à 5 observations/jours pour quelques mois

| Mois | NbS | 0 | 1 | 2 | 3 | 4 | 5 | Total de jour/mois |
|----------|-----|----|----|---|---|---|---|--------------------|
| 2010/10 | 7 | 27 | 1 | 0 | 2 | 0 | 0 | 30 |
| 2012/08* | 22 | 16 | 10 | 3 | 2 | 0 | 0 | 31 |
| 2013/02 | 5 | 24 | 3 | 1 | 0 | 0 | 0 | 28 |
| 2013/05 | 3 | 28 | 3 | 0 | 0 | 0 | 0 | 31 |
| 2015/12 | 10 | 23 | 7 | 0 | 1 | 0 | 0 | 31 |

* 22 observations sont datées, la 23^e est signalée le 14 août mais pas précisément datée dans le mois



Probabilité d'une distribution de Poisson

| Mois | Test Khi ² |
|----------|-----------------------|
| 2010/10 | 1,44021E-016 |
| 2012/08* | 0,78043533 |
| 2013/02 | 0,8401651196 |
| 2013/05 | 0,9967804572 |
| 2015/12 | 0,1215029905 |

Figure 5 : distribution de fréquence du nombre de signalements

En dehors du mois d'octobre 2010, on ne peut pas écarter la probabilité d'une distribution de Poisson pour la fréquence du nombre de signalements dans une journée. Les probabilités sont bien supérieures au seuil habituel de 5 % et même à celui de 1 %.

Pour l'ensemble de notre série NbS, et donc le nombre de signalements mensuels, la distribution n'est pourtant pas semblable et ne suit probablement pas une distribution de Poisson, ni une distribution Normale mais probablement une distribution Gamma (Figure 5 – $p=0,7246$).

En statistique, on utilise Poisson pour la fréquence à temps constant d'un événement, Gamma est utilisé lorsque le temps est variable entre la survenance d'un événement et un autre (par exemple en industrie on utilise ce modèle pour l'événement « survenance d'une panne »).

Rappelons que le cadre de cet article n'est pas déterministe. Nous n'utiliserons pas ces résultats pour spéculer sur la nature du ou des phénomènes qui se trouvent décrits par notre variable NbS mais simplement comme une variable à observer et décrire empiriquement ou probabilistiquement.



2.3. Méthode COBEPS

La méthode proposée par Abraini est hybride et cette situation est probablement due en partie au nombre trop réduit de données dont il disposait. Finalement, Abraini, ne pouvant trancher entre une distribution normale ou de poisson, a utilisé la distribution de fréquence observée et le percentile 95 comme seuil. Il y ajoute ensuite la correction saisonnière déterminée par X12.

Nous avons vu que pour nos données, la procédure X12-Arima aboutit à des statistiques peu fiables même si le caractère saisonnier est indéniable son ajustement est délicat et de mauvaise qualité. La distribution générale des données si l'on considère la fréquence du nombre mensuel de signalements se rapproche de la distribution gamma alors que la fréquence du nombre journalier de signalements dans le mois est plutôt de type Poisson.

Nous proposons donc de considérer chaque mois NbS_{ij} comme une distribution de Poisson (ce qui semble le cas au niveau de la distribution journalière à l'intérieur d'un mois), de prendre la moyenne du nombre d'observations mensuelles du mois 'i' (\overline{NbS}_i) comme centre de gravité et de relever comme valeur seuil, le nombre de signalements qui théoriquement s'approchera de la borne supérieure de 1 %. Le seuil de 5 % est trop peu contraignant. Il s'agit de la méthode que nous avons appelée méthode COBEPS.

Voici l'exemple du calcul pour le mois de janvier

$$NbS_i = (0, 7, 13, 8, 6)$$

$$\overline{NbS}_i = 6,8$$

La probabilité se calcule par la formule suivante :
$$p_x = \frac{e^{-\overline{NbS}_i} \times \overline{NbS}_i^x}{x!}$$

Tableau 4 : calcul des probabilités pour les mois de janvier

| Nombre de signalements (x) | Probabilité d'occurrence selon Poisson (p _x) | Nombre de signalements (x) | Probabilité d'occurrence selon Poisson (p _x) |
|----------------------------|--|----------------------------|--|
| 0 | 0,00 | 13 | 0,01 |
| 1 | 0,01 | 14 | 0,01 |
| 2 | 0,03 | 15 | 0,00 |
| 3 | 0,06 | 16 | 0,00 |
| 4 | 0,10 | 17 | 0,00 |
| 5 | 0,13 | 18 | 0,00 |
| 6 | 0,15 | 19 | 0,00 |
| 7 | 0,15 | 20 | 0,00 |
| 8 | 0,13 | 21 | 0,00 |
| 9 | 0,10 | 22 | 0,00 |
| 10 | 0,06 | 23 | 0,00 |
| 11 | 0,04 | 24 | 0,00 |
| 12 | 0,02 | 25 | 0,00 |

Nous prenons, pour être très conservateur, la valeur de 'x' la plus élevée pour une probabilité proche de 1 % soit 14. Si NbS en janvier est supérieur à 14, le seuil d'alerte est atteint. La saisonnalité éventuelle est intégrée dans la variabilité intrinsèque du mois, d'année en année. Ce qui permet d'obtenir un schéma mensuel de seuils valable en 2016 et qui devra être réévalué annuellement.

Tableau 5 : valeurs de seuils par la méthode de Poisson 1 %

| Mois | NbS _{1%} |
|-----------|-------------------|
| Janvier | 14 |
| Février | 8 |
| Mars | 11 |
| Avril | 11 |
| Mai | 12 |
| Juin | 14 |
| Juillet | 16 |
| Aout | 24 |
| Septembre | 19 |
| Octobre | 11 |
| Novembre | 13 |
| Décembre | 15 |

3. Résultats

Au terme de ce travail nous allons comparer les résultats obtenus par calcul et extrapolation pour 2016 et par les deux méthodes présentées ci-dessus : la méthode Abraini, la méthode COBEPS. Nous allons adjoindre les prévisions du logiciel X12-ARIMA (Reg-ARIMA – avant désaisonnalisation – Annexe 1, tableau A1) et les NbS maxima observés durant les cinq années de données disponibles.

3.1. Seuil d'alerte

La figure 6 présente les seuils d'alertes prévisionnels. Pour rappel : pour Abraini, le percentile au seuil de 95 % et pour X12, les probabilités au seuil de 95 % ont été retenues et à 99 % pour COBEPS. Pour Abraini nous n'avons pas calculé un nouveau percentile pour chaque mois. Nous avons pris le percentile 95 % de l'ensemble de la distribution comme base. Sa valeur est de 15.

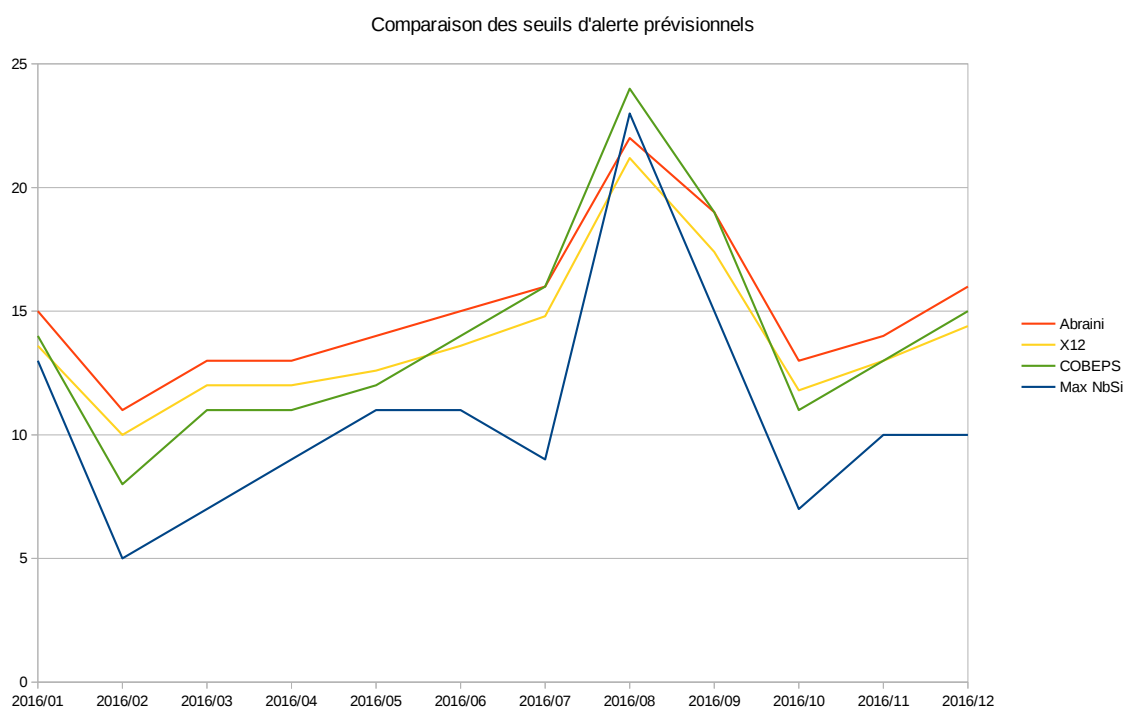


Figure 6 : seuil d'alerte prévisionnels pour 2016 comparaison des méthodes

Première constatation pour tous les mois, à l'exception du mois d'août pour les méthodes Abraini et X12, les seuils d'alertes sont supérieurs aux maxima obtenus après 5 années d'observations de NbS. Ce qui signifie que ces indicateurs sont assez « conservateurs », le plus conservateur étant l'indicateur de la méthode Abraini qui est au-dessus ou égal à tous les autres estimateurs pour onze mois sur douze. Ils garantissent cependant tous, que l'alerte ne sera pas donnée trop précocement. Pour les méthodes Abraini et X12, le mois d'août 2012 doit être considéré comme « anormal » et déclenche le système d'alerte. C'est le seul de la série des 60.

Tous les indicateurs respectent globalement la saisonnalité de NbS, toutefois les variations sont plus amorties pour les indicateurs Abraini et X12 (variance entre 8,71 et 8,14). La méthode COBEPS respecte davantage la structure de la variance observées de max. NbS_i tout en conservant une tendance centrale similaire aux autres indicateurs (moyenne et médianes du tableau 6).

Tableau 6 : tendance centrale et variabilité des seuils

| | Abraini | X12 | COBEPS | Max NbSi |
|-----------------|----------------|------------|---------------|-----------------|
| Moyenne | 15,08 | 13,86 | 14,00 | 10,83 |
| Médiane | 14,50 | 13,30 | 13,50 | 10,00 |
| Variance | 8,81 | 8,74 | 18,00 | 21,97 |

Déterminer le meilleur indicateur dépend des objectifs poursuivis et de la définition de ce que l'on appelle une vague d'ovni. Pour notre part, suite à l'expérience de la grande vague belge, et si l'on se contente des aspects quantitatifs, les vagues sont massives et prolongées. Massives : c'est-à-dire qu'elles présentent toutes, une rupture brutale du nombre de signalements (sur quelques semaines parfois quelques jours). À cette fin, la principale qualité d'un indicateur est d'être conservateur. Ils le sont tous, même si celui d'Abraini l'est davantage sur l'ensemble d'une année. Prolongée : c'est-à-dire qu'elle se poursuit sur plusieurs mois. C'est pourquoi, il nous semble important qu'elle suive au plus près les éventuelles variabilités, y compris celles qui résultent d'un sursaut de signalements ponctuels dus à la saison et/ou à des événements de type fortuit comme une rentrée atmosphérique signalée massivement. Il apparaît de toute façon nécessaire de ne pas déclencher une alerte sans regarder la distribution journalière. Lorsque la grande majorité des signalements concerne un seul et même moment, il est probable que l'alerte n'est pas nécessaire et qu'une explication conventionnelle du phénomène soit donnée rapidement. Il est donc important aussi que l'indicateur suive étroitement la variabilité des signalements mais pour autant l'alerte ne doit être donnée que lorsque le seuil est franchi sur minimum deux mois.

Enfin, l'indicateur doit aussi être en phase avec l'histoire connue de fréquence des vagues du phénomène. Abraini signale que choisir un niveau de probabilité de 1 % revient à dire qu'une vague revient une fois tous les cent mois ; une fois tous les vingt mois pour une probabilité de 5 %. Il apparaît, en tous les cas en Belgique, que les vagues ne se conforment pas à ces fréquences. On peut relever une vague en 1954, une autre en 1971-1972 et une troisième en 1989-1993 (Boitte, 2014). Les autres événements ne sont que des frémissements par rapport à ces périodes. La fréquence est donc plutôt de l'ordre d'une vague tous les 240 mois. Il est donc recommandé de choisir un seuil de probabilité inférieur à 5 %.

Signalons que les indicateurs probabilistes comme celui du COBEPS sont probablement plus adéquats que ceux qui sont basés sur le principe d'auto-corrélation telle la méthode X12. Ils se calculent donc tenant compte d'une loi de distribution et pas d'un lien possible entre des observations.

En conclusion, la méthode Abraini paraît tout à fait acceptable. Mais, étant donné la complexité de sa mise en oeuvre et le fait que pour les données belges, les statistiques X12 ne sont pas de bonne qualité, nous retenons l'indicateur COBEPS comme le plus intéressant.

3.2. Indicateur mensuel d'activité ufologique

Nous retenons l'indicateur COBEPS, pour lequel nous avons calculé les probabilités d'occurrences de n signalements mensuels sur la base de la distribution de Poisson, comme au tableau 4. La figure 7 montre la distribution pour le mois d'août. Nous retiendrons que l'activité est **faible** lorsque $p \approx 0,05$ pour les nombres d'observations situés à gauche de la première ligne rouge sur le graphique, qu'elle est **forte** pour les nombres situés à droite de la seconde ligne rouge et **normale** entre les deux. L'activité dépasse le seuil d'**alerte** à droite de la ligne orange ($p < 0,01$, $n = 24$).

La figure 8 montre les domaines de l'activité ufologique selon la répartition décrite ci-dessus et pour les 12 mois de 2016. Le tableau 7 montre les chiffres des limites.

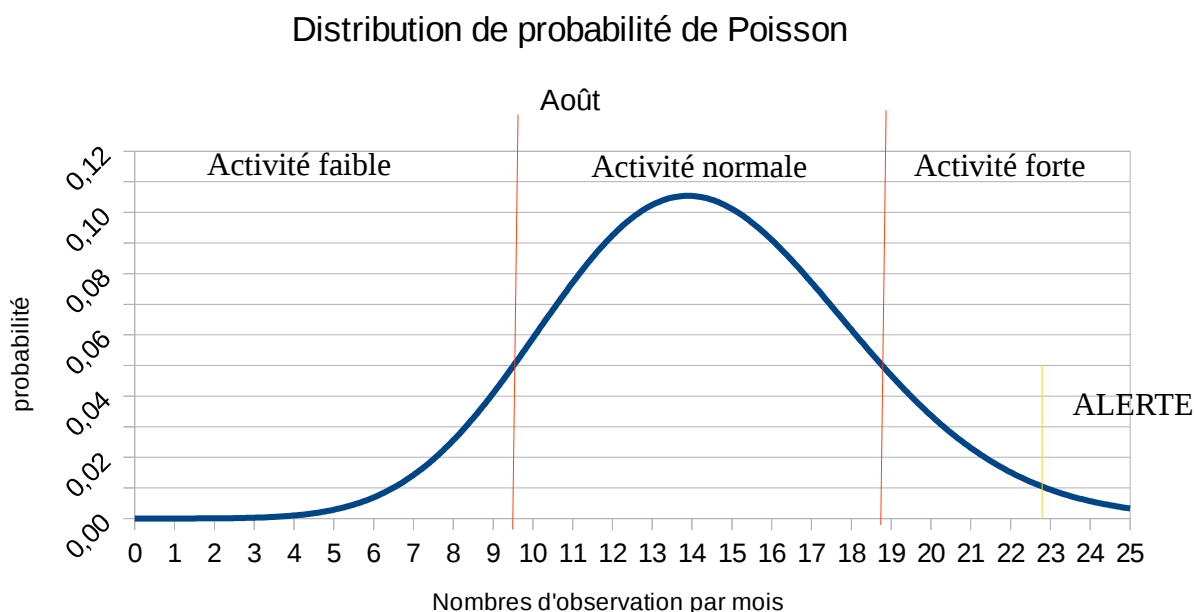


Figure 7 : bornes de probabilité $p \approx 0,05$ (5%) pour la distribution de Poisson pour le mois d'août

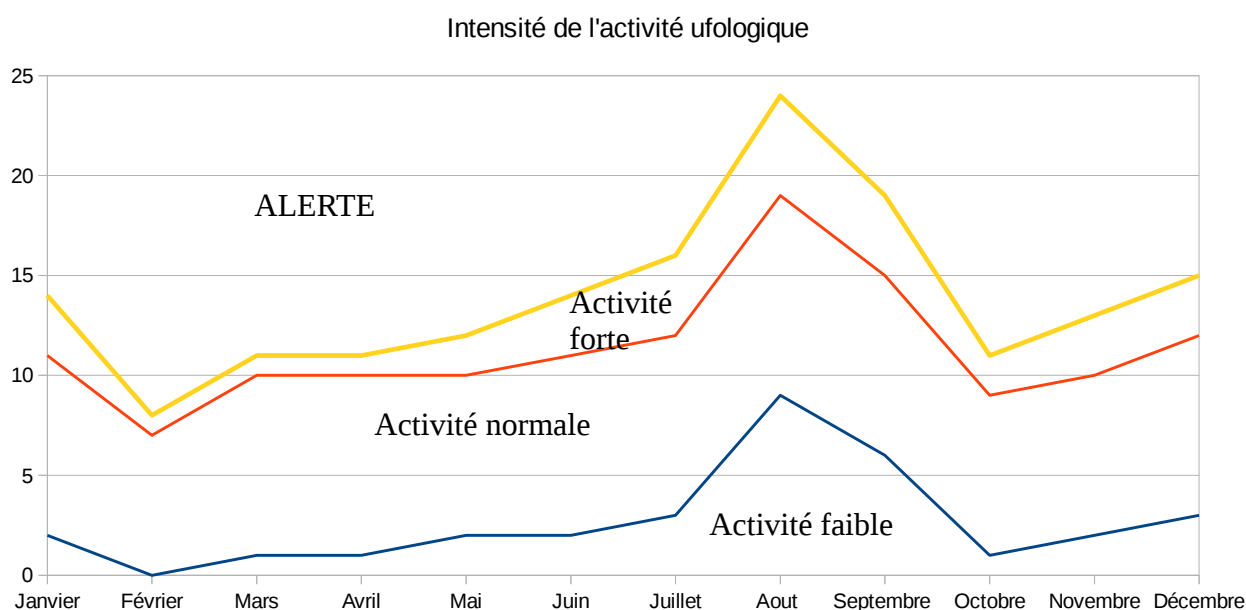


Figure 8 : indicateurs d'activité ufologique

Tableau 7 : chiffres des limites

| Mois | Faible | Forte | Mois | Faible | Forte |
|---------|--------|-------|-----------|--------|-------|
| Janvier | 2 | 11 | Juillet | 3 | 12 |
| Février | 0 | 7 | Aout | 9 | 19 |
| Mars | 1 | 10 | Septembre | 6 | 15 |
| Avril | 1 | 10 | Octobre | 1 | 9 |
| Mai | 2 | 10 | Novembre | 2 | 10 |
| Juin | 2 | 11 | Décembre | 3 | 12 |

Conclusion

On ne peut pas encore ni décrire ni prévoir les signalements d'ovni par les méthodes de séries temporelles. Le gros problème est la grande variabilité de NbS au fil du temps, à des saisons imparfaitement marquées d'une année à l'autre et à l'amplitude des irrégularités qui sont généralement équivalentes à la composante saisonnière. Ceci est certainement dû au caractère agrégé de NbS qui rassemble plusieurs phénomènes. Il est probable qu'il faudra attendre encore quelques années afin d'obtenir suffisamment de données pour augmenter la qualité des statistiques. D'autre part, les distributions journalières au sein d'un mois sont bien conformes à la loi de Poisson pour les signalements en Belgique francophone. Nous avons vu que l'on peut aussi définir des seuils et des niveaux en utilisant cette distribution de probabilité.

Nous avons calculé les prévisions du nombre de signalements mensuels pour 2016 à partir de la méthode préconisée par Abraini et celle dite du COBEPS. Ces deux méthodes permettent effectivement de définir des chiffres mensuels et, en les comparant aux maxima mensuels observés, on peut constater leur caractère conservateur et saisonnier similaire à la variable NbS. Etant donné la fragilité des statistiques de notre série temporelle NbS, nous préconisons l'utilisation de la méthode COBEPS. Elle sera utilisée pour mettre en ligne un tableau de bord de l'activité ufologique en Belgique francophone. Il nous semble aussi que l'alerte ne peut être significative si l'ensemble des signalements porte sur un événement ponctuel (dans le temps). Elle ne serait effective que si les seuils sont dépassés pendant deux mois consécutifs et pour une bonne distribution des observations à l'intérieur des mois concernés.

Remerciements

- Aux membres du COBEPS en particulier Eddy Louchez & Vincent De Bareaecker pour Leurs vérifications statistiques et relecture des contenus.
- A Franck Boitte pour sa relecture et suggestions
- A Giorgio Abraini pour son travail et les réponses à mes questions.
- A Messieurs Eduardo Russo et Vicente Juan Ballester-Olmos pour les relais vers les textes originaux et les auteurs.

Bibliographie

- Abraini giorgio, *How to monitor UFO waves*, communication personnelle, rédigé en 2005
- Ballester-Olmos V.J. , "On UFO and the Poisson distribution", *UFO Phenomena*, II.1, 1977
- Berger E., "Are UFOs Poisson-distributed?", *UFO Phenomena*, I.1, 1976
- Boitte Franck, *Catalogue des Observations Belge*, communication personnelle, 11 septembre 2011
- Box, George; Jenkins, Gwilym (1970). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
- Cottrell Allin (Wake Forest University), Lucchetti Riccardo "Jack" (Università Politecnica delle Marche), <http://gretl.sourceforge.net/> , 2016
- Droesbeke Jean-Jacques, Dehon Catherine, Vermandele Catherine, *Eléments de statistique*, Collection : Statistique et mathématiques appliquées, 2015
- Gomez and Maravall, *Automatic Modeling Methods for Univariate Series" - A Course in Time Series (Edited by D. Pena, G. C. Tiao, R. S. Tsay)*, New York : J. Wiley and Sons, 2000
- Ladiray Dominique et Quenneville Benoît, *Comprendre la méthode x11*, 1999, <http://www.census.gov/ts/papers/x11doc.pdf>
- Nicolau Florence, *Statistique : Introduction aux séries chronologiques, support de cours et de travaux pratiques sur les séries chronologiques*, Menton, France, 2008 <http://www.nicolau.nom.fr/publications.html>
- Office for National Statistics of United Kingdom, *Guide to Seasonal Adjustment with X-12-ARIMA **DRAFT***, 2007, disponible à l'adresse www.ons.gov.uk/ons/guide.../time.../guide-to-seasonal-adjustment.pdf .
- Roustant o., *Introduction aux séries chronologiques – Axes, méthodes statistiques et applications*, Ecole Nationale Supérieure des Mines de Saint-Etienne, 2008 http://www.emse.fr/~roustant/documents/polycopie_serie_temporelles_2008_2009.pdf
- U.S. Census Bureau, *X12-Arima Manual*, 2011 <https://www.census.gov/ts/x12a/v03/x12adocV03.pdf>