

**Evaluation de critères d'alerte
pour la détection et l'investigation
des épidémies de salmonelloses
communautaires**

Christophe BILLY

Cette étude a été réalisée par :

Le Réseau National de Santé Publique :

Ch. BILLY
JC. DESENCLOS
D. COULOMBIER
Ph. QUENEL
V. VAILLANT
E. DELAROCQUE - ASTAGNEAU

Le Centre National des Salmonelles et des Shigelles :

Ph. BOUVET
P.A.D. GRIMONT

Cette étude a fait l'objet d'un mémoire de DEA (Epidémiologie et Intervention en Santé Publique, Universités Bordeaux II - Tours) soutenu par Christophe Billy le 11 septembre 1998.

Cette étude a bénéficié du soutien de la Fondation pour la Recherche Médicale.

Avertissement au lecteur : les sorties écrans du logiciel SSS1 sont en anglais ; les semaines y sont indiquées en "décalages" ou "lags" chronologiques comptabilisées à partir de la date d'origine fixée au 01/01/1992 (semaine du 01/01/92 au 07/01/92 = 9201 = décalage 1).

Liste des figures et tableaux

- Figure 1** Séries chronologiques hebdomadaires correspondant au profil 1 des sérotypes Paratyphi B (PB), Brandenburg (BB) et Goldcoast (GC) pour la période 1992-1996.
- Figure 2** Séries chronologiques hebdomadaires correspondant au profil 2 des sérotypes Typhimurim (TM) et Hadar (HR) pour la période 1992-1996.
- Figure 3** Séries chronologiques hebdomadaires correspondant au profil 3 des sérotypes Dublin (DU), Infantis (IF) et Indiana (ID) pour la période 1992-1996.
- Figure 4** Graphique de la moyenne – étendue de la série originale *S. Paratyphi B* (SPB) pour la période 1992-1996.
- Figure 5** Superposition de la série *S. Paratyphi B* (SPB) et de la série écrêtée (SPBe) pour la période 1992-1996.
- Figure 6** Graphique de la moyenne – étendue de la série écrêtée *S. Paratyphi B* (SPBe) pour la période 1992-1996.
- Figure 7** Série écrêtée (SPBe) sur les 261 semaines de la période 1992-1996 (261 décalages)
- Figure 8** Graphes de la fonction d'autocorrélation (ACF) et de la fonction d'autocorrélation partielle (PACF) de la série SPBe.
- Figure 9** Série transformée Y_t (SPBe après différenciation d'ordre 52).
- Figure 10** Graphes des fonctions d'autocorrélations (ACF, PACF) de la série Y_t .
- Figure 11** Série transformée Z_t (SPBe après double différenciation d'ordres 1 et 52).
- Figure 12** Fonctions d'autocorrélation (ACF, PACF) de tendance (Trns) et de saison (Trns seasonal) de la série Z_t .
- Figure 13** Fonctions d'autocorrélation (ACF, PACF) des résidus
- Figure 14** Ajustement du modèle SARIMA (0.1.1)x(0.1.1) à la série écrêtée (SPBe) pour la période 1992-1996.
- Figure 15** Superposition de la série écrêtée SPBe et des prédictions moyennes du modèle SARIMA (0.1.1)x(0.1.1) sur la période juillet 1993 à décembre 1996.
- Figure 16** Série originale SPB et seuil Box & Jenkins à 95 % pour *S. Paratyphi B* sur la période juillet 1993 – décembre 1996.
- Figure 17** Série originale SPB et tendance linéaire estimée pour la période 1992-1996.
- Figure 18** Périodogramme de la série *S. Paratyphi B* (SPB)
- Figure 19** Modèles estimés par analyse spectrale, avant et après retrait des points épidémiques de la série SPB pour la période 1992-1996.
- Figure 20** Série originale *S. Paratyphi B* (SPB) et seuil Serfling à 95 % pour la période 1992-1996.
- Figure 21** Seuil CNR, nombres hebdomadaires et cumulés mensuels de souches *S. Paratyphi B* pour la période juillet 1993 – juin 1995.
- Figure 22** Seuil Box & Jenkins et seuil Serfling superposés au nombre hebdomadaire observé de souches *S. Paratyphi B* pour la période juillet 1993 – juin 1995.
- Figure 23** Séries hebdomadaires de souches *S. Paratyphi B* reçues au CNR pour les régions Poitou-Charentes, Pays-de-la-Loire, Ile-de-France et Rhône-Alpes en 1992-1996.
- Figure 24** Séries hebdomadaires de souches *S. Paratyphi B* reçues au CNR par classes d'âge pour la période 1992-1996

- Figure 25** Graphique de la moyenne étendue de la série originale *S. Typhimurium* (STM) pour la période 1992-1996.
- Figure 26** Série originale de *S. Typhimurium* (STM) pour les 261 semaines de la période 1992-1996 (261 décalages)
- Figure 27** Graphes de la fonctions d'autocorrélation (ACF) et de la fonction d'autocorrélation partielle (PACF) de la série STM.
- Figure 28** Série transformée Y_t .
- Figure 29** Graphes des fonctions d'autocorrélations (ACF, PACF) de la série Y_t .
- Figure 30** Série transformée Z_t .
- Figure 31** Fonctions d'autocorrélation (ACF, PACF) de tendance (Trns) et de saison (Trns seasonal) de la série Z_t .
- Figure 32** Fonctions d'autocorrélation (ACF, PACF) des résidus.
- Figure 33** Ajustement du modèle SARIMA (1.1.1)x(0.1.1) à la série originale *S. Typhimurium* (STM).
- Figure 34** Comparaison de la série *S. Typhimurium* observée et des prédictions moyennes du modèle SARIMA pour l'année 1997.
- Figure 35** Série originale et seuil Box & Jenkins à 95 % pour *S. Typhimurium* sur la période 1994-1997
- Figure 36** Série originale *S. Typhimurium* (STM) et tendance linéaire estimée pour la période 1992-1996.
- Figure 37** Périodogramme de la série STM
- Figure 38** Modèles estimés par analyse spectrale, avant et après retrait des points épidémiques de la série STM pour la période 1992-1996.
- Figure 39** Série originale *S. Typhimurium* et seuil Serfling à 95 % pour la période 1992-1997.
- Figure 40** Seuil CNR et nombre cumulé mensuel de souches *S. Typhimurium* par semaine pour la période 1996-1997.
- Figure 41** Seuil Box & Jenkins et seuil Serfling superposés au nombre hebdomadaire observé de souches *S. Typhimurium* pour la période 1996-1997.
- Figure 42** Séries hebdomadaires de souches *S. Typhimurium* reçues au CNR pour les régions Aquitaine, Poitou-Charentes, Ile-de-France, Centre et Franche-Comté, en 1992-1997.
- Figure 43** Séries hebdomadaires de souches *S. Typhimurium* reçues au CNR pour le département du Jura (dept 39) en 1992-1997.
- Figure 44** Séries hebdomadaires de souches *S. Typhimurium* reçues au CNR pour le département du Lot-et-Garonne (dept 47) en 1992-1997.
- Figure 45** Séries hebdomadaires de souches *S. Typhimurium* reçues au CNR par classe d'âge pour la période 1992-1997.
- Figure 47** Superposition de la série originale *S. Brandenburg* (SBB) et de la série écrêtée (SBBe) sur la période 1992-1996.
- Figure 48** Superposition de la série originale *S. Goldcoast* (SGC) et de la série écrêtée (SGCe) sur la période 1992-1993.
- Figure 49** Comparaison de la série originale *S. Hadar* (SHR) et de la série originale après retrait de la tendance (SHRe) sur la période 1992-1996.
- Figure 50** Périodogramme de la série *S. Brandenburg* (SBB) pour la période 1992-1996.

- Figure 51** Seuil Serfling à 95 % pour la série *S. Brandenburg* sur la période 1994-1995.
- Figure 52** Seuil CNR pour *S. Brandenburg* sur la période 1994-1995.
- Figure 53** Périodogramme de la série *S. Goldcoast (SGC)* pour la période 1992-1996.
- Figure 54** Seuil Serfling à 95 % pour la série *S. Goldcoast* sur la période 1994-1995.
- Figure 55** Seuil CNR pour *S. Goldcoast* sur la période 1994-1995.
- Figure 56** Périodogramme de la série *S. Hadar (SHR)* pour la période 1992-1996.
- Figure 57** Seuil Serfling à 95 % pour la série *S. Hadar* sur la période 1994-1995.
- Figure 58** Seuil CNR pour *S. Hadar* sur la période 1994-1995.

- Tableau I** Points épidémiques détectés et remplacés sur la série originale SPB
- Tableau II** Diagnostic des paramètres et matrice des coefficients de corrélation
- Tableau III** Diagnostic des résidus
- Tableau IV** Tests statistiques sur les résidus
- Tableau V** Paramètres de l'analyse spectrale pour la série *S. Paratyphi B*
- Tableau VI** Paramètres optimisés pour l'analyse spectrale de la série *S. Paratyphi B*.
- Tableau VII** Nombre de souches de *S. Paratyphi B* reçues au CNR sur la période 94-34 à 94-48 par région et par département.
- Tableau VIII** Nombre de souches de *S. Paratyphi B* reçues au CNR sur la période 93-33 à 93-47 par région et par département.
- Tableau IX** Diagnostic des paramètres et matrice des coefficients de corrélation
- Tableau X** Diagnostic des résidus.
- Tableau XI** Tests statistiques sur les résidus
- Tableau XII** Paramètres de l'analyse spectrale pour la série *S. Typhimurium*
- Tableau XIII** Paramètres optimisés pour l'analyse spectrale de la série *S. Typhimurium*
- Tableau XIV** Résultats des critères de jugement pour les 3 sérotypes *S. Brandenburg*, *S. Goldcoast* et *S. Hadar*.
- Tableau XV** Sensibilité (Se) et valeur prédictive positive (vpp) estimées pour les seuils CNR, Box & Jenkins et Serfling en fonction de la référence "brute".
- Tableau XVI** Sensibilité (Se) et valeur prédictive positive (vpp) estimées pour les seuils CNR, Box & Jenkins et Serfling en fonction de la référence "corrigée".

SOMMAIRE

RESUME	11
1 – INTRODUCTION	13
1-1 Contexte : les épidémies communautaires de salmonelloses	13
1-2 Justification de l'étude	13
2 – OBJECTIFS	15
2-1 Objectif général	15
2-2 Objectifs spécifiques	15
3 – MATERIELS ET METHODES	17
3-1 Schéma général de l'étude	17
3-2 Sources d'informations	17
3-3 Matériel	17
3-4 Elaboration des critères d'alerte (signaux)	21
3-5 Choix d'une référence pour la comparaison des seuils	33
3-6 Outils d'analyse statistique	34
4 – RESULTATS	35
4-1 Résultats pour <i>S. Paratyphi B</i>	35
4-2 Résultats pour <i>S. Typhimurium</i>	56
4-3 Généralisation des seuils pour les autres sérotypes	75
4-4 Evaluation des critères de jugement en fonction de la référence "corrigée"	83
5 – DISCUSSION	87
5-1 Résultats principaux	87
5-2 Référence "brute" versus référence "corrigée"	88
5-3 Qualité des données	89
5-4 Choix des modèles, avantages / inconvénients	89
5-5 Propositions	91
6 – CONCLUSION	95
REFERENCES	97
ANNEXES	101

RESUME

Introduction

Les épidémies communautaires d'infections à salmonelles non typhiques peuvent être rapidement diffusées et avoir des conséquences graves sur la santé humaine. La détection de ces épidémies et le déclenchement de l'alerte reposent sur le système de surveillance du Centre National de Référence des Salmonelles et des Shigelles (CNR) qui réalise le sérotypage des souches de salmonelles envoyées par les laboratoires d'analyses biologiques et médicales (LABM). Lorsqu'une alerte épidémique est déclarée par le CNR, le Réseau National de Santé Publique (RNSP) conduit les enquêtes épidémiologiques pour identifier les facteurs favorisant l'épidémie et recommander les mesures à prendre. Cette étude a pour objectif d'explorer les performances de critères décisionnels pour détecter et investiguer les épidémies communautaires à salmonelles en France.

Méthodes

L'étude a évalué rétrospectivement la validité de signaux ou seuils d'alerte épidémique hebdomadaires établis systématiquement versus des semaines classées épidémiques ou non épidémiques. Les données de surveillance du CNR ont été croisées avec les résultats des investigations d'épidémies du RNSP pour cinq sérotypes de salmonelles (Paratyphi B, Goldcoast, Brandenburg, Typhimurium, Hadar) sur la période 1992-1997. Les séries chronologiques correspondant à chaque sérotype ont été reconstituées sur la base de la date de réception des souches de salmonelles au CNR. Trois approches méthodologiques ont été utilisées pour construire les seuils d'alerte : la méthode des "moyennes historiques" (seuil CNR) utilisée en routine par le CNR, la méthode Box et Jenkins (seuil Box & Jenkins) et la méthode Serfling (seuil Serfling). Des analyses géographiques et par tranches d'âge ont été détaillées pour repérer a posteriori des épisodes épidémiques qui seraient passés inaperçus. La comparaison des seuils a été réalisée selon deux références successives : une référence "brute" définie a priori, sur la base des épidémies connues et investiguées ; une référence "corrigée" définie sur la base des épidémies supplémentaires repérées par les analyses complémentaires temps, lieu, personnes. Les critères de jugement ont porté sur la sensibilité, la valeur prédictive positive (vpp), la réactivité et en envisageant des délais d'alerte variables de une à plusieurs semaines.

Résultats

Un seuil Serfling et un seuil Box & Jenkins ont pu être élaborés respectivement pour les cinq séries de sérotypes et pour deux séries de sérotypes (Paratyphi B, Typhimurium). Le seuil CNR mensuel obtenu pour les cinq sérotypes a été reconstruit à un format hebdomadaire comparable. Les analyses complémentaires ont permis de repérer pour les sérotypes Paratyphi B et Typhimurium des épidémies qui étaient passées inaperçues. Les performances des seuils sont très variables d'un sérotype à l'autre et semblent dépendre du profil général de chaque série. Néanmoins, tous sérotypes confondus, les trois seuils ont montré des performances équivalentes. Le seuil Box & Jenkins est apparu le plus précis (vpp = 0.73 à 1). Les seuils CNR et Serfling ont les qualités les plus similaires en sensibilité et vpp, avec un avantage au seuil Serfling pour ses capacités de réactivité et de stabilité de détection malgré le changement de référence.

Conclusion

L'élaboration d'un algorithme décisionnel requiert la combinaison de plusieurs critères prenant en compte les dimensions temps, lieu et personnes. Cette étude expose les avantages et les limites de disposer de seuils d'alerte quantitatifs et la nécessaire association à des critères qualitatifs.

Mots-clefs : salmonelles, épidémies communautaires, séries chronologiques, Box et Jenkins, Serfling, santé publique, seuils d'alerte épidémique.

1 - INTRODUCTION

1-1 Contexte : les épidémies communautaires de salmonelloses

Les salmonelloses non typhiques, dites mineures par opposition à la fièvre typhoïde, sont les premières causes de maladies d'origine alimentaire dans les pays développés (1). Leur nombre est en constante augmentation depuis une quinzaine d'années (2, 3).

Ce sont des infections à bactéries entéropathogènes invasives du genre *Salmonella*. Elles peuvent se manifester de façon sporadique sous la forme de cas isolés de gastro-entérite fébrile, parfois responsable d'hospitalisation et de décès (4, 5). Elles sont responsables de toxi-infections alimentaires collectives (TIAC) et représentent en France plus de 80% des cas de TIAC déclarés (6). Elles peuvent aussi être à l'origine d'épidémies communautaires diffuses (foyers diffus) pouvant toucher plusieurs départements ou régions (7), voire plusieurs pays (8). Ces foyers épidémiques dépassent fréquemment les 300 cas (9). Ils correspondent habituellement à la contamination d'un aliment distribué à grande échelle, le plus souvent contaminé lors de la production, la préparation ou le conditionnement (10). Dans d'autres situations, plus rares, il s'agit de la contamination d'une filière de production toute entière : *Salmonella* sérotype Enteritidis (oeufs) (11, 12), *Salmonella* sérotype Hadar (volaille) (13, 14).

La surveillance au niveau national par le Centre National de Référence des Salmonelles et des Shigelles de l'Institut Pasteur (CNR) permet de détecter les épidémies communautaires qui ne sont pas identifiées au niveau régional (6, 10). Des enquêtes épidémiologiques sont alors menées par le Réseau National de Santé Publique (RNSP) pour mettre en évidence la source de contamination et le véhicule ainsi que certains facteurs favorisants. Leurs résultats sont indispensables pour cibler au mieux les mesures de contrôle et améliorer la prévention.

Le CNR répond à toutes les demandes de sérotypage de souches de salmonelles isolées et provenant des laboratoires d'analyses biologiques et médicales (LABM). Environ un tiers des LABM publics et privés envoient volontairement leurs souches accompagnées de renseignements épidémiologiques (15). Parmi les 18 000 souches d'origine humaine enregistrées en 1995, plus de la moitié (9892, 52%) ont été reçues au CNR et sérotypées, le reste provenant d'informations épidémiologiques sur les souches complètement étudiées par les laboratoires collaborateurs. Grâce au système de surveillance exercée par le CNR, sept épidémies régionales ou nationales ont été détectées entre 1993 et 1997 (*Salmonella* sérotypes Typhimurium, Virchow, Infantis, Brandenburg, Goldcoast, Hadar, Dublin) (6). L'exhaustivité de ce système est de 50% pour les TIAC à salmonelles (16). Parmi les souches de salmonelles analysées par le CNR, 99 % proviennent de cas isolés de salmonelloses non impliqués dans les petites épidémies survenant dans un contexte familial ou de restauration collective (17).

1-2 Justification de l'étude

Les investigations pour rechercher l'aliment responsable sont lourdes. Elles nécessitent des moyens importants, doivent être conduites avec la plus grande rigueur et faire l'objet de protocoles standardisés et coordonnés. Leur efficacité dépend également de la collaboration large de partenaires multiples (Directions Départementales des Affaires Sanitaires et Sociales (DDASS), Direction des Services Vétérinaires (DSV), Direction Générale de la Santé (DGS), ministère de l'Agriculture, CNR, RNSP) (10).

La finalité de la surveillance est l'application à des actions de prévention (18). S'il est indispensable de détecter ces épidémies communautaires très précocement pour espérer limiter la morbidité au maximum, il n'est pas souhaitable de déclencher une procédure d'alerte inutilement. De 1993 à 1995 des seuils épidémiques mensuels ont été construits à partir de l'approche Box et Jenkins pour un nombre limité de sérotypes (17, 19, 20). Pour les autres sérotypes l'alerte est fondée sur le repérage visuel d'une augmentation qui semble inhabituelle. Sur cette base de fonctionnement, le délai entre le début de l'épidémie et sa détection varie de 3 à 8 semaines (10). Ce délai correspond au fonctionnement même du CNR (délai entre les dates d'isolement, d'envoi de la souche et de sérotypage) auquel il faut ajouter un recul suffisant pour la constatation de l'augmentation du nombre de souches avant de déclencher une investigation. Il est apparu en particulier que le seuil de détection mensuel avait une réactivité insuffisante, ce qui était en faveur d'une unité de temps d'analyse hebdomadaire (21).

Des critères d'alerte sensibles sont donc nécessaires pour détecter rapidement un état épidémique, raccourcir le délai de déclaration effective d'alerte épidémique et déclencher une investigation épidémiologique précoce (19). Cependant des critères sensibles mais non spécifiques peuvent amener à détecter des fausses épidémies (faux positifs) et entraîner des enquêtes inutiles. Les critères d'alerte doivent donc non seulement se fonder sur des seuils de détection mais aussi sur d'autres critères décisionnels. La définition d'un algorithme prenant en compte un ensemble de critères plus larges, et pas seulement sur le nombre arithmétique de cas, devrait permettre de rationaliser la décision de l'investigation d'une épidémie.

2 - OBJECTIFS

2-1 Objectif général

Explorer les critères potentiels d'un algorithme décisionnel pour la détection et l'investigation des épidémies communautaires à salmonelles non typhiques.

2-2 Objectifs spécifiques

- Analyser trois différents signaux d'alerte générés par trois approches méthodologiques (Box et Jenkins, Serfling et moyennes historiques) ;

- Déterminer les plus performants en terme de sensibilité, spécificité, valeur prédictive positive, réactivité et prédiction de la fin d'épidémie.

3 - MATERIELS ET METHODES

3-1 Schéma général de l'étude

Il s'agit d'une étude de validité rétrospective sur les données de surveillance de salmonelles du CNR de 1992 à 1997. L'étude a évalué trois types de signaux hebdomadaires établis systématiquement versus des semaines classées épidémiques ou non épidémiques.

3-2 Sources d'informations

3-2-1 Les données de surveillance

Le Centre National de Référence des Salmonelles et des Shigelles de l'Institut Pasteur (CNR) reçoit pour sérotypage les souches de salmonelles des laboratoires d'analyses biologiques et médicales (LABM), et assure la surveillance de l'augmentation inhabituelle d'un sérotype. Les données de surveillance des souches de salmonelles humaines couvrant la période de 1992 à 1996-1997 ont été obtenues du CNR. Pour chaque souche de salmonelle les renseignements épidémiologiques ont été relevés sur la fiche standardisée du CNR : sérotype, date d'isolement au laboratoire et date de réception au CNR, type de prélèvement (sang, selles, LCR ...), groupe d'âge du malade, département du laboratoire (annexe 1).

3-2-2 Les données d'investigations

Le Réseau National de Santé Publique (RNSP) est chargé de mener les enquêtes épidémiologiques en collaboration avec les DDASS, en vue de l'identification du véhicule et de la source de l'épidémie. Chaque alerte du CNR donnant lieu à une investigation fait l'objet d'un rapport par le RNSP. Les caractéristiques de chaque épidémie investiguée ont été recueillies à partir des rapports d'enquêtes du RNSP.

La définition opérationnelle retenue pour considérer une observation comme "épidémique" était l'augmentation importante du nombre de cas d'une salmonellose d'un sérotype particulier, sur une période de temps et une zone géographique données, par rapport à ce qui était attendu. Lors des sept dernières années, une vingtaine d'épidémies nationales ou régionales ont été détectées par le CNR. De 1993 à 1997, une dizaine de ces épidémies ont fait l'objet d'une enquête épidémiologique approfondie, menée par le RNSP en relation avec les médecins inspecteurs de Santé Publique (MISP) des départements concernés et le CNR.

3-3 Matériel

Les informations épidémiologiques ont été relevées pour huit sérotypes de salmonelles ayant donné lieu à des épidémies investiguées :

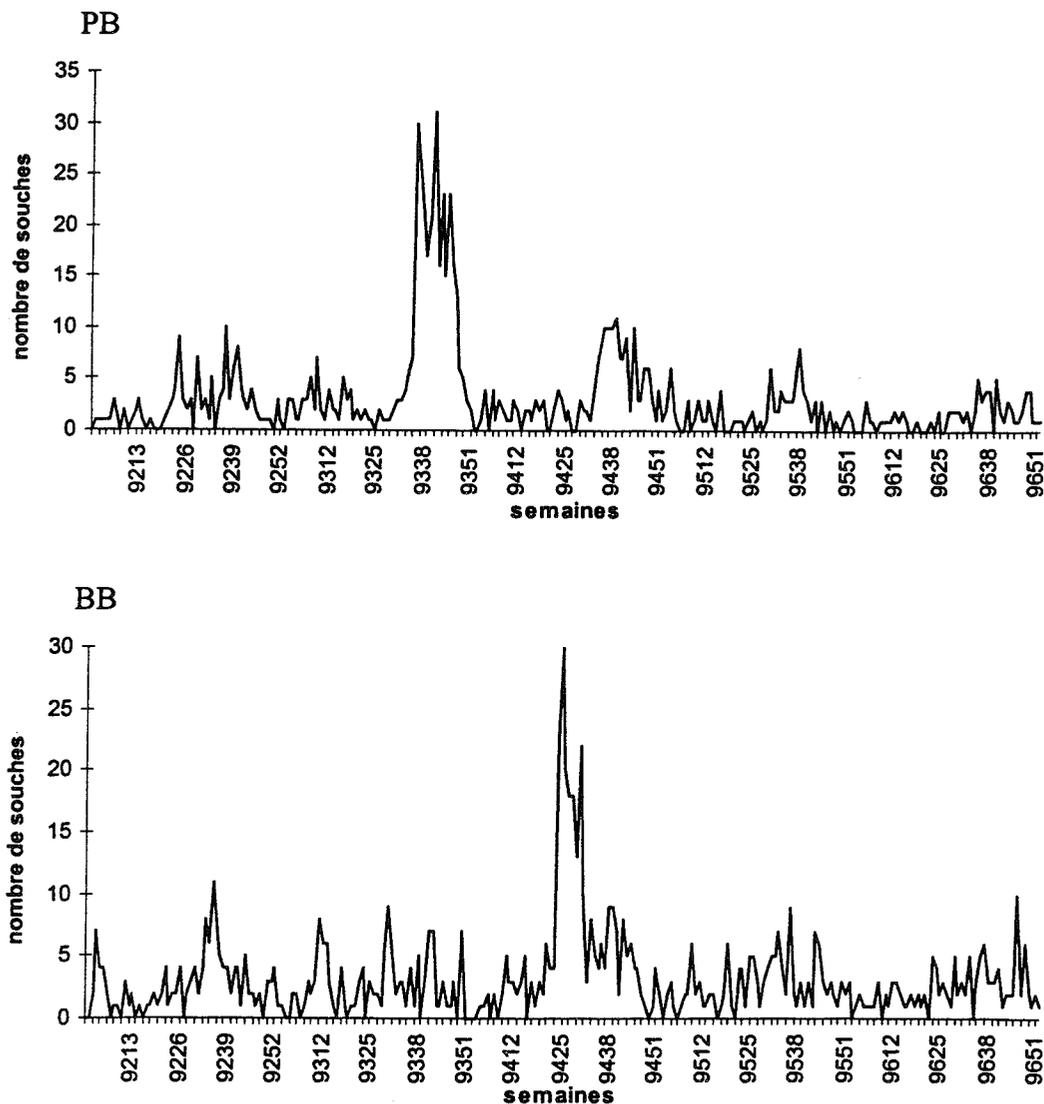
- *Salmonella* sérotype Brandenburg (*S. Brandenburg*),
- *Salmonella* sérotype Dublin (*S. Dublin*),
- *Salmonella* sérotype Goldcoast (*S. Goldcoast*),
- *Salmonella* sérotype Hadar (*S. Hadar*),
- *Salmonella* sérotype Indiana (*S. Indiana*),

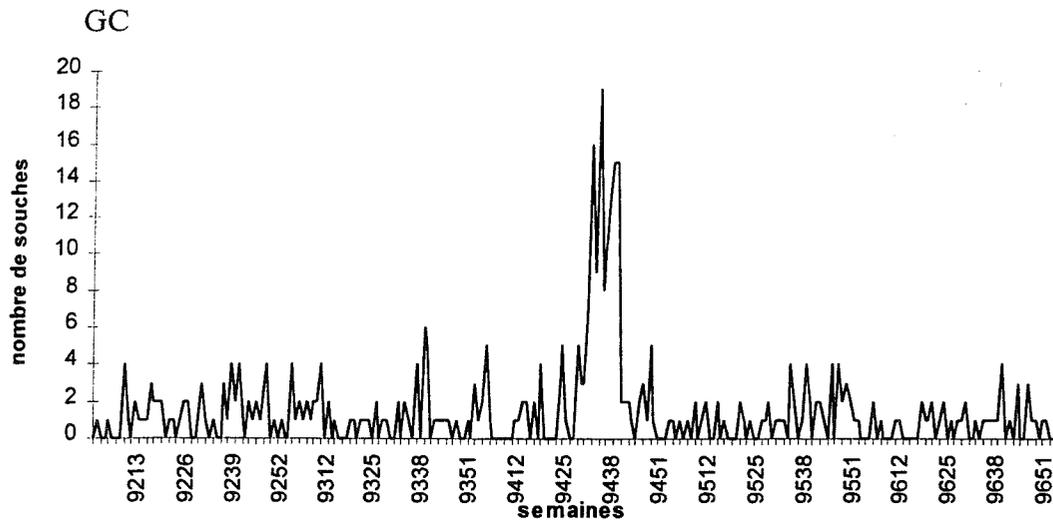
- *Salmonella* sérotype Infantis (*S. Infantis*),
- *Salmonella* sérotype Paratyphi B (*S. Paratyphi B*),
- *Salmonella* sérotype Typhimurium (*S. Typhimurium*).

Les séries chronologiques correspondant à chaque sérotype ont été reconstituées sur la base de la date de réception des souches au CNR. L'observation visuelle des huit courbes a permis de repérer des caractéristiques communes. Trois types de profils chronologiques ont été différenciés.

• Profil 1 (figure1) : profil avec saisonnalité et sans tendance, qui regroupe les sérotypes Paratyphi B (PB), Brandenburg (BB), Goldcoast (GC). Ces séries montrent une ligne de base accidentée dont l'ensemble ne présente pas de tendance particulière sur le long terme. Une saisonnalité plus ou moins marquée est repérable visuellement. Un pic se détache du reste de la courbe correspondant à l'épidémie connue et investiguée.

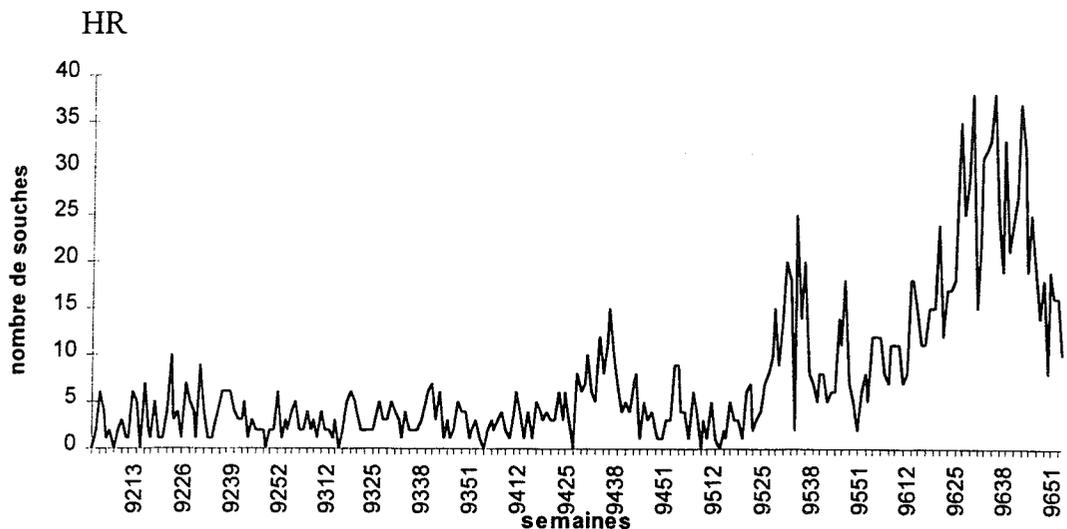
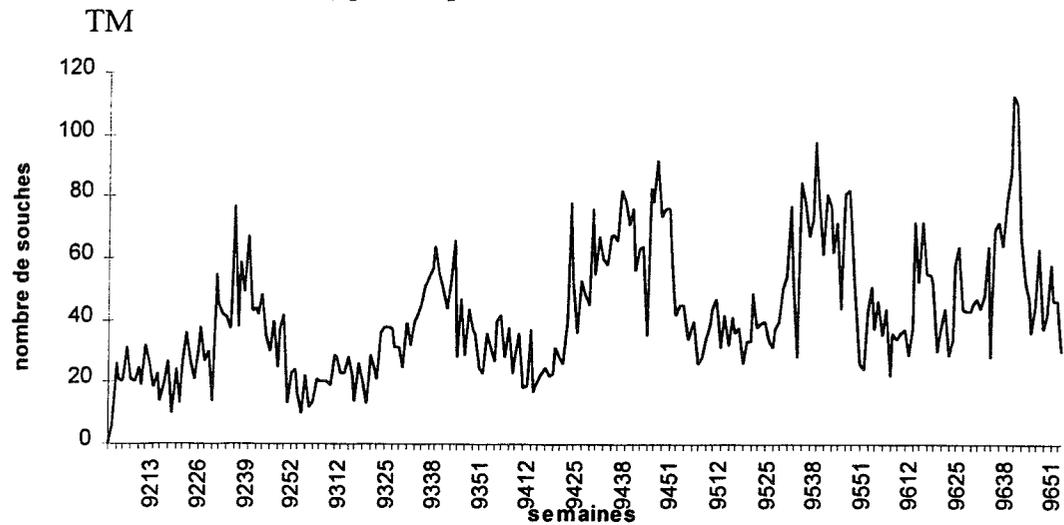
Figure 1 : Séries chronologiques hebdomadaires correspondant au profil 1 des sérotypes Paratyphi B (PB), Brandenburg (BB) et Goldcoast (GC) pour la période 1992-1996.





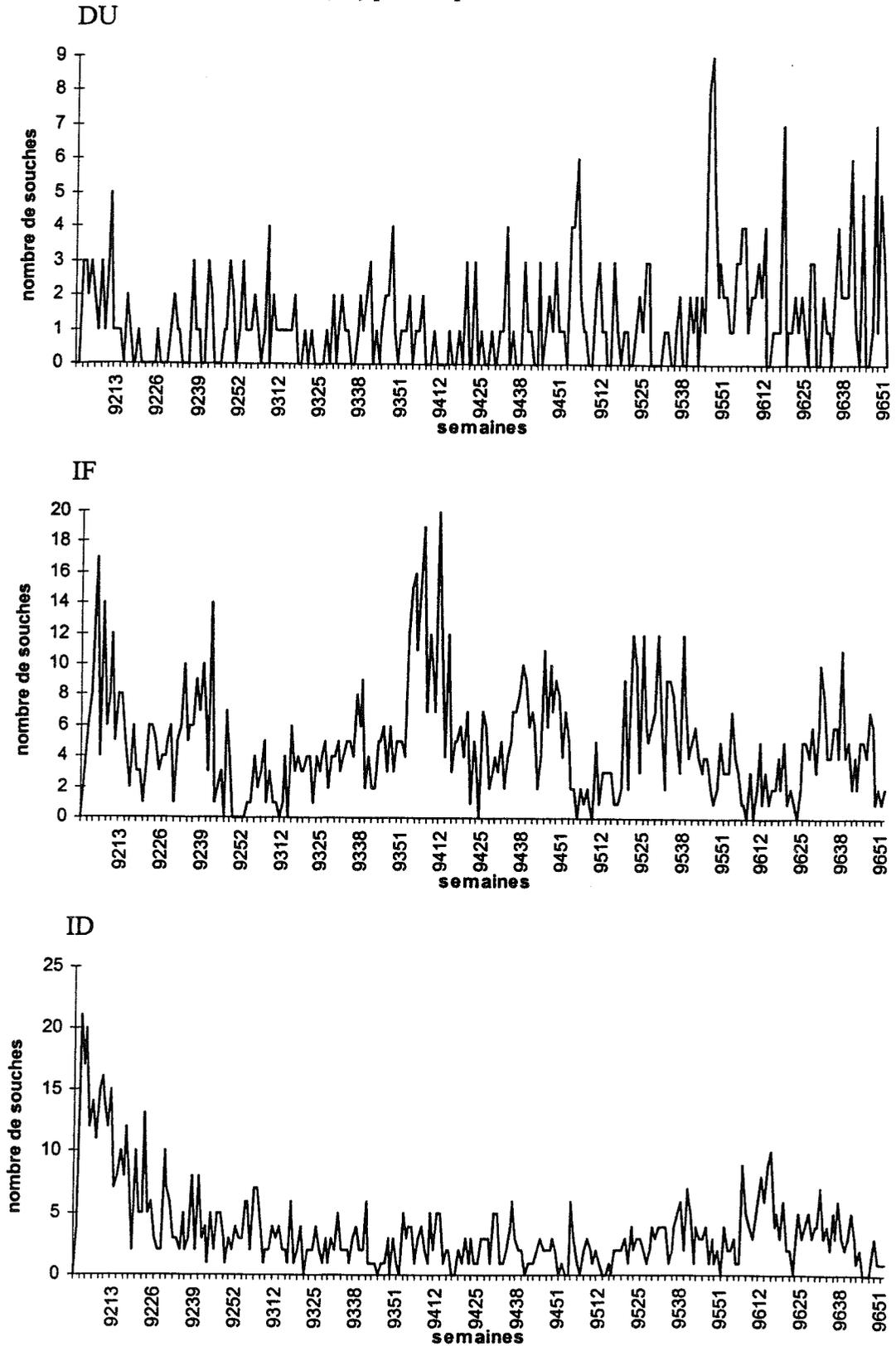
- Profil 2 (figure2) : profil avec saisonnalité et avec tendance, qui regroupe les sérotypes Typhimurium (TM) et Hadar (HR). *S. Typhimurium* présente une tendance linéaire croissante à partir de 1994 et une saisonnalité annuelle. *S. Hadar* présente une tendance exponentielle croissante et une saisonnalité annuelle.

Figure 2 : Séries chronologiques hebdomadaires correspondant au profil 2 des sérotypes Typhimurium (TM) et Hadar (HR) pour la période 1992-1996.



• Profil 3 (figure 3) : profil sans saisonnalité et sans tendance, qui regroupe les sérotypes Dublin (DU), Infantis (IF) et Indiana (ID). La ligne de base accidentée ne présente pas de tendance ou montre un changement de niveau rapide. L'analyse visuelle ne permet pas de repérer de saisonnalité particulière. La série apparaît très aléatoire.

Figure 3 : Séries chronologiques hebdomadaires correspondant au profil 3 des sérotypes Dublin (DU), Infantis (IF) et Indiana (ID) pour la période 1992-1996.



3-4 Elaboration des critères d'alerte (signaux)

La première étape a été la construction et la validation rétrospective d'un seuil d'alerte épidémique. Un seuil d'alerte épidémique est une valeur attendue maximale qui lorsqu'elle est dépassée par la valeur observée déclenche l'alerte. Un tel seuil sert de signal, d'indicateur de surveillance. C'est un critère de premier plan pour détecter l'augmentation anormale d'un sérotype. Dans notre travail, trois approches méthodologiques ont été utilisées pour construire un seuil d'alerte : la méthode des "moyennes historiques" (seuil CNR), la méthode Box et Jenkins (seuil Box & Jenkins) et la méthode Serfling (seuil Serfling).

3-4-1 Méthode des "moyennes historiques" (seuil CNR)

Une augmentation anormale d'un sérotype peut être comparée avec les données historiques correspondant aux données mensuelles des années précédentes. Pour chaque sérotype, un seuil d'alerte mensuel (ou seuil de surveillance) est calculé. C'est une moyenne mensuelle d'une série de 15 mois, qui prend en compte le mois courant, le mois précédent et le mois suivant, sur les 5 années précédentes. Les valeurs mensuelles historiquement élevées (dues aux épisodes épidémiques) qui risquent de rendre le seuil insensible sont minorées (ramenées à la moyenne). Une moyenne corrigée peut ainsi être calculée.

Cette méthode a été développée par le CNR depuis 1992 (6). Elle est utilisée aujourd'hui comme une aide pour repérer les augmentations anormales du nombre de salmonelles. Une base de données informatisée et actualisée, répertoriant plus de 80 sérotypes de *Salmonella* parmi les plus fréquents, permet d'obtenir un relevé de surveillance hebdomadaire qui indique le dépassement éventuel du seuil (annexe 2). Les souches reçues au CNR sont additionnées semaine après semaine pour un même mois, donnant un nombre cumulé mensuel. Chaque semaine du mois en cours, ce nombre cumulé est comparé au nombre moyen attendu pour le mois entier.

Par exemple pour calculer la valeur attendue ou seuil du mois de mai 1998. Les valeurs des mois de mai des 5 années précédentes, de 1993 à 1997 inclus sont relevées. A partir de ces valeurs, une moyenne mensuelle (M5) et un écart type (σ) sont calculés :

$$M5 = \frac{\text{mai 1993} + \dots + \text{mai 1997}}{5}$$

Chacune des 5 valeurs précédemment relevées est ensuite comparée à (M5) la valeur type. Une valeur est détectée comme "point épidémique" si elle dépasse une limite supérieure à $(M5 + 1\sigma)$. Ces valeurs "points épidémiques" sont alors remplacées par la valeur type (M5). Ce procédé appliqué sur chaque mois permet de corriger ou "lisser" la série.

Une moyenne mensuelle corrigée (M5c) est ensuite calculée, déterminant la valeur attendue pour mai 1998. (M5c) porte sur les valeurs (M5) respectives de 15 mois, soit une fenêtre de 3 mois centrée sur le mois de mai et déplacée sur les 5 années successives qui précèdent 1998 :

$$M5c = \frac{(\text{avril 93} + \text{mai 93} + \text{juin 93}) + \dots + (\text{avril 97} + \text{mai 97} + \text{juin 97})}{15}$$

(M5c) est la valeur moyenne attendue du nombre de salmonelles reçues au CNR pour le mois de mai 1998. Une moyenne corrigée (M5c) peut être ainsi calculée pour chaque mois et pour chaque sérotype.

D'autres indicateurs quantitatifs ont été déterminés. Ainsi un relevé d'alerte complet intègre les comparaisons avec les données historiques à 3 et 5 ans (annexe 2). Ce relevé fournit trois valeurs :

- la moyenne mensuelle (M5c) d'une série de 15 mois (le mois courant + les mois précédent et suivant pendant 5 ans). Les valeurs mensuelles minimale, maximale et maximale corrigée sont également disponibles.
- la moyenne des sommes des valeurs du mois courant et du mois précédent pour les 3 années précédentes (M3 = moyenne "2 mois x 3").
- la moyenne des sommes des valeurs du mois courant et des 11 mois précédents pour les 3 années précédentes (M12 = moyenne "12 mois x 3").

Les dépassements éventuels de ces différentes valeurs sont calculés et signalés sur le relevé grâce à des indicateurs visuels qui quantifient en pourcentage le niveau de dépassement. Ces indicateurs permettent de détecter plus facilement une augmentation immédiate (indicateur M5c), une augmentation récente (indicateur M3) ou une augmentation plus discrète mais continue (indicateur M12) pour un sérotype particulier.

La valeur de référence qui sert de seuil d'alerte habituel est la valeur calculée pour chaque mois (M5c + 20%). Pour le CNR, il s'agit surtout d'un "seuil de sensibilisation" (signal) pour détecter une augmentation inhabituelle d'un sérotype particulier.

Une analyse régionale du nombre de souches reçues au CNR est réalisée au cas par cas. Un découpage en régions administratives ou en régions géographiques (départements limitrophes) n'est utilisé que ponctuellement lorsqu'un problème semble émerger. D'autres analyses sont également réalisables à partir des informations épidémiologiques recueillies pour chaque souche reçue sur la fiche standardisée (cf. 3.2.1 et annexe 1).

3-4-2 Introduction aux séries chronologiques

Une série chronologique est un ensemble d'observations enregistrées à intervalles de temps réguliers (semaine, mois...). Une particularité des données de séries chronologiques est que les observations successives ne sont pas indépendantes les unes des autres et que l'ordre dans lequel les observations ont été recueillies joue un rôle important. Par conséquent, l'analyse des séries chronologiques nécessite des techniques statistiques appropriées. De nombreux exemples de séries chronologiques existent dans des domaines variés tels que l'économie, les sciences physiques, la météorologie, la démographie et l'épidémiologie (22).

Les objectifs communs à toute analyse sont la description des variations, la construction d'un modèle et la prédiction des valeurs dans le futur. Les étapes successives qui ont été utilisées dans notre travail sont (23) :

- La représentation graphique des données en fonction du temps pour essayer de retrouver des particularités repérables telles qu'une tendance, une saisonnalité, des pics et des

creux. Cette méthode graphique est indispensable à la fois pour la description et la formulation du modèle.

- Les deux outils principaux complémentaires à l'analyse graphique, la fonction d'autocorrélation et le spectrum. La modélisation des séries chronologiques basée essentiellement sur la fonction d'autocorrélation est souvent décrite comme une approche temporelle type Box et Jenkins, tandis qu'une approche fréquentielle correspond à l'analyse spectrale type Serfling.

L'approche temporelle concerne les relations entre les observations à différents temps, avec en particulier des corrélations entre les intervalles de temps ("lags" ou décalages). Pour l'approche fréquentielle, les observations sont considérées comme étant composées de termes périodiques à différentes fréquences et l'analyse porte sur les amplitudes relatives aux différentes fréquences, comme la décomposition du spectre de la lumière (24). Les deux approches sont en fait mathématiquement équivalentes et peuvent être dérivées l'une de l'autre. Cependant elles font ressortir des caractéristiques différentes des séries chronologiques ce qui justifie de les considérer séparément.

3-4-3 Méthode Box et Jenkins (seuil Box & Jenkins) : analyse des variations temporelles et modèles ARIMA

3-4-3-1 Principe

La modélisation par la méthode Box et Jenkins a été développée dans les années 70 et appliquée initialement aux données économiques (25). Plus tard Armitage l'a appliquée à des données de surveillance épidémiologique de maladies infectieuses (26).

Ces modèles nécessitent des séries stationnaires. La stationnarité est un concept important qui implique que la structure de la série ne se modifie pas avec le temps. Une série ou un processus $\{X_t\}$ est dit stationnaire si son espérance (sa moyenne) et sa variance sont constantes, et que ses autocorrélations dans le temps ont été enlevées par différenciation (dérivation mathématique), soit :

$$E[X_t] = \mu$$

$$\text{Var}[X_t] = \gamma(0)$$

$$\text{Cov}[X_t, X_{t+\tau}] = \gamma(\tau), \text{ pour un décalage temporel } = \tau$$

En remplaçant une observation au temps t par la différence de valeurs observées au temps $(t-1)$ et t (différenciation d'ordre 1), les valeurs individuelles sont indépendantes de la période d'observation.

Ces modèles sont identifiables à l'aide d'outils qui sont la fonction d'autocorrélation et la fonction d'autocorrélation partielle. L'autocorrélation mesure la corrélation entre des observations prises avec un certain décalage dans le temps. Cela mesure la dépendance entre une observation à un temps donné et des observations passées.

La fonction d'autocorrélation (ACF : AutoCorrelation Function) mesure l'autocorrélation entre X_t et $X_{t+\tau}$. La fonction d'autocorrélation partielle (PACF : Partial AutoCorrelation Function) mesure la corrélation entre X_t et X_{t-k} conditionnellement à $\{X_{t-1}, \dots, X_{t-k+1}\}$.

Une fois la série rendue stationnaire, il reste un résidu appelé bruit blanc Gaussien $\{\varepsilon_t\}$ qui correspond à une suite de variables gaussiennes indépendantes, d'espérance nulle et de variance σ^2 . C'est le "bruit aléatoire" résiduel de distribution normale.

Quatre étapes ont été nécessaires pour l'étude des variations temporelles des séries et l'élaboration d'un modèle (27) :

- 1) Transformer la série pour la rendre stationnaire
- 2) Identifier le modèle
- 3) Estimer les paramètres du modèle
- 4) Vérifier l'adéquation et justifier le choix du modèle retenu

Une étape supplémentaire a permis de calculer les prévisions et leur intervalle de confiance dans un but prédictif.

3-4-3-2 Première étape : rendre la série stationnaire

Pour enlever les sources de variation non stationnaires, une série est décomposée en une tendance, une variation saisonnière cyclique et d'autres fluctuations "irrégulières" stationnaires.

La représentation graphique des observations en fonction du temps a permis de repérer visuellement les caractéristiques importantes des variations telles qu'une tendance, un effet périodique, l'existence de valeurs aberrantes ou de discontinuités. Le tracé de la moyenne - étendue ("mean - range plot") a aidé également au repérage des périodes aux valeurs extrêmes qui devaient être corrigées avant de commencer la modélisation. Les séries avec des valeurs corrigées ont été appelées "séries écrêtées" (cf. 3.4.5.2).

La tendance séculaire correspond à "un changement à long terme" du niveau moyen de la série. Si la série présente une tendance, il est recommandé de transformer les données. Par exemple une transformation logarithmique est le plus souvent nécessaire, en particulier si la variance augmente de façon proportionnelle avec l'espérance. Avec des données saisonnières on a pu estimer la tendance sous-jacente par une moyenne mobile de portée au moins égale à l'ordre du décalage saisonnier. Cela correspond au tracé de la moyenne - étendue . Pour rendre la série stationnaire il faut soustraire cette tendance estimée. Un "filtrage" utile pour enlever une tendance est de différencier la série jusqu'à obtention d'une série stationnaire. Par exemple, une tendance linéaire peut être ôtée par une différenciation d'ordre 1 ; dans ce cas on remplace les valeurs au temps $(t+1)$, X_{t+1} par la différence au temps t et $(t+1)$:

$$\nabla_1 X_t = X_{t+1} - X_t .$$

La saisonnalité correspond à un effet cyclique de période T . Pour enlever une composante saisonnière, une différenciation correspondant à l'intervalle entre les saisons a été utilisée. Par exemple, pour des données hebdomadaires présentant une saisonnalité annuelle, une différenciation d'ordre 52 a été utilisée :

$$\nabla_{52} X_t = X_t - X_{t-52}$$

3-4-3-3 Deuxième étape : identifier le modèle

Box et Jenkins ont proposé de nombreux modèles théoriques (25) qui sont fondés sur deux processus différents : le processus autorégressif AR (p) et le processus moyenne mobile MA (q).

Le processus autorégressif AR (p) (AutoRegressive process)

Un processus $\{X_t\}$ est dit autorégressif d'ordre (p) si

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t$$

où $\{\phi_i\}$ sont des constantes, paramètres du modèle et $\{\varepsilon_t\}$ le "bruit aléatoire" résiduel.

Cette écriture ressemble à celle d'une régression multiple mais les paramètres $\{X_t\}$ sont régressés sur des valeurs de son passé, d'où le terme autorégressif. Ce processus montre des dépendances temporelles à certains intervalles entre les observations.

Un AR (p) peut s'écrire sous une autre forme équivalente :

$$\Phi(B) X_t = \varepsilon_t \quad \text{où } B^n X_t = X_{t-n} \text{ définit l'opérateur retard et}$$
$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p, \text{ est un polynôme en } B.$$

Dans un modèle autorégressif pur d'ordre p, l'ACF décroît exponentiellement vers 0, et la PACF est nulle à partir du décalage (p+1).

Le processus moyenne mobile MA (q) (Moving Average process)

Un processus $\{X_t\}$ est dit moyenne mobile d'ordre (q) si

$$X_t = \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

où $\{\theta_i\}$ sont des constantes, paramètres du modèle et $\{\varepsilon_t\}$ le "bruit aléatoire" résiduel.

En règle générale θ_0 est égal à 1. X_t est une combinaison linéaire des variations aléatoires présentes et passées dans les données, mais non des valeurs $\{X_{t-i}\}$ comme dans un processus autorégressif.

Un MA (q) peut s'écrire sous une autre forme équivalente :

$$X_t = \Theta(B) \varepsilon_t \quad \text{où } B^n \varepsilon_t = \varepsilon_{t-n} \text{ définit l'opérateur retard et}$$
$$\Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q, \text{ est un polynôme en } B.$$

Dans un modèle moyenne mobile MA pur d'ordre q, la PACF décroît exponentiellement vers 0, et l'ACF est nulle à partir du décalage q+1.

Modèles mixtes ou autorégressif - moyenne mobile ARMA (p,q) (AutoRegressive Moving Average process)

Les modèles mixtes combinent les processus autorégressif et moyenne mobile. La dualité des comportements de l'ACF et de la PACF pour chaque processus AR (p) et MA (q), guide l'identification du processus sous-jacent à la série.

Un processus $\{X_t\}$ est dit ARMA d'ordre (p,q) si

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

où $\{\phi_i, \theta_i\}$ sont des constantes, et $\{\varepsilon_t\}$ le "bruit aléatoire" résiduel.

Un ARMA (p,q) peut s'écrire sous une autre forme équivalente :

$$\Phi(B) X_t = \Theta(B) \varepsilon_t \quad \text{où } \Phi(B) \text{ et } \Theta(B) \text{ sont définis par :}$$

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

$$\Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$$

Dans ce cas les ACF et PACF décroissent exponentiellement vers 0. Un outil complémentaire à l'identification de l'ordre d'un modèle de type ARMA a été proposé et développé pour les séries chronologiques : l'autocorrélation inverse (IACF : Inverse AutoCorrelation Function).

Processus ARIMA (p,d,q) (AutoRegressive Integrated Moving Average)

Le processus ARIMA intègre un effet de tendance. C'est un processus ARMA dans lequel une différenciation (d) a été effectuée pour ôter une tendance et stabiliser la moyenne. Ainsi la série observée $\{X_t\}$ présentant une tendance est remplacée par la série Y_t :

$$Y_t = \nabla^d X_t = (1 - B)^d X_t$$

En règle générale, d vaut 1 ou 2. La série Y_t est ensuite modélisée selon un processus ARMA (p,q) . La série originale X_t s'écrit alors :

$$\Phi(B) (1 - B)^d X_t = \Theta(B) \varepsilon_t$$

Processus SARIMA (p,d,q) (P,D,Q)s (Seasonal AutoRegressive Integrated Moving Average)

Le processus SARIMA intègre un effet saisonnier et / ou un effet de tendance. Il s'agit de faire intervenir dans un modèle ARIMA, des décalages adaptés au profil saisonnier de la série (exemple : des décalages multiples de 52) en évitant une augmentation excessive du nombre de paramètres. Le principe de parcimonie fait opter pour le modèle comportant le moins de termes.

La première étape consiste à modéliser par un modèle ARIMA (P,D,Q) les observations présentant une saisonnalité de période S (distance entre les observations). Le modèle s'écrit :

$$\Phi_p(B^S) \nabla_S^D X_t = \Theta_Q(B^S) u_t \quad \text{où} \quad \nabla_S^D = (1 - B^S)^D$$

La seconde étape consiste à modéliser par un modèle ARIMA (p,d,q) les liaisons à court terme entre les erreurs u_t :

$$\Phi_p(B) \nabla^d u_t = \Theta_q(B) \varepsilon_t \quad \text{où} \quad \nabla^d = (1 - B)^d$$

La résultante de ces deux modèles conduit à :

$$\nabla^d \Phi_p(B) \Phi_p(B^S) \nabla_S^D X_t = \Theta_q(B) \Theta_Q(B^S) \varepsilon_t$$

La sélection du modèle approprié se fait par une exploration visuelle des corrélogrammes d'ACF et de PACF. Les corrélogrammes indiquent les valeurs de la corrélation pour différents décalages temporels. Leur forme donne des indications sur le processus sous-jacent à l'origine des données et sur les paramètres à inclure dans le modèle.

3-4-3-4 Troisième étape : estimer les paramètres du modèle

Plusieurs méthodes existent mais nous avons utilisé la méthode des moindres carrés, selon un critère calculé conditionnellement à p valeurs initiales et q valeurs du bruit blanc. Cette étape a été effectuée par le logiciel SSS1 (cf. 3.6).

3-4-3-5 Quatrième étape : vérifier l'adéquation du modèle

L'évaluation de l'adéquation du modèle nécessite de tester les paramètres du modèle $\{\phi_i, \theta_i\}$ et l'hypothèse selon laquelle les ε_t forment un bruit blanc résiduel de distribution normale. Les paramètres ont été testés par un test de Student. La vérification du caractère aléatoire des valeurs résiduelles ε_t a été réalisé par exploration de l'ACF et de la PACF. L'absence d'autocorrélation des valeurs résiduelles ε_t a été vérifiée par les tests de Ljung-Box et Box-Pierce (28, 29).

Si le modèle n'est pas adéquat, il est possible de le modifier au vu des ACF et PACF des valeurs résiduelles (22). Tout paramètre non significatif est retiré et de nouveaux paramètres peuvent être ajoutés. Ainsi le choix des paramètres (p, d, q, P, D, Q) est fait par des essais successifs. Une procédure récursive a été utilisée jusqu'à identification d'un modèle adéquat. Le critère informatif d'Akaike (AIC), fondé sur le principe de parcimonie, est utilisé en complément si nécessaire pour comparer plusieurs modèles équivalents (30). Le logiciel SSS1 ne donne pas la valeur de l'AIC mais un critère approché (cf. 3.6) : MAPE (Mean Absolute Percent Error). Pour chaque intervalle de temps, le logiciel détermine l'importance des résidus par rapport à la série modélisée et calcule un ratio. La somme des ratio sur l'ensemble des périodes de la série est moyennée et transformée en pourcentage pour donner le MAPE. Il donne une estimation approximative de l'adéquation du modèle mais oriente surtout le choix quand plusieurs modèles sont possibles. Plus le MAPE est élevé, meilleur est le modèle.

3-4-3-6 Cinquième étape : prévoir les valeurs futures et calculer leur intervalle de confiance

A partir du modèle choisi, des prévisions des valeurs futures de la série ont été calculées. Les données prédites sont fonctions des observations immédiatement précédentes. Pour un risque α donné, un intervalle de confiance autour de ces prédictions a été estimé : intervalle de prévision qui a la particularité d'augmenter avec le temps.

Une règle de la modélisation de Box et Jenkins est qu'il ne faut pas chercher à prévoir à trop long terme. Un cycle saisonnier est le maximum. En théorie, seul le point suivant de la série peut être prédit avec confiance (31).

3.4.4 Méthode Serfling (seuil Serfling) : Analyse spectrale

3-4-4-1 Principe

L'analyse spectrale consiste en une exploration des composantes cycliques des données. Le but est de décomposer une série chronologique (signal), qui présente des oscillations cycliques, en une série de courbes sinusoïdales de certaines amplitudes et périodes (32, 33). Telle la décomposition du "spectre" d'un rayon de lumière blanche à travers un prisme, les données de surveillance sont "filtrées" pour identifier les différentes longueurs d'ondes. Une longueur d'onde ou période correspond au nombre d'unités temporelles séparant deux pics dans une sinusoïde. La fonction pour exprimer une composante périodique dans un signal s'écrit :

$$Y_t = R \cos(\omega t + \theta)$$

où ω est la fréquence de la variation périodique ($= 2\pi / \text{période}$) ou nombre de fois qu'un événement se répète dans un intervalle de temps donné ;

R est l'amplitude de la variation cyclique ;

θ est la phase, correspondant au point de départ de la vague.

En 1807 le mathématicien Jean Baptiste Fourier montra que tout signal périodique peut être décomposé en une série de courbes sinusoïdales. Il emploie une série trigonométrique ("série de Fourier") pour exprimer une fonction discontinue en une somme de sinus et cosinus (32). De manière générale, dans un signal de n données on peut faire passer $n/2$ sinusoïdes (exemple : dans 256 points de données on peut faire passer 128 courbes sinusoïdales).

La transformée de Fourier convertit un signal exprimé comme une fonction du temps en une fonction de fréquences. Aucune information du signal n'est perdue lors de la transformation. La transformée de Fourier calcule deux coefficients (coefficients de Fourier) α et β qui permettent de calculer l'amplitude et la phase de la courbe pour chaque fréquence testée :

$$Y_t = R \cos(\omega t + \theta)$$

$$\text{or } \cos(\alpha + \beta) = \cos(\alpha) \cos(\beta) - \sin(\alpha) \sin(\beta)$$

donc Y_t peut être développé et s'écrire :

$$Y_t = R (\cos \omega t \cos \theta - \sin \omega t \sin \theta) \quad \text{où} \quad \alpha = R \cos \theta \quad \text{et} \quad \beta = -R \sin \theta$$

L'expression devient alors :

$$Y_t = \alpha \cos \omega t + \beta \sin \omega t$$

α et β contiennent l'information concernant la phase et l'amplitude. C'est le format d'expression de la transformée de Fourier pour chaque fréquence. Ce format s'écrit sous forme de nombres complexes, permettant l'expression de α et β sous une forme unique :

$(\alpha + \beta i)$ où α est appelé la composante réelle du nombre complexe,
 β est appelé la composante imaginaire et
 i représente le nombre imaginaire "racine de (-1)".
 α et β sont des nombres réels, et $i = \sqrt{-1}$, racine de l'équation $(x^2 + 1 = 0)$.

Pour chaque coefficient de Fourier, une courbe sinusoïdale peut être construite sous la forme : $Y_t = \alpha \cos \omega t + \beta \sin \omega t$

Par exemple, pour 256 points de données, la transformée de Fourier renvoie 128 nombres complexes de la forme $(\alpha + \beta i)$. La somme de toutes les courbes sinusoïdales reproduit le signal original, sans perte d'information.

Serfling a utilisé la méthode des séries de Fourier pour analyser des données de mortalité et a introduit le concept de "seuil épidémique" (33, 34).

3-4-4-2 Identification d'un modèle par décomposition du signal original : étapes nécessaires au calcul d'un coefficient de Fourier sur une série chronologique

Un coefficient de Fourier montre l'amplitude des contributions d'une fréquence spécifiée dans les données. Quatre étapes ont été nécessaires pour calculer un coefficient de Fourier pour une fréquence donnée (33). Il a fallu mettre en "résonance" le signal original et la courbe sinusoïdale à tester, puis évaluer l'importance de cette résonance.

La première étape est de dessiner une courbe sinusoïdale d'une fréquence donnée. L'équation d'une courbe sinusoïdale simple est de la forme : $Y_t = \cos \omega t$.

or la fréquence ω s'exprime sous la forme : $\omega = 2 \pi / \text{période}$

d'où : $Y_t = \cos (2 \pi t / \text{période})$

Si l'exploration visuelle de la série chronologique montre une saisonnalité {S} importante, cette valeur est testée en priorité comme période d'oscillation.

La deuxième étape consiste à multiplier le signal que constitue la série originale, par la courbe cosinus obtenue précédemment.

La troisième étape estime la surface sous la courbe réalisée par les produits (signal x courbe cosinus).

La quatrième étape ajuste la phase pour maximiser la surface sous la courbe des produits. Cette étape requiert de déplacer la courbe cosinus jusqu'à obtenir la valeur maximale possible pour la somme des produits. Lorsque la somme est maximale, les deux courbes sont en "phase".

Afin d'éviter la répétition des étapes, nous avons utilisé la transformée rapide de Fourier du logiciel Excel (FFT = Fast Fourier Transform) (cf. 3.6) pour calculer en une fois les contributions de toutes les fréquences possibles dans les données. Les coefficients de Fourier sont exprimés sous forme de nombres complexes. L'algorithme de calcul de la transformée de Fourier est très performant si le nombre de points est une puissance de 2. Pour n données, il y a n coefficients. Les n/2 derniers coefficients correspondent à une image en miroir des n/2 premiers.

L'information recherchée est la période {P} (semaines ...) qui a la plus forte contribution dans les données. Elle a été recherchée par l'analyse du périodogramme. Le périodogramme est un graphique de distribution des contributions des périodes. Pour le représenter il faut calculer "l'énergie" {E} qui est la racine carré de la somme des carrés de la composante réelle $\{\alpha\}$ et imaginaire $\{\beta\}$ du nombre complexe. Soit : $E = \sqrt{(\alpha^2 + \beta^2)}$

Les résultats des calculs donnent la période pour laquelle {E} a une valeur maximale et permettent de tracer le périodogramme. L'exploration visuelle du périodogramme peut montrer la période {P} pour laquelle la contribution des oscillations est la plus forte. Il peut aussi indiquer des oscillations appelées "harmoniques", correspondant à des sous multiples de {P} (1/2, 1/3, 1/4...). Lorsque la période décroît vers 0, correspondant à des fréquences de plus en plus élevées, l'énergie décroît vers 0.

3-4-4-3 Identification d'un modèle spectral et élaboration d'un seuil d'alerte Serfling

A partir de l'information extraite du périodogramme il est possible de construire un seuil épidémique pour la détection de valeurs excessives dans les données de surveillance (27, 33). Trois étapes successives sont nécessaires :

- Prendre en compte la tendance sur la période ;
- Introduire les termes correspondant aux fréquences à inclure dans le modèle ;
- Calculer l'intervalle de confiance de ce modèle.

a) Première étape : définir la tendance

Une tendance linéaire se définit par une pente et par sa valeur au croisement de l'axe des ordonnées Y (l'intercept). Elle s'exprime sous la forme d'une équation telle que :

$$Y_t = a + b.t, \text{ où } b \text{ est la pente et } a \text{ est la valeur au croisement à l'axe } Y (t = 0).$$

De façon similaire, une tendance exponentielle se définit par une équation de la forme :

$$Y_t = a + c.exp(b.t).$$

L'adéquation de l'estimation de la tendance aux données peut être réalisée en minimisant la somme des carrés des écarts entre le signal original et la tendance. Ceci revient à minimiser la variance résiduelle. Le coefficient de régression est le rapport entre la covariance des deux séries (signal et tendance) divisé par la racine carrée du produit des deux variances de séries :

$$R^2 = \frac{\text{covariance (signal, tendance)}}{\sqrt{(\text{Variance du signal} \times \text{Variance de la tendance})}}$$

La valeur de R^2 donne la proportion de la variance totale de la série originale prise en compte ou "expliquée" par le modèle, en l'occurrence ici par la tendance.

La fonction "Solveur" du logiciel Excel a permis de calculer les valeurs des coefficients a, b et c (cf. 3.6).

b) Deuxième étape : prise en compte des contributions cycliques et prédictions du modèle

A ce stade, il faut introduire dans le modèle les composantes cycliques identifiées sur le périodogramme au cours de la première phase d'analyse de la série (cf. 3.4.4.2). La période et les harmoniques, chacune définie par son amplitude {A} et sa phase {φ}, sont retenues comme paramètres du modèle et associées à la tendance déjà estimée.

La fonction "Solveur" du logiciel Excel a permis à nouveau de calculer les valeurs optimales, donnant une valeur du coefficient R^2 la plus proche possible de 1. On a obtenu ainsi le modèle qui "explique" le mieux la variance originale. La part de la variance restante, soit $(1 - R^2)$, qui n'est pas expliquée par des termes saisonniers, correspond à des facteurs aléatoires.

Le modèle final identifié, qui est la combinaison des estimations de la tendance et des composantes cycliques, peut être répété à l'infini pour obtenir les prédictions des saisons suivantes. Il est possible également de déterminer la limite supérieure du modèle qui servira de seuil pour détecter les phénomènes épidémiques.

c) Troisième étape : optimisation des paramètres

L'analyse spectrale donne la même importance aux données sur l'ensemble de la période d'étude. Elle utilise l'ensemble des données comme un tout. Dans le but de réaliser un seuil épidémique pour une détection précoce de passage en phase épidémique, il faut construire, à l'origine, un modèle qui ne prend pas en compte les données des périodes épidémiques observées dans le passé. Cette sélection repose sur l'exclusion des valeurs excédant la moyenne + 1,64 écart type. Le choix de 1,64 écart type correspond à une marge d'exclusion unilatérale dans laquelle la probabilité d'observer des valeurs sur l'ensemble de la série est inférieure à 5%. A ce stade de la modélisation, toutes les valeurs excédant ce seuil sont donc éliminées pour ne pas les prendre en compte dans le calcul des paramètres du modèle.

Ce niveau de seuil peut être modifié en fonction de la sensibilité et de la spécificité recherchées. Plus ce seuil est bas et plus on a de "chance" de détecter des épidémies, mais moins ces épidémies correspondent à de véritables excès de cas.

3-4-5 Les choix méthodologiques

3-4-5-1 *Le temps*

Compte tenu de l'objectif de réactivité précoce du système d'alerte et de la périodicité hebdomadaire du relevé de surveillance établi par le CNR, il a été choisi de construire et d'analyser des seuils d'alerte sur des séries hebdomadaires. Les séries ont donc été établies en semaine de réception des souches de salmonelles au CNR.

L'origine du temps pour chaque série a été fixée au 1er janvier 1992. La semaine du 1er au 7 janvier 1992 est la semaine 9201 ou décalage 1. L'année 1992 comportait 53 semaines, les années 1993 à 1997 comportaient 52 semaines. L'étendue de la période de 1992 jusqu'en 1996 ou 1997 représente donc des séries sur 261 ou 313 semaines respectivement.

3-4-5-2 *Les options retenues pour chaque type de modélisation*

Pour pouvoir modéliser une série chronologique, il est nécessaire d'étudier une longue série de valeurs en régime stable. Or, dans le cas des salmonelles comme d'autres maladies infectieuses, on observe la superposition d'un phénomène stable et la présence de pics épidémiques. Il faut donc, dans un premier temps, détecter les points dits "épidémiques" qui correspondent à des valeurs exceptionnellement fortes, pour pouvoir les remplacer dans un deuxième temps. A cette étape, le but est d'appréhender la structure stable sous-jacente pour la modéliser tout en restant le plus proche de la série originale.

Pour l'approche temporelle Box et Jenkins, une définition quantitative opérationnelle a été utilisée pour identifier les points dits "épidémiques". Cette définition a été fondée d'une part sur la mise en évidence d'année(s) exceptionnelle(s) sur le graphe de la moyenne étendue et d'autre part sur l'observation séquentielle de la série elle-même.

Le graphe de la moyenne étendue représente, pour chaque année (soit une saison), l'étendue des valeurs observées (différence entre la valeur maximale et la valeur minimale sur une année donnée) en fonction de la moyenne annuelle. Ce graphique permet de repérer les années qui se détachent de l'ensemble, suggérant l'existence de points "épidémiques".

Dans le cadre d'une série observée à intervalle régulier d'une semaine pendant plusieurs années et qui ne présentait pas de tendance, un point a été dit "épidémique" si le nombre observé était supérieur à la moyenne des maxima des années antérieures considérées comme normales, additionnée de 1.64 fois l'écart type de ces maxima. Après examen de plusieurs possibilités, il a été décidé de prendre les quatre valeurs maximales par année pour le calcul de la moyenne. Le 1.64 correspond au quantile unilatéral d'ordre 95% de la loi normale $N(0,1)$. Ces deux choix ont permis le juste équilibre entre ne pas enlever trop de points observés et rendre la série stable. La définition opérationnelle a identifié systématiquement les épisodes "épidémiques" passés.

Dans un deuxième temps, il a été choisi de remplacer ces valeurs "épidémiques" par des valeurs "normales" attendues. La valeur "normale" attendue a été pour chaque intervalle de temps la moyenne des valeurs correspondantes aux mêmes intervalles pour les années précédentes considérées comme normales.

Ces différentes étapes (détection et remplacement des points épidémiques) ont permis de disposer d'une série écrêtée stable suffisamment longue et par-delà la possibilité d'une modélisation globale de la série par un modèle de type ARIMA.

Pour l'approche fréquentielle de type Serfling, la modélisation n'a pas nécessité d'écrêter préalablement la série. Le processus d'ajustement par analyse spectrale effectue un lissage systématique de la série qui substitue aux valeurs observées des valeurs représentables par une courbe continue et sans point anguleux. Les points "épidémiques", détectés et remplacés en cours de modélisation, ont été les valeurs qui excédaient la valeur seuil correspondant à la moyenne de la série additionnée de 1.64 fois l'écart type. Ici, il a été choisi également d'exclure des valeurs ayant une probabilité de moins de 5% d'être observées sur l'ensemble de la série, dans une optique unilatérale. Lors de l'étape d'optimisation des paramètres (cf supra), les valeurs excédant la valeur seuil n'ont pas été prises en compte dans l'estimation du modèle final.

3-4-5-3 Détermination d'un seuil d'alerte

A partir de la modélisation adéquate de la structure de la série chronologique, un seuil d'alerte a été élaboré.

Le modèle ARIMA a été identifié sur la série complète de 5 années, de 1992 à 1996. Il s'agit soit de la série écrêtée après détection et remplacement des "points épidémiques"(cf 3.4.5.2), soit de la série originale si aucune année épidémique n'a été détectée. Le modèle a été ensuite appliqué rétrospectivement, pas à pas sur les données disponibles, pour obtenir les

prédictions moyennes année par année sur la période 1993-1997. La limite unilatérale à 95% des prédictions moyennes a déterminé le seuil d'alerte Box & Jenkins.

Le modèle spectral a été identifié sur la série originale complète de 5 années, de 1992 à 1996. Les estimations des paramètres ont été optimisées par une étape finale d'ajustement du modèle, après retrait des points épidémiques au seuil $\alpha = 5\%$ unilatéral du modèle (cf. 3.4.5.2). Le seuil d'alerte Serfling a été défini comme la limite unilatérale à 95% du modèle final.

Chaque seuil superposé à la série initiale, nous a amené à observer à quel moment les épidémies existantes durant la période 1993-1997 auraient été détectées.

3-4-5-4 *Analyses complémentaires*

Une analyse détaillée des dépassements de seuil identifiés a été effectuée à partir des renseignements épidémiologiques disponibles dans la base de données du CNR. Son but a été d'explorer des arguments en faveur d'une épidémie, tels que des regroupements géographiques, départementaux ou régionaux, ou touchant une classe d'âge particulière.

3-5 **Choix d'une référence pour la comparaison des seuils**

Les critères de jugement pour analyser la validité des critères d'alerte ont porté sur la sensibilité, la valeur prédictive positive, la réactivité (délai entre la détection et le début réel de l'épidémie) et la fin de l'épidémie (35).

3-5-1 Référence "brute"

La référence (référence "brute") a été constituée par les épidémies validées, survenues entre 1992 et 1997, et ayant donné lieu à une investigation. Cependant une épidémie est définie parce qu'elle a été détectée et investiguée. De fait, les périodes épidémiques passées inaperçues n'ont pas été considérées comme épidémiques.

La difficulté à définir un gold standard absolu pour une épidémie nous a obligé à prendre une option méthodologique opérationnelle qui a permis de classer les semaines en épidémiques (au-dessus du seuil) et non épidémiques (en dessous du seuil). Dans un but de détection précoce des épidémies, le début d'une épidémie a été défini comme la première semaine au cours de laquelle le seuil a été dépassé.

La comparaison a été réalisée avec les semaines effectivement investiguées ou non investiguées par le RNSP, correspondant respectivement aux périodes "épidémiques" ou "non épidémiques" définies a priori. Ces critères ont permis de catégoriser chaque semaine en vrai positif (VP), faux positif (FP), vrai négatif (VN) ou faux négatif (FN).

Le tableau d'évaluation pour chaque seuil a été établi selon la présentation suivante :

Seuil	Semaine "épidémique" investiguée	Semaine "non épidémique" non investiguée	
Signal +	VP	FP	vpp*
Signal -	FN	VN	vpn*

* Se = sensibilité, Sp = spécificité, vpp = valeur prédictive positive, vpn = valeur prédictive négative.

Pour chaque sérotype, la comparaison des seuils a porté sur deux années consécutives dont l'une comportait une année avec épidémie reconnue. La comparaison a été réalisée à partir du second semestre 1993, date de collaboration opérationnelle entre le CNR et le RNSP.

3-5-2 Référence "corrigée"

Sur la base des analyses complémentaires temps, lieu, personnes (cf. 3.4.5.4), d'un jugement pragmatique et indépendamment des seuils construits par les trois méthodes, des semaines ont été reclassées en épidémiques vraies. Une nouvelle référence (référence "corrigée") a été définie. Elle intègre ces semaines reclassées en épidémies vraies qui correspondent à des augmentations importantes prolongées, présentant une concentration géographique ou par classe d'âge. A partir de cette référence "corrigée", une nouvelle estimation de la sensibilité et de la valeur prédictive positive a été calculée.

Pour les deux références "brute" et "corrigée", les estimations de sensibilité et valeur prédictive positive ont été déterminées en faisant varier le délai d'alerte de 1, 2 ou 3 semaines.

3.6 Outils d'analyse statistique

La modélisation Box et Jenkins a été réalisée à l'aide du logiciel SSS1 (Statistical Software for Public Health Surveillance, CDC Atlanta, Georgia 30333, USA, 1994).

La modélisation par analyse spectrale a été réalisée sur le logiciel Excel (Version 5.0, Microsoft Corporation, 1993), en utilisant la fonction Transformée Rapide de Fourier (FFT) et la fonction Solveur d'optimisation (27).

La gestion des fichiers, les analyses de répartition géographique et par âge ont été réalisées sur le logiciel Epi Info (Version 6.03, CDC Atlanta, Georgia 30333, USA, 1995).

4 - RESULTATS

4.1 Résultats pour *S. Paratyphi B*

4-1-1 Descriptif de la série originale et de l'épidémie reconnue et investiguée

La série originale hebdomadaire de *S. Paratyphi B* (SPB) est une série sans tendance apparente sur le long terme et avec une saisonnalité d'environ 52 semaines (figure 5, p 36). Son profil est comparable à deux autres sérotypes étudiés : *S. Brandenburg* et *S. Goldcoast* (cf. 3.3, figure 1, p 6).

Sur une ligne de base qui oscille entre 0 et 10 souches de salmonelles par semaine, se détache nettement un pic correspondant à l'épidémie nationale survenue en été automne 1993. Le maximum du pic compte 31 souches de salmonelles à la semaine 93-41 [décalage 94 = 13/10/93 - 19/10/93].

L'épidémie a débuté au cours de la troisième semaine du mois d'août 1993 (semaine 93-33 = décalage 86), et s'est terminée mi-novembre 1993 (semaine 93-47 = décalage 100). La lysotypie des souches indiquaient que la plupart (80%) appartenaient au lysotype 1 var 3, le plus fréquent des lysotypes de *S. Paratyphi B* en France. A la demande de la DGS le 02/11/93 (semaine 93-44 = décalage 97), le RNSP a initié une enquête épidémiologique en collaboration avec les MISP des DDASS (36).

Au total, le CNR qui a détecté l'épidémie, a comptabilisé 273 souches de ce sérotype pendant les trois mois. La région Poitou-Charentes était la plus touchée (TA = 32/million), suivie par les régions limitrophes puis le reste du territoire. Tous les groupes d'âge étaient touchés mais le taux d'attaque était plus élevé chez les enfants de moins de 5 ans (20/million) et le plus bas chez les 6-14 ans (7/million). Trente sept pour-cent des personnes ont été hospitalisés et un décès est survenu.

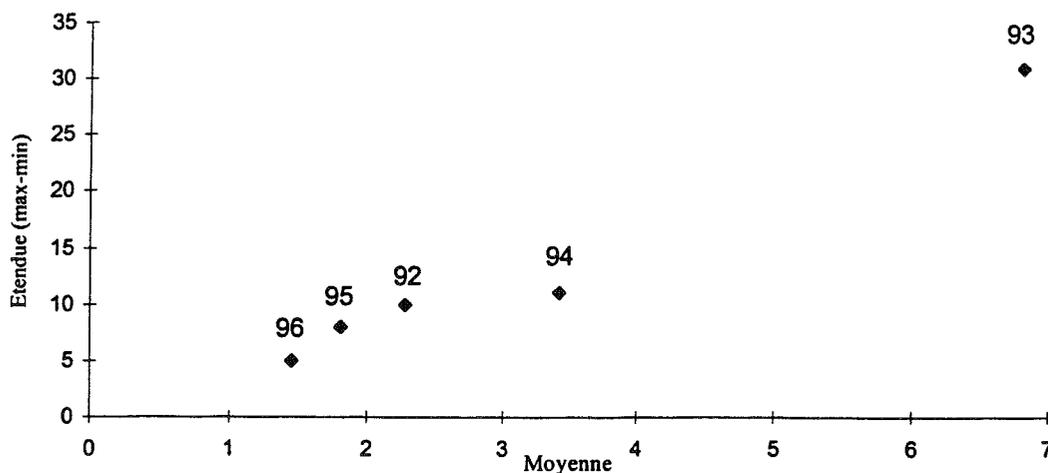
Sur la base des informations obtenues lors de l'investigation d'une épidémie à *S. Paratyphi B* survenue en 1990 dans la région Poitou-Charentes, l'enquête s'est orientée vers une origine laitière caprine. L'enquête cas-témoins a débuté le 08/11/93 (semaine 93-45 = décalage 98) et a porté principalement sur la consommation de produits laitiers et de fromage. Le risque de salmonellose à *S. Paratyphi B* était significativement plus élevé chez les consommateurs d'un fromage de chèvre de forme ronde et de taille moyenne de marque X (OR = 12 ; IC à 95% : 1,6 - 92,3) (15).

4-1-2 Modélisation ARIMA et seuil Box & Jenkins pour *S. Paratyphi B*

4-1-2-1 *Détection et remplacement des "points épidémiques"*

Le graphique de la moyenne-étendue montre que l'année 1993 se détache des autres années.

Figure 4 : Graphique de la moyenne - étendue de la série originale *S. Paratyphi B* (SPB) pour la période 1992-1996.



L'application de la méthode de détection des "points épidémiques" sur l'année 1993 permet de repérer quels sont les semaines ayant un nombre de salmonelles inhabituellement excessifs par rapport aux années considérées comme "normales" (cf. 3.4.5.2). La valeur limite au-delà de laquelle un point a été déclaré épidémique est égal à 11 pour l'année 1993. Douze points épidémiques ont été identifiés et ont été remplacés, comme indiqués dans le tableau I.

Tableau I : Points épidémiques détectés et remplacés sur la série originale SPB.

Semaines* (décalages)	Valeurs détectées	Valeurs de remplacement
93-36 (89)	30	10
93-37 (90)	25	3
93-38 (91)	20	6
93-39 (92)	17	8
93-40 (93)	21	4
93-41 (94)	31	3
93-42 (95)	16	2
93-43 (96)	23	4
93-44 (97)	15	2
93-45 (98)	23	1
93-46 (99)	16	1
93-47(100)	13	1

* Semaine d'origine 01/01/1992 - 07/01/1992 = semaine 92-01 = décalage 1

La série écrêtée SPBe est plus régulière (figure 5) et le graphique de la moyenne - étendue correspondant ne montre plus d'année exceptionnelle (figure 6). La série SPBe plus stable peut à présent être modélisée selon les étapes d'une approche Box et Jenkins.

Figure 5 : Superposition de la série originale *S. Paratyphi B* (SPB) et de la série écrêtée (SPBe) pour la période 1992-1996.

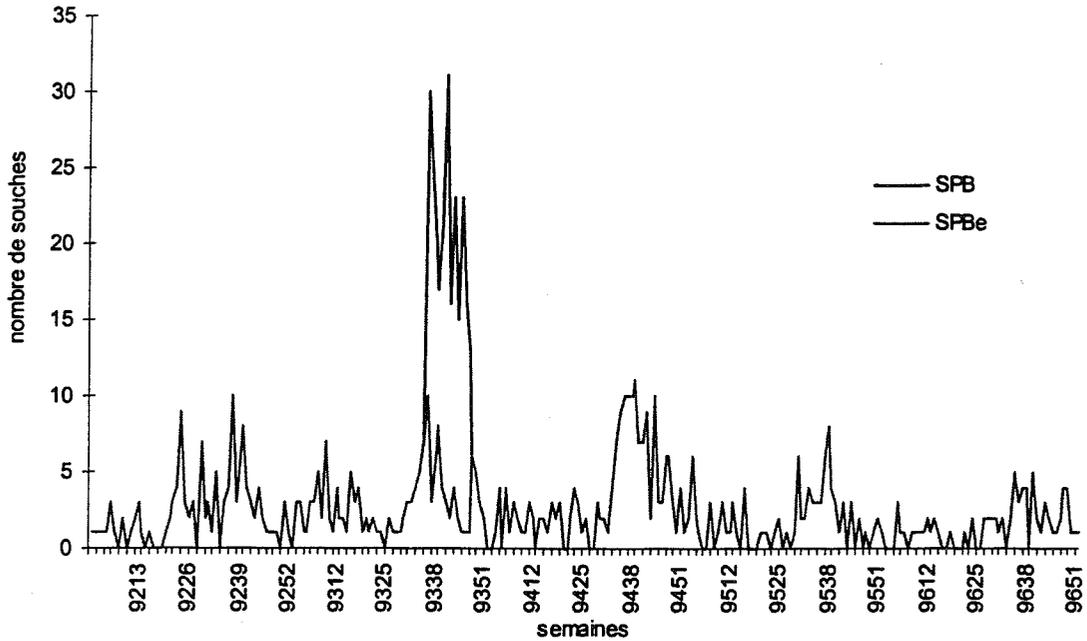
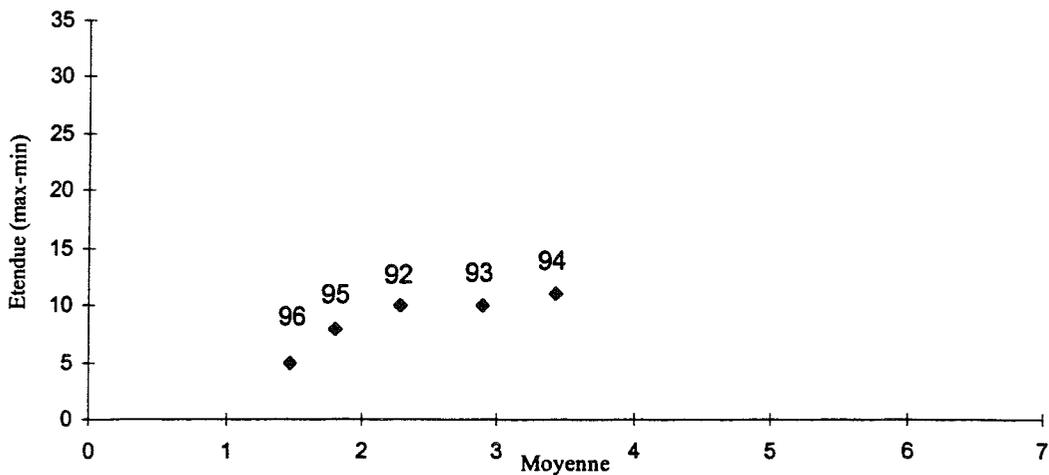


Figure 6 : Graphique de la moyenne - étendue de la série écrêtée *S. Paratyphi B* (SPBe) pour la période 1992-1996.



4-1-2-2 Identification du modèle ARIMA

A ce stade, il faut stabiliser la moyenne de la série SPBe par un processus de différenciation. La série SPBe (figure 7) et sa fonction d'autocorrélation (figure 8) montrent une tendance linéaire et une saisonnalité de 52 semaines. En effet, l'ACF décroît vers 0 dans les dix premiers intervalles de temps (décalages) et remonte de façon significative, telle une courbe sinusoïde, autour de 52 semaines.

Figure 7 : Série échantillonnée (SPBe) sur les 261 semaines de la période 1992-1996 (261 décalages).

Original
Series

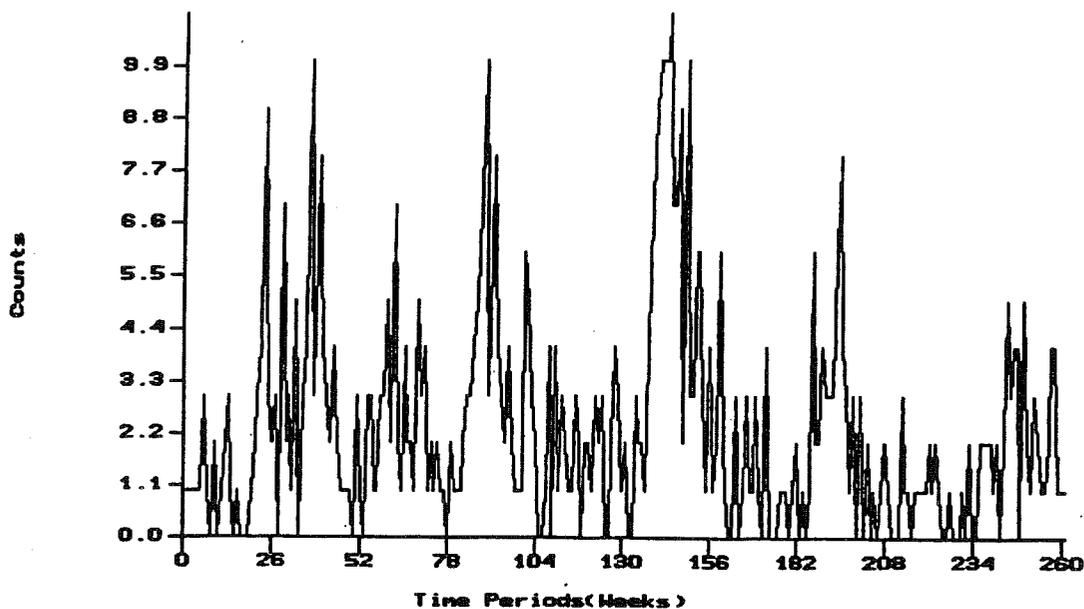
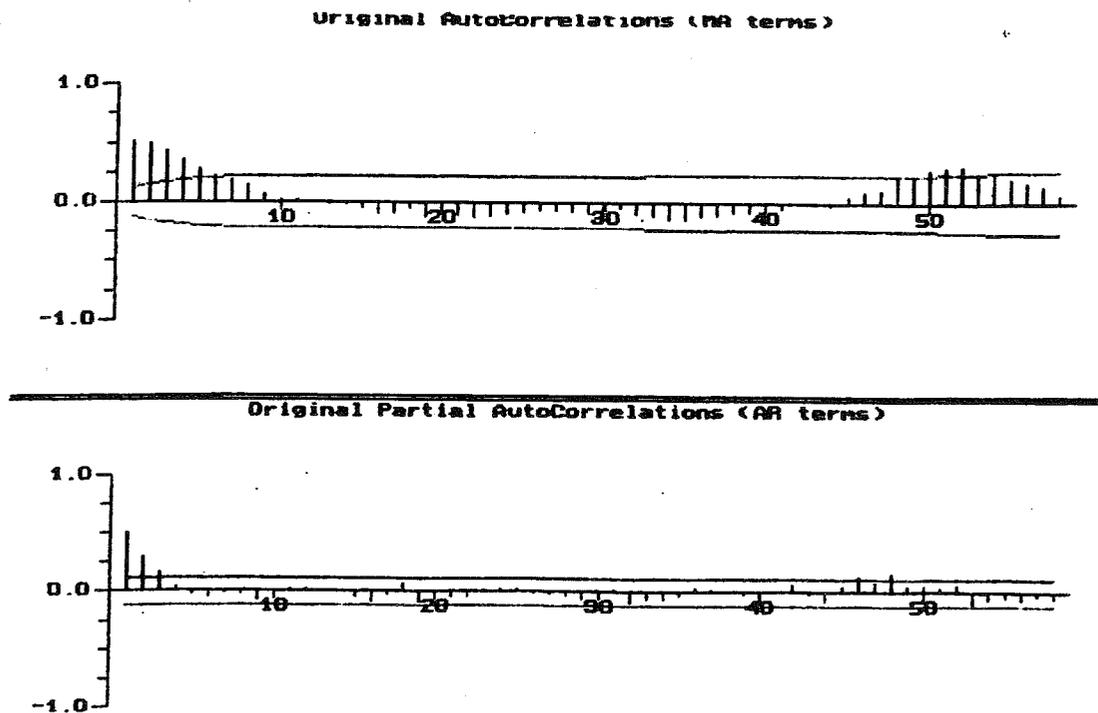


Figure 8 : Graphes de la fonction d'autocorrélation (ACF) et de la fonction d'autocorrélation partielle (PACF) de la série SPBe.



Dans un premier temps, nous avons effectué une différenciation d'ordre 52, qui réalise une série Y_t , transformée sur 52 semaines, telle que : $Y_t = \nabla_{52} X_t = X_t - X_{t-52}$ (figure 9).

Figure 9 : Série transformée Y_t (SPBe après différenciation d'ordre 52).

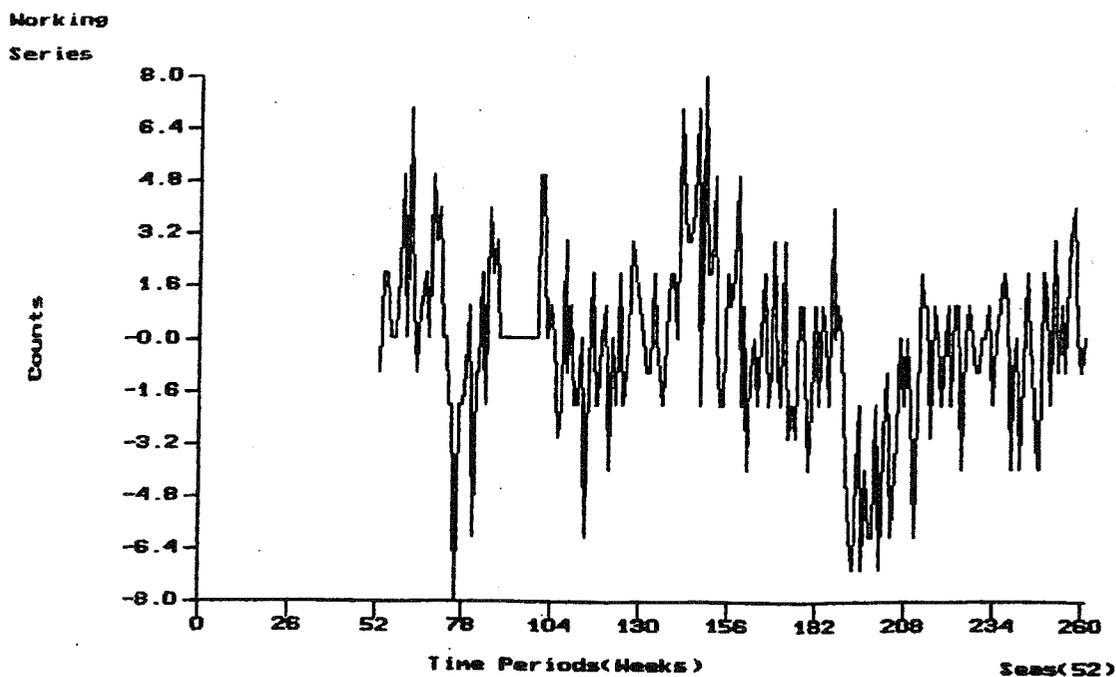
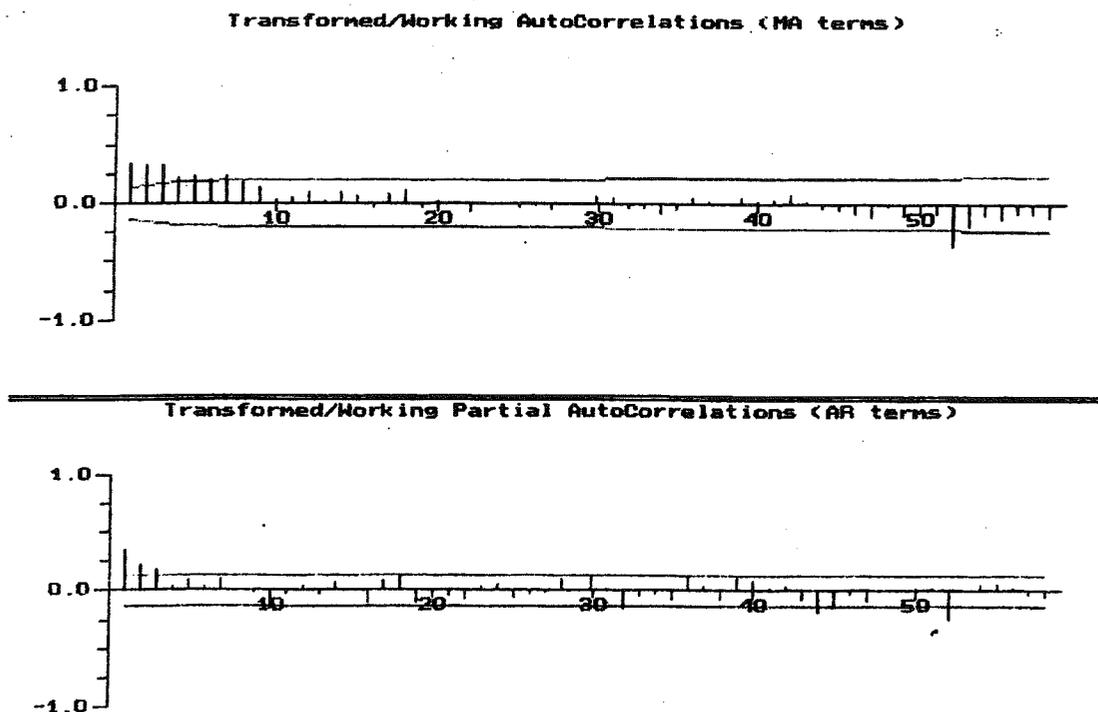


Figure 10 : Graphes des fonctions d'autocorrélation (ACF, PACF) de la série Y_t .



L'ACF de la série Y_t (figure 10) affiche qu'il persiste des pics significatifs dans les 10 premiers décalages, indiquant la nécessité d'ajouter une différenciation d'ordre 1, qui réalise une série Z_t telle que :

$$Z_t = \nabla_1 \nabla_{52} X_t = Y_t - Y_{t-1} = (X_t - X_{t-52}) - (X_{t-1} - X_{t-53}) .$$

La série Z_t , résultante d'une double différenciation, d'ordre 52 et d'ordre 1, est centrée sur 0 et semble osciller aléatoirement autour de cette valeur (figure 11). Le contrôle des ACF et PACF vérifie sa stabilisation (figure 12). Pour la tendance, l'ACF montre une autocorrélation significative au décalage 1 et une décroissance exponentielle de la PACF sur les six premiers décalages. Pour la saison, l'ACF est significatif au décalage 52. En comparaison avec le corrélogramme initial (figure 8), ces ACF et PACF confirment que la série Z_t est stationnaire.

Figure 11 : Série transformée Z_t (SPBe après double différenciation d'ordres 1 et 52).

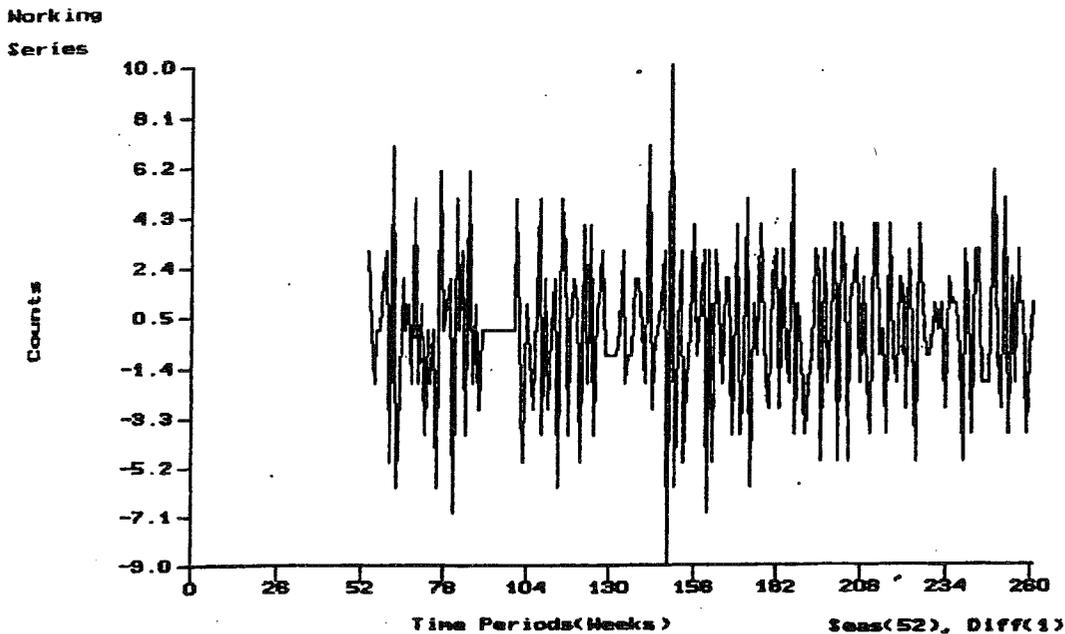
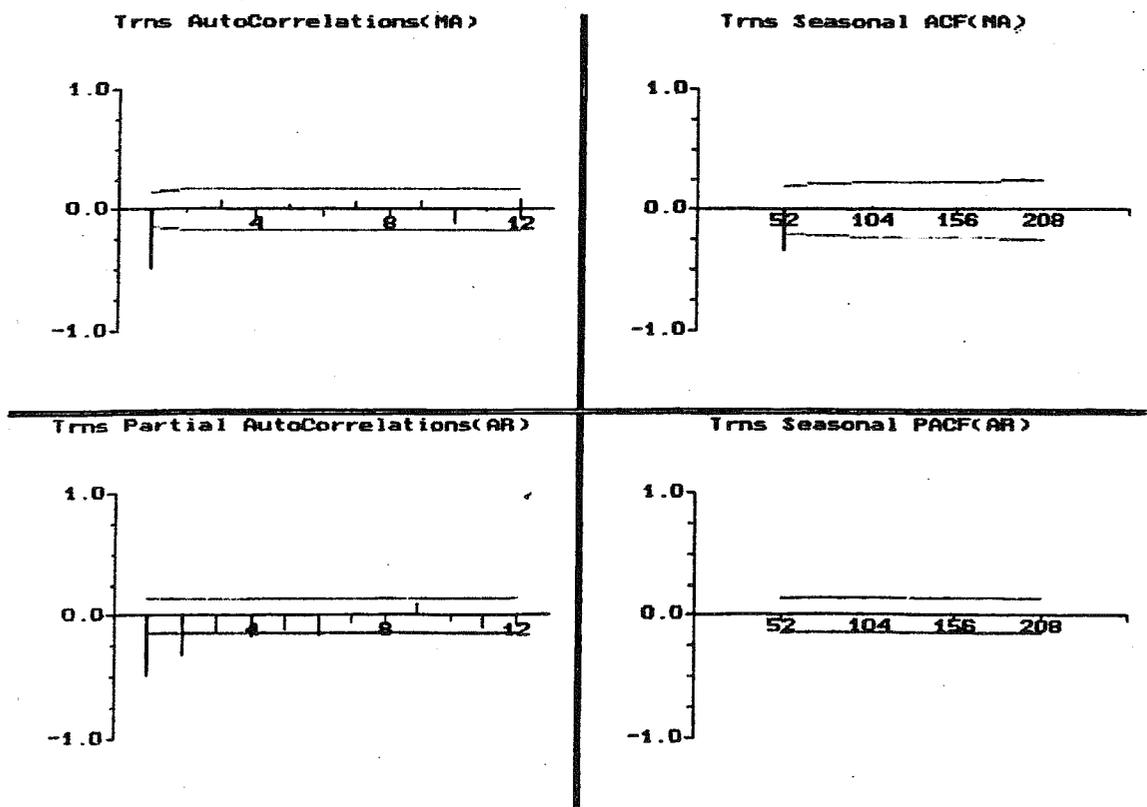


Figure 12 : Fonctions d'autocorrélation (ACF, PACF) de tendance (Trns) et de saison (Trns seasonal) de la série Z_t .



Sur les deux corrélogrammes de tendance, la PACF présente une décroissance exponentielle vers 0 et l'ACF présente un pic significatif isolé au décalage 1. Ceci correspond à un processus MA d'ordre 1. Sur les deux corrélogrammes de composantes saisonnières, un pic au décalage 52 est observé pour l'ACF et aucun pour la PACF indiquant un processus MA d'ordre 52. Le modèle suggéré est un modèle SARIMA (0.1.1) (0.1.1)₅₂ qui s'écrit :

$$\nabla_1 \cdot \nabla_{52} \cdot X_t = (1 - \theta B) (1 - \Theta B^{52}) \varepsilon_t$$

4-1-2-3 Estimations des paramètres et analyse des résidus

L'estimation des paramètres et les tests statistiques sont présentés dans le tableau II.

Tableau II : Diagnostic des paramètres et matrice des coefficients de corrélation

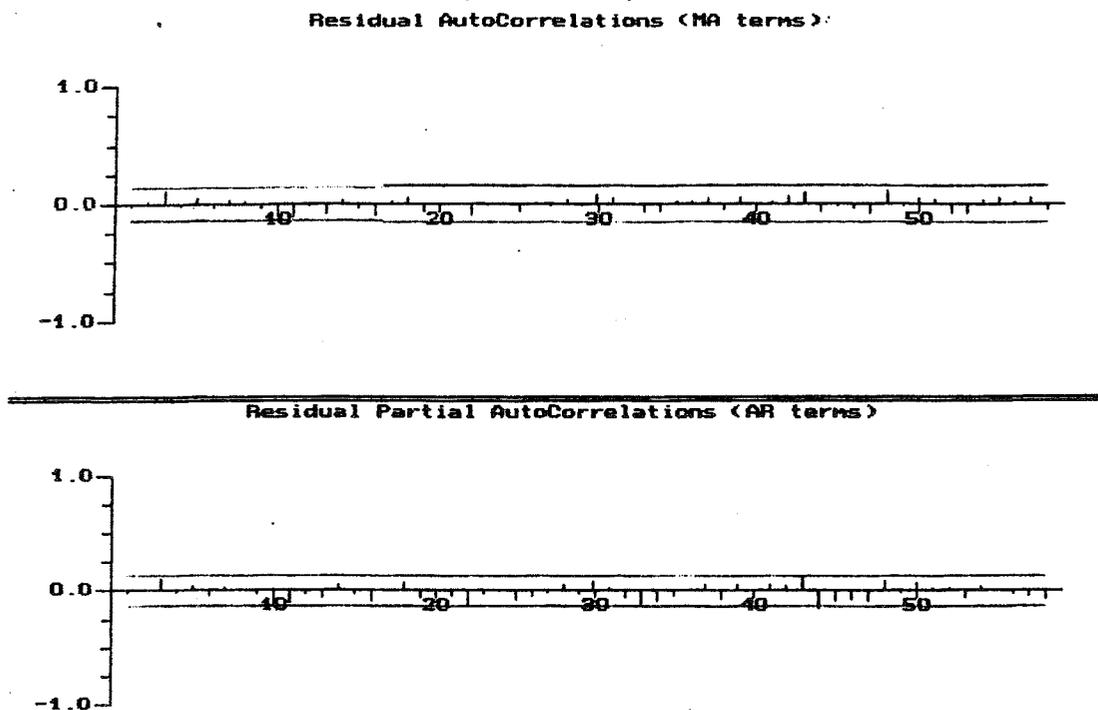
Type de paramètre (ordre)	Estimation	Ecart type	IC à 95%	t de Student	ddl	p
MA(1)	0.789	0.038	[0.714 - 0.864]	20.6	206	< 10 ⁻⁶
MA(52)	0.829	0.059	[0.714 - 0.944]	14.1	206	< 10 ⁻⁶

IC à 95% : intervalle de confiance à 95%, ddl : degrés de liberté, p : degré de significativité.

Matrice	MA(1)	MA(52)
MA(1)	1	
MA(52)	-2.59 x 10 ⁻²	1

Les deux paramètres MA(1) et MA (52) sont significativement différents de 0 (t de Student > 1.96 et 0 ∉ IC à 95%) et sont indépendants (coefficients de corrélation < 0.9). Par conséquent ils ont été conservés tous les deux dans le modèle. Le corrélogramme des valeurs résiduelles n'a montré aucun pic significatif (figure 13).

Figure 13 : Fonctions d'autocorrélation (ACF, PACF) des résidus.



Les résultats de l'analyse des résidus et les tests statistiques correspondants sont présentés dans les tableaux III et IV.

Tableau III : Diagnostic des résidus.

Moyenne résiduelle (μ_e)	0.008
Ecart réduit de la moyenne résiduelle	0.122
t de Student de la moyenne résiduelle	0.070
Variance résiduelle	3.101
Déviation standard résiduelle (σ_e)	1.761
Nombre de résidus négatifs	113
Nombre de résidus positifs	95
Nombre de croisements de l'axe	103
Valeur résiduelle minimale	-5.386
Valeur résiduelle maximale	4.507
Etendue des résidus	9.894

Tableau IV : Tests statistiques sur les résidus.

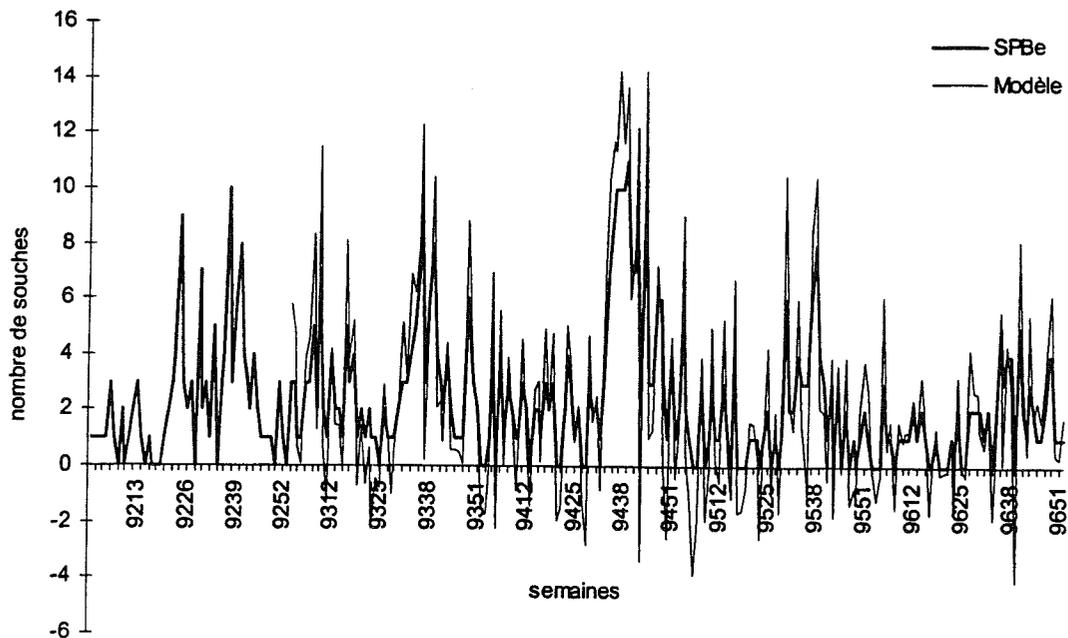
Tests	décalages	statistique Q	X ² (5%)	ddl	p
Box-Pierce	6	2.66	9.48	4	0.61
	12	7.90	18.30	10	0.63
	18	12.32	26.29	16	0.72
	24	15.07	33.92	22	0.85
Ljung-Box	6	2.73	9.48	4	0.60
	12	8.30	18.30	10	0.59
	18	13.11	26.29	16	0.66
	24	16.21	33.92	22	0.80

ddl : degrés de liberté, p : degré de significativité.

La moyenne des valeurs résiduelles est proche de 0 ($\mu_e = 0.008$) et montre une variation faible autour de cette valeur (écart type $\sigma_e = 1.761$).

Les tests de corrélation résiduelle de Box-Pierce et de Ljung-Box ne sont pas significatifs, indiquant une distribution aléatoire des valeurs résiduelles. La figure 14 montre l'ajustement du modèle aux données.

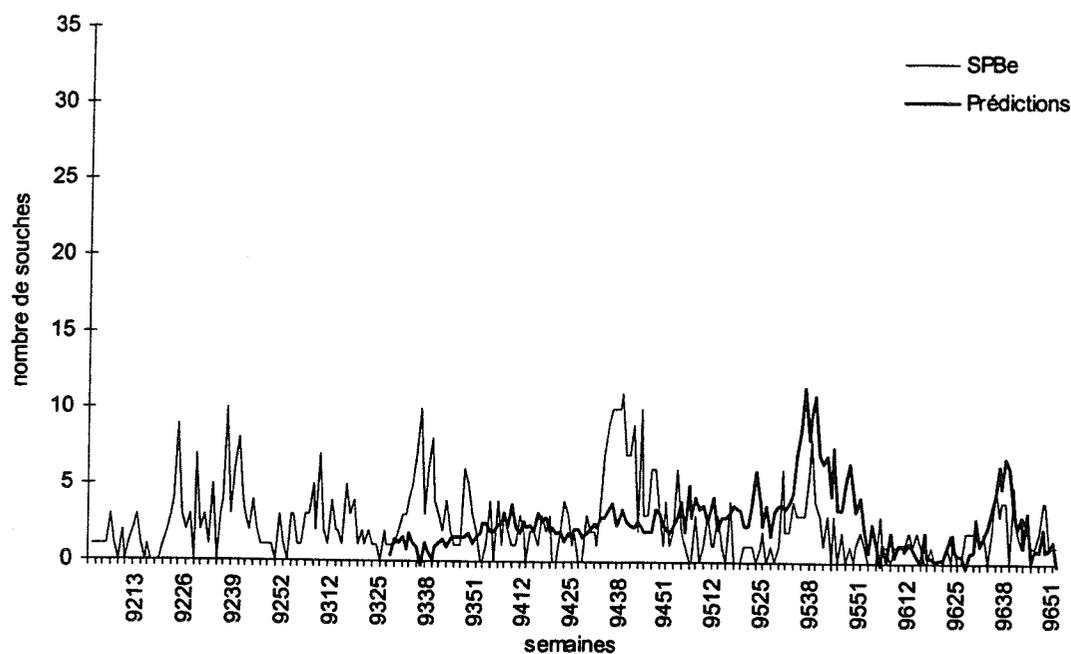
Figure 14 : Ajustement du modèle SARIMA (0.1.1)x(0.1.1)_{s2} à la série écriée (SPBe) pour la période 1992-1996.



4-1-2-4 Détermination du seuil d'alerte Box & Jenkins

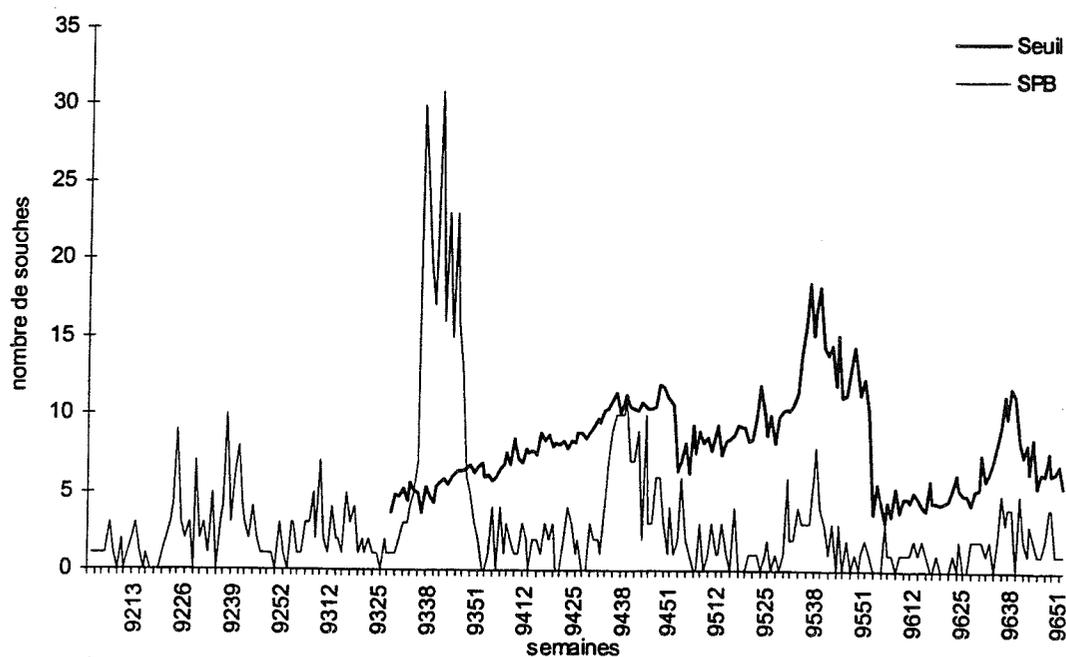
A partir du modèle SARIMA identifié, les valeurs prédites de chaque semaine $X_t(s)$ ont été calculées pas à pas pour chaque année. A chaque pas, le modèle a été appliqué sur les données de la série SPBe. Ainsi les prédictions de chaque pas sont basées sur les données disponibles des années antérieures. La figure 15 donne un aperçu visuel de la comparaison entre les données observées et les prédictions moyennes du modèle.

Figure 15 : Superposition de la série écrêtée SPBe et des prédictions moyennes du modèle SARIMA $(0.1.1) \times (0.1.1)_{52}$ sur la période juillet 1993 à décembre 1996.



La limite unilatérale à 95% des prédictions du modèle représente le seuil d'alerte pour *S. Paratyphi B* sur la période 1993-1996.

Figure 16 : Série originale SPB et seuil Box & Jenkins à 95% pour *S. Paratyphi B* sur la période juillet 1993 - décembre 1996.



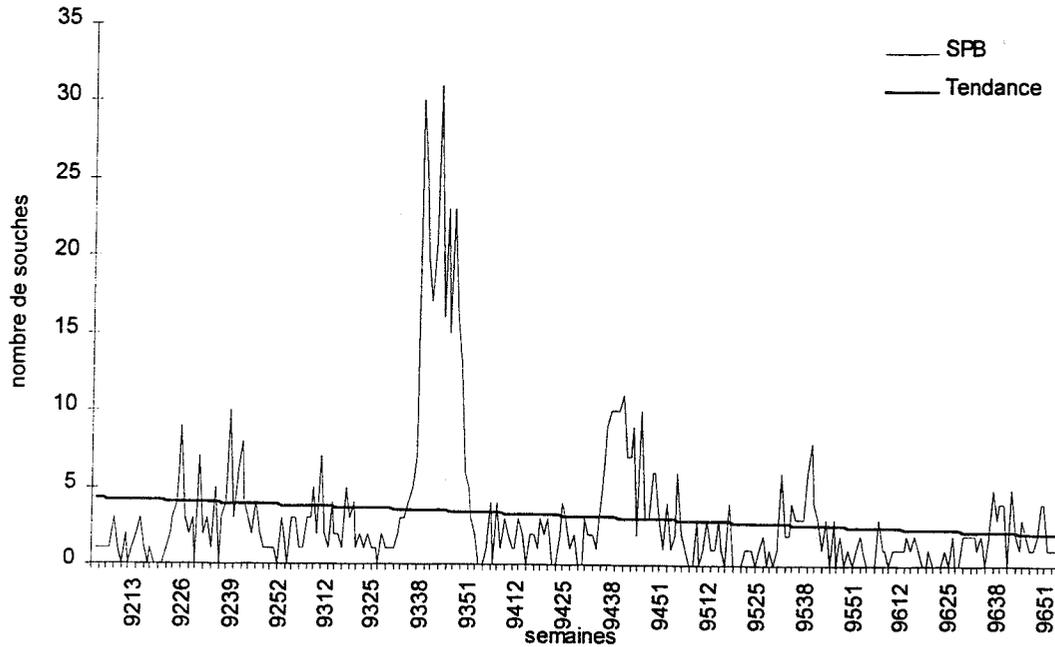
4-1-3 Modélisation spectrale pour *S. Paratyphi* B

4-1-3-1 Identification du modèle spectral

La première étape est l'estimation de la tendance linéaire de la série initiale SPB. La droite de régression obtenue s'écrit : $Y_t = 4.26 - 0.008.t$

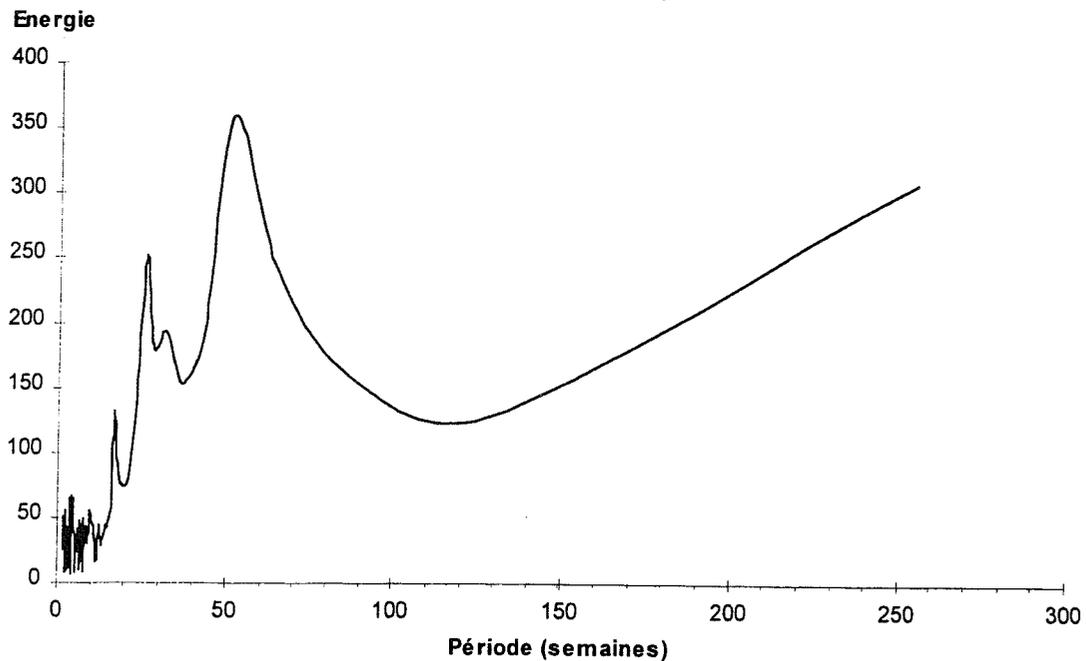
La pente de la tendance linéaire est faible et décroissante (figure 17). Le coefficient de régression R^2 est égal à 13.6%.

Figure 17 : Série originale SPB et tendance linéaire estimée pour la période 1992-1996.



Le périodogramme de la série SPB montre une périodicité de 52 semaines et des harmoniques significatives sur 32, 26 et 17 semaines.

Figure 18 : Périodogramme de la série *S. Paratyphi* B (SPB).



Pour l'ajustement du modèle, nous avons retenu :

- les quatre composantes cycliques 52, 32, 26 et 17, qui sont définies par leur amplitude (A1, A2, A3, A4) et leur phase ($\varphi_1, \varphi_2, \varphi_3, \varphi_4$) respectives;
- la tendance linéaire.

Le modèle s'écrit :

$$X_t = a + b.t + A_1.\cos(2\pi (t - \varphi_1)/52) + A_2.\cos(2\pi (t - \varphi_2)/32) + A_3.\cos(2\pi (t - \varphi_3)/26) + A_4.\cos(2\pi (t - \varphi_4)/17) + \varepsilon$$

La méthode des moindres carrés pour estimer les paramètres a donné les résultats regroupés dans le tableau V.

Tableau V : Paramètres de l'analyse spectrale pour la série S. Paratyphi B.

Tendance		Composantes cycliques Amplitude (A) / Phase (φ)			
a	b	52	32	26	17
4.986	-0.014	3.01 / -11.30	1.52 / -3.13	-1.92 / 2.01	-1.05 / -1.05

Avec ces estimations de paramètres, 62% de la variance de la série initiale est prise en compte par le modèle.

L'étape d'optimisation des paramètres exclue les valeurs "épidémiques" ayant une probabilité de moins de 5% d'être observées sur l'ensemble de la série (valeur "épidémique" > moyenne + 1.64 écart type). La valeur seuil de détection a été calculée à 10.76. Les valeurs "épidémiques" exclues ont été les semaines 93-36 à 93-47 et la semaine 94-39.

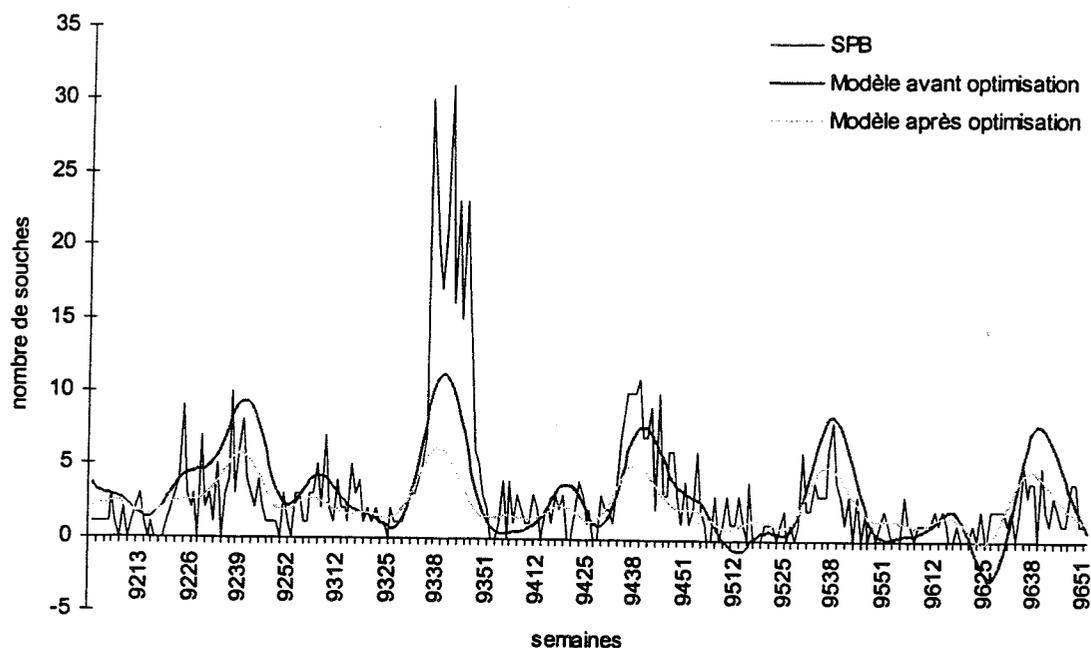
La méthode des moindres carrés appliquée à nouveau pour estimer les paramètres optimisés a donné les résultats regroupés dans le tableau VI.

Tableau VI : Paramètres optimisés pour l'analyse spectrale de la série S. Paratyphi B.

Tendance		Composantes cycliques Amplitude (A) / Phase (φ)			
a	b	52	32	26	17
3.166	-0.006	1.52 / -12.35	0.39 / -4.83	-0.97 / 0.21	0.60 / -11.02

Avec ces estimations de paramètres, 58.6% de la variance de la série initiale est prise en compte par le modèle final.

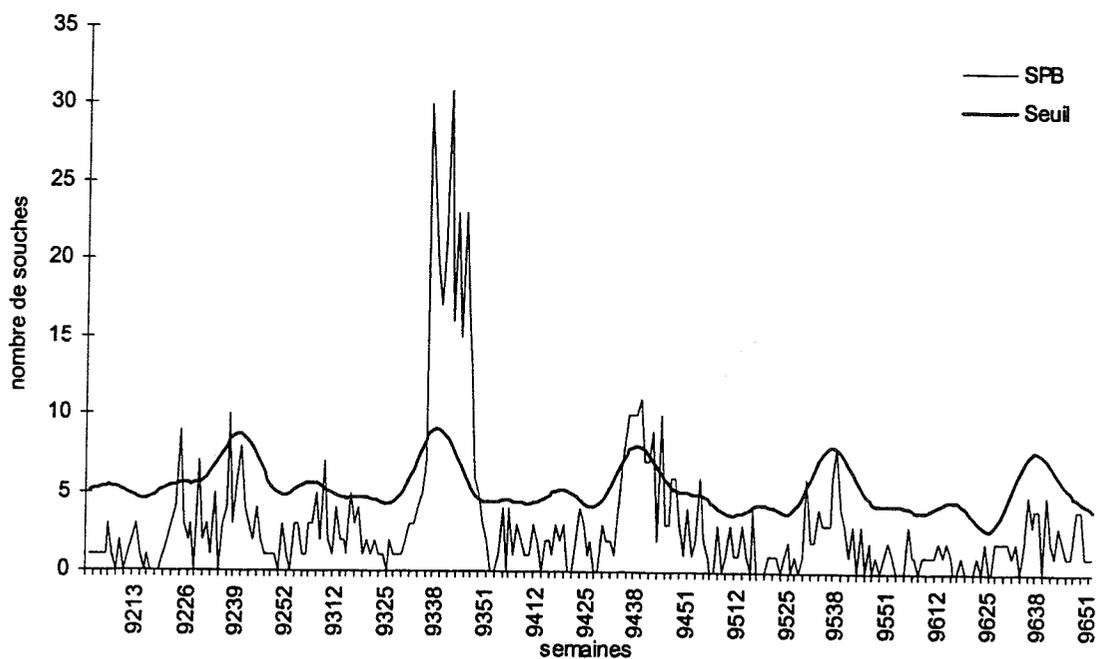
Figure 19 : Modèles estimés par analyse spectrale, avant et après retrait des points épidémiques de la série SPB pour la période 1992-1996.



4-1-3-2 Détermination du seuil d'alerte Serfling

Le seuil d'alerte Serfling est déterminé par le calcul de la limite unilatérale à 95% du modèle final obtenu après optimisation des paramètres (figure 20).

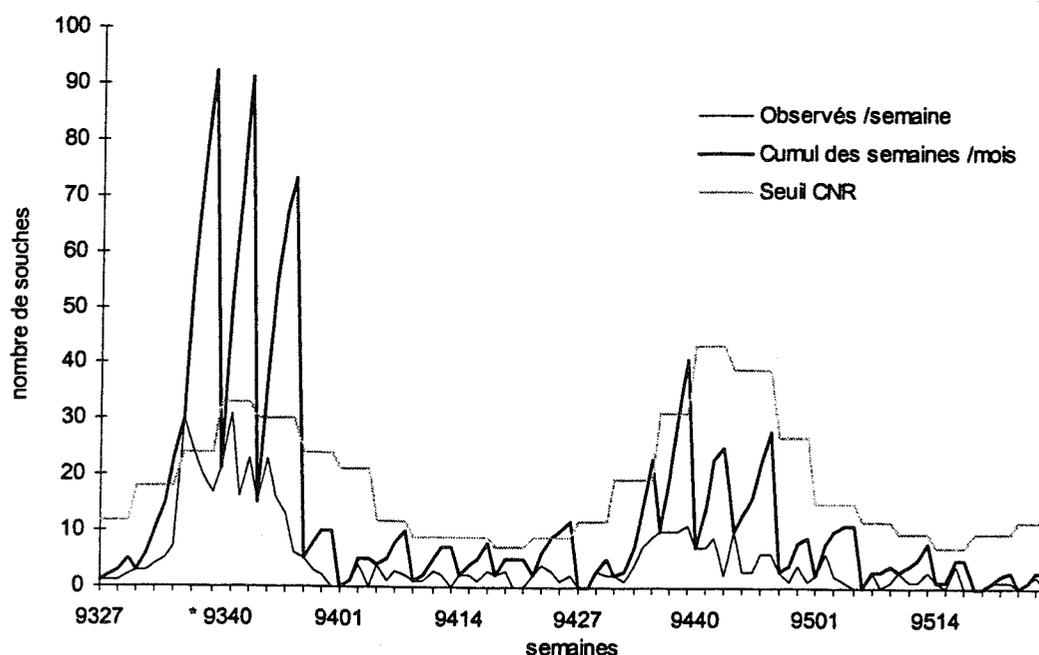
Figure 20 : Série originale *S. Paratyphi B* (SPB) et seuil Serfling à 95% pour la période 1992-1996.



4-1-4 Détermination du seuil CNR par méthode des moyennes historiques

Le seuil CNR correspond à une moyenne mensuelle attendue maximale qui est fixée à 20% de la moyenne mensuelle corrigée (M5c) : soit $M5c + 20\%$ (cf. 3.4.1). Chaque semaine, le nombre cumulé mensuel de souches est comparé au seuil. Le seuil CNR et les observations cumulées de *S. Paratyphi B* ont été reconstruites pour la période de deux années à analyser 1993-1995 afin de les rendre comparables aux seuils hebdomadaires Box - Jenkins et Serfling (figure 21).

Figure 21 : Seuil CNR, nombres hebdomadaires et cumulés mensuels de souches *S. Paratyphi B* pour la période juillet 1993 - juin 1995.



4-1-5 Détection des semaines épidémiques par chaque seuil et détermination des critères de jugement

Les trois seuils construits pour *S. Paratyphi B* ont été analysés et comparés sur la période 93-27 à 95-26 (annexe 3).

Le seuil CNR est dépassé sur les semaines (figure 21) :

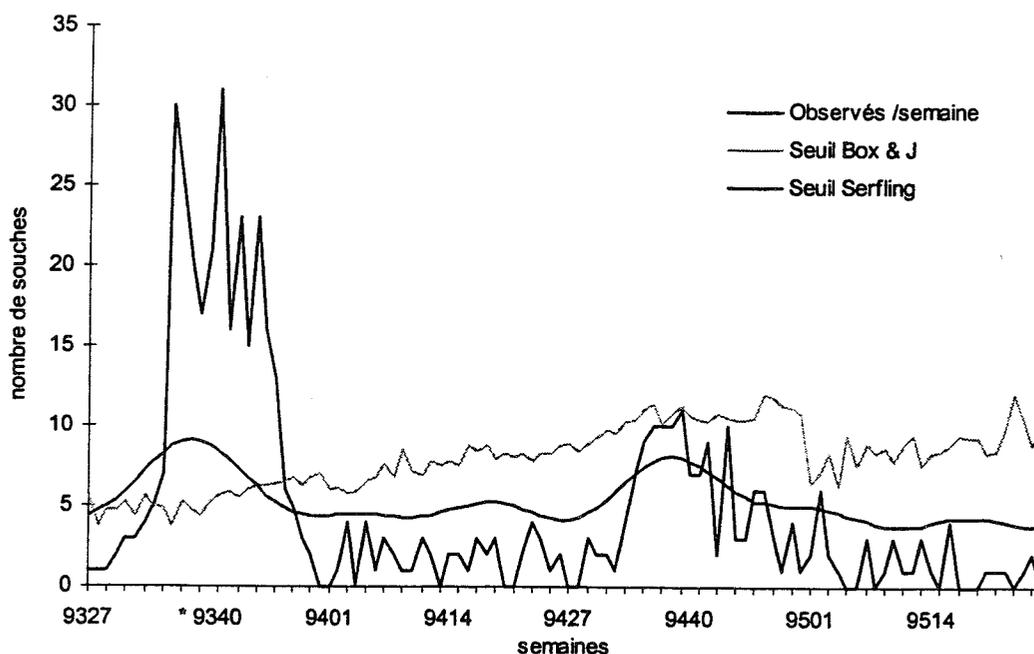
- 35 à 39, 41 à 43 et 45 à 48 en 1993 ;
- 24 à 26, 35 et 39 en 1994.

Le seuil Box & Jenkins oscille entre 3.2 et 18.7 (figure 22). Les semaines 34 à 47 de 1993 dépassent le seuil.

Le seuil Serfling oscille entre 2.8 et 9 (figure 22). Le seuil est dépassé pour les semaines:

- 36 à 49 en 1993 ;
- 34 à 39, 42, 44, 47 et 48 en 1994 ;
- 2 en 1995.

Figure 22 : Seuil Box & Jenkins et seuil Serfling superposés au nombre hebdomadaire observé de souches *S. Paratyphi B* pour la période juillet 1993 - juin 1995.



Sur les deux années qui couvrent la période du second semestre 1993 au premier semestre 1995 (semaines 93-27 à 95-26), il existe une épidémie à *S. Paratyphi B*. La période épidémique de *S. Paratyphi B* est définie de la semaine 93-33 à 93-47.

Sur la période épidémique de 15 semaines reconnue en 1993 (annexe 3), le seuil CNR déclenche l'alerte sur la semaine 93-35 et la maintient jusqu'à la semaine 93-48, excepté les semaines 93-40 et 93-44 correspondant aux premières semaines des mois d'octobre et de novembre respectivement. Une autre période est en alerte sur trois semaines consécutives en juin 1994, semaines 94-24 à 94-26. Deux semaines 94-35 et 94-39 sont en alerte de façon isolée.

Le seuil Box & Jenkins maintient une alerte constante de la semaine 93-34 à 93-47. Aucune autre semaine n'est détectée comme épidémique sur les deux années analysées.

Le seuil Serfling maintient une alerte en continu de la semaine 93-36 à 93-49. Une période de six semaines consécutives en août et septembre 1994 est détectée comme épidémique, semaines 94-34 à 94-39. Sur les mois d'octobre 1994 à janvier 1995, 5 semaines éparses sont détectées comme épidémiques : 94-42, 94-44, 94-47 et 94-48, 95-02.

A partir de la référence "brute" choisie, définissant les semaines "épidémiques" et "non investiguées" (cf. 3.5), pour comparer les trois seuils, nous avons établi trois tableaux de contingence suivant :

Pour le seuil CNR :

	Semaines "épidémiques"	Semaines "non épidémiques"	
Signal +	12	6	vpp = 0.67
Signal -	3	83	vpn = 0.97
	Se = 0.80	Sp = 0.93	

Total = 104

Pour le seuil
Box & Jenkins :

	Semaines "épidémiques"	Semaines "non épidémiques"	
Signal +	14	0	vpp = 1.00
Signal -	1	89	vpn = 0.99
	Se = 0.93	Sp = 1.00	

Total = 104

Pour le seuil Serfling :

	Semaines "épidémiques"	Semaines "non épidémiques"	
Signal +	12	11	vpp = 0.52
Signal -	3	78	vpn = 0.96
	Se = 0.80	Sp = 0.88	

Total = 104

Sur la période épidémique connue à *S. Paratyphi B*, le seuil Box & Jenkins est le plus réactif en détectant l'épidémie dès la semaine 93-34, versus 93-35 pour le seuil CNR et 93-36 pour le seuil Serfling. Il a également une meilleure prédiction de fin d'épidémie, prévoyant la fin d'épidémie à la semaine 93-47, versus 93-48 pour le seuil CNR et 93-49 pour le seuil Serfling. L'alerte maintenue par le seuil Box & Jenkins correspond au mieux à la période définie a priori comme épidémique.

Hors épidémie connue, le seuil Box & Jenkins n'est jamais dépassé. Sous l'hypothèse d'absence d'épidémie sur le reste des deux années, le seuil Box & Jenkins ne déclenche donc pas d'alerte injustifiée. En conséquence, ce seuil pour *S. Paratyphi B* semble réunir les meilleurs qualités de sensibilité et réactivité, de spécificité et de valeur prédictive.

4-1-6 Recherche de critères complémentaires

Le pic estival en 1994 est plus élevé que sur les autres années qui constitue la série chronologique. La courbe a dépassé le seuil Serfling de façon prolongée (6 semaines) mais n'a pas été détecté par le seuil CNR. L'alerte n'a pas été déclenchée. A posteriori, ce pic est suspect d'être une épidémie passée inaperçue. Nous avons analysé en détail la série chronologique selon des critères géographiques et de classes d'âge pour essayer de retrouver des caractéristiques susceptibles de confirmer le caractère épidémique ou non épidémique des variations de la série.

4-1-6-1 *Analyse par zones géographiques*

L'analyse régionale par régions administratives a montré la prédominance de quatre régions sur les 22 régions que comptent la France métropole : les régions Pays-de-Loire, Poitou-Charentes, Rhône-Alpes et Paris-Ile-de-France. Sur la période correspondant aux dépassements de seuils les plus fréquents (94-34 à 94-48), le tableau VII indique la répartition des cas par région et département et la part relative par rapport au total national.

Tableau VII : Nombre de souches de *S. Paratyphi B* reçues au CNR sur la période 94-34 à 94-48 par région et par département.

Régions	Départements	Nombre cumulé (%)	Nombre cumulé par région (%)	Taux d'incidence par région / 10 ⁶ (*)
Poitou-Charentes	16	0 /	13 (11.8)	8.0
	17	9 (8.2)		
	79	1 (0.9)		
	86	3 (2.7)		
Pays-de-la-Loire	44	1 (0.9)	18 (16.3)	5.7
	49	2 (1.8)		
	53	1 (0.9)		
	72	1 (0.9)		
	85	13 (11.8)		
Paris-Ile-de-France	75	2 (1.8)	19 (17.2)	1.7
	77	1 (0.9)		
	78	0 /		
	91	4 (3.6)		
	92	2 (1.8)		
	93	2 (1.8)		
	94	2 (1.8)		
	95	6 (5.5)		
Rhône-Alpes	1	0 /	16 (14.5)	2.9
	7	0 /		
	26	0 /		
	38	11 (10.0)		
	42	0 /		
	69	4 (3.6)		
	73	0 /		
	74	1 (0.9)		

*Taux d'incidence calculés sur la base des données de recensement démographiques régionales de 1995.

Dans la région Ouest, les départements 17 (Charente-Maritime) et 85 (Vendée) contigus concentrent la majorité des cas. Le cumul des deux régions administratives limitrophes (Poitou-Charentes et Pays-de-la-Loire) représente 28.1% (31 cas) du total national sur la même période.

Considérant la région géographique constituée par les départements 17 et 85, et les départements contigus 16, 24, 33, 44, 49 et 79, le nombre de cas cumulés est de 27 cas soit 24.5%.

Le tableau VIII indique les proportions de cas pour les mêmes régions lors de l'épidémie de 1993.

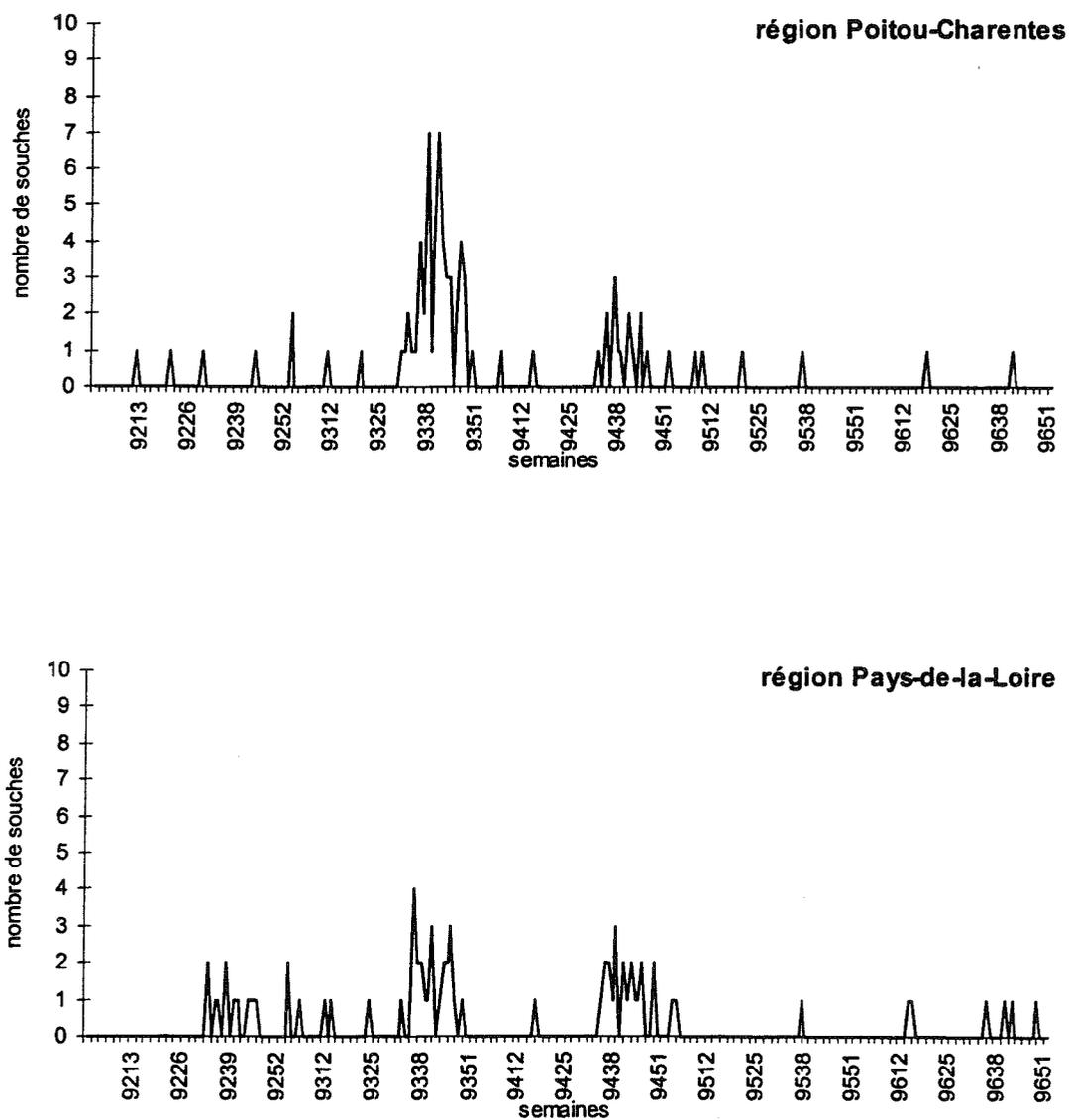
Tableau VIII : Nombre de souches de *S. Paratyphi B* reçues au CNR sur la période 93-33 à 93-47 par région et par département.

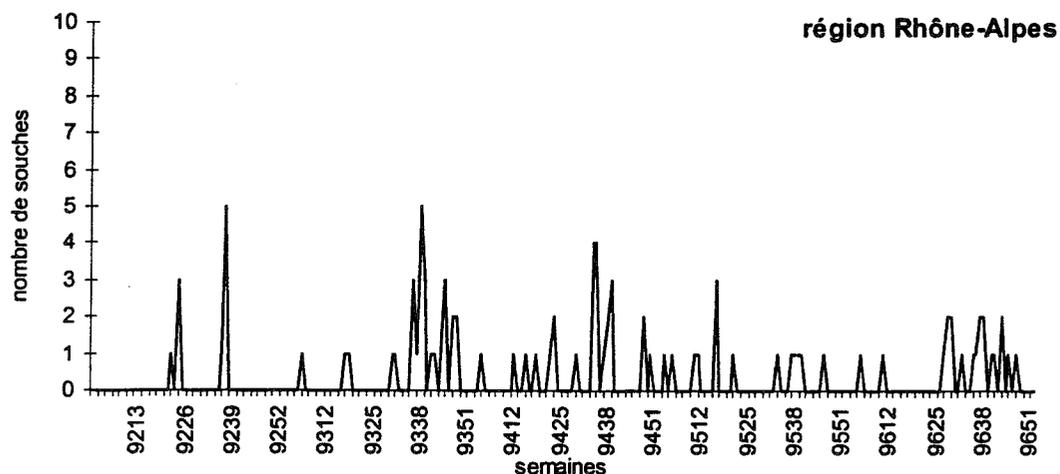
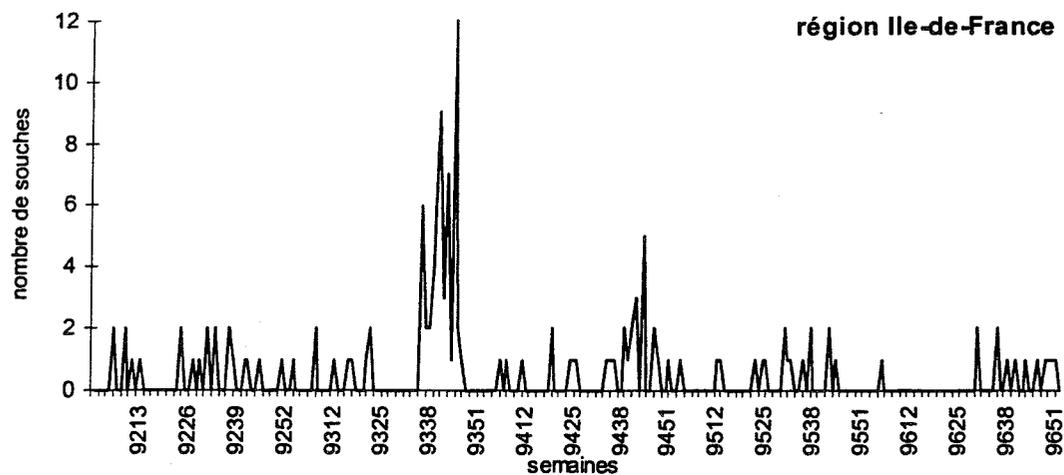
Régions	Départements	Nombre cumulé (%)	Nombre cumulé par région (%)	Taux d'incidence par région / 10 ⁶ (*)
Poitou-Charentes	16	3 (1.1)	45 (16.9)	27.8
	17	12 (4.5)		
	79	14 (5.3)		
	86	16 (6.0)		
Pays-de-la-Loire	44	3 (1.1)	23 (8.6)	7.3
	49	15 (5.6)		
	53	0		
	72	2 (0.8)		
	85	3 (1.1)		
Paris-Ile-de-France	75	10 (3.8)	55 (20.7)	5.0
	77	2 (0.8)		
	78	12 (4.5)		
	91	3 (1.1)		
	92	16 (6.0)		
	93	3 (1.1)		
	94	5 (1.9)		
	95	4 (1.5)		
Rhône-Alpes	1	1 (0.4)	21 (8.0)	3.8
	7	1 (0.4)		
	26	1 (0.4)		
	38	7 (2.6)		
	42	2 (0.8)		
	69	7 (2.6)		
	73	0		
	74	2 (0.8)		

*Taux d'incidence calculés sur la base des données de recensement démographiques régionales de 1995.

La figure 23 montre les séries hebdomadaires obtenues sur la période 1992-1996 pour les quatre régions. Les graphiques temporels font apparaître pour chaque région les concentrations de cas inhabituelles en comparaison avec le reste de la série correspondante.

Figure 23 : Séries hebdomadaires de souches *S. Paratyphi B* reçues au CNR pour les régions Poitou- Charentes, Pays-de-la-Loire, Ile-de-France et Rhône-Alpes en 1992-1996.

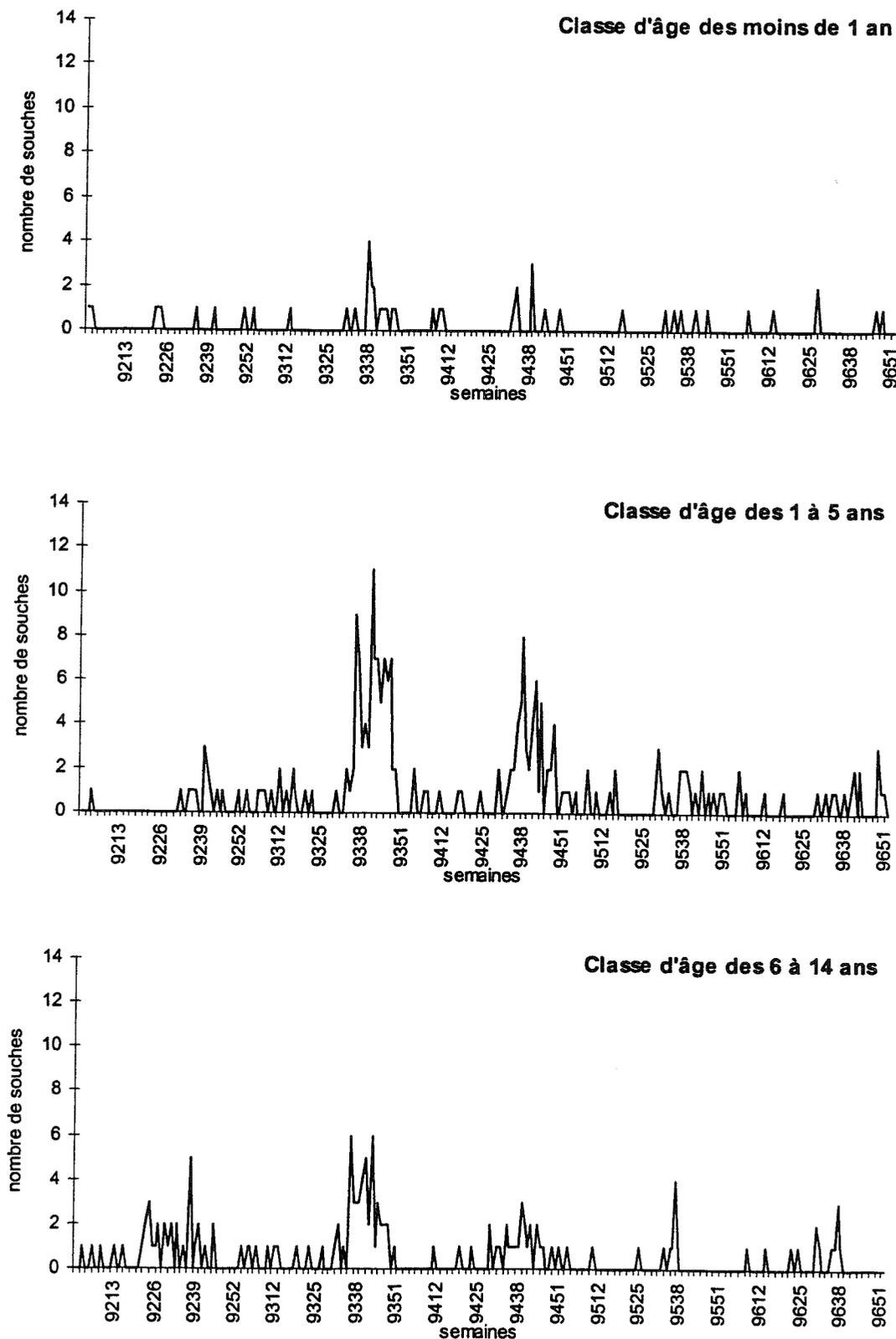


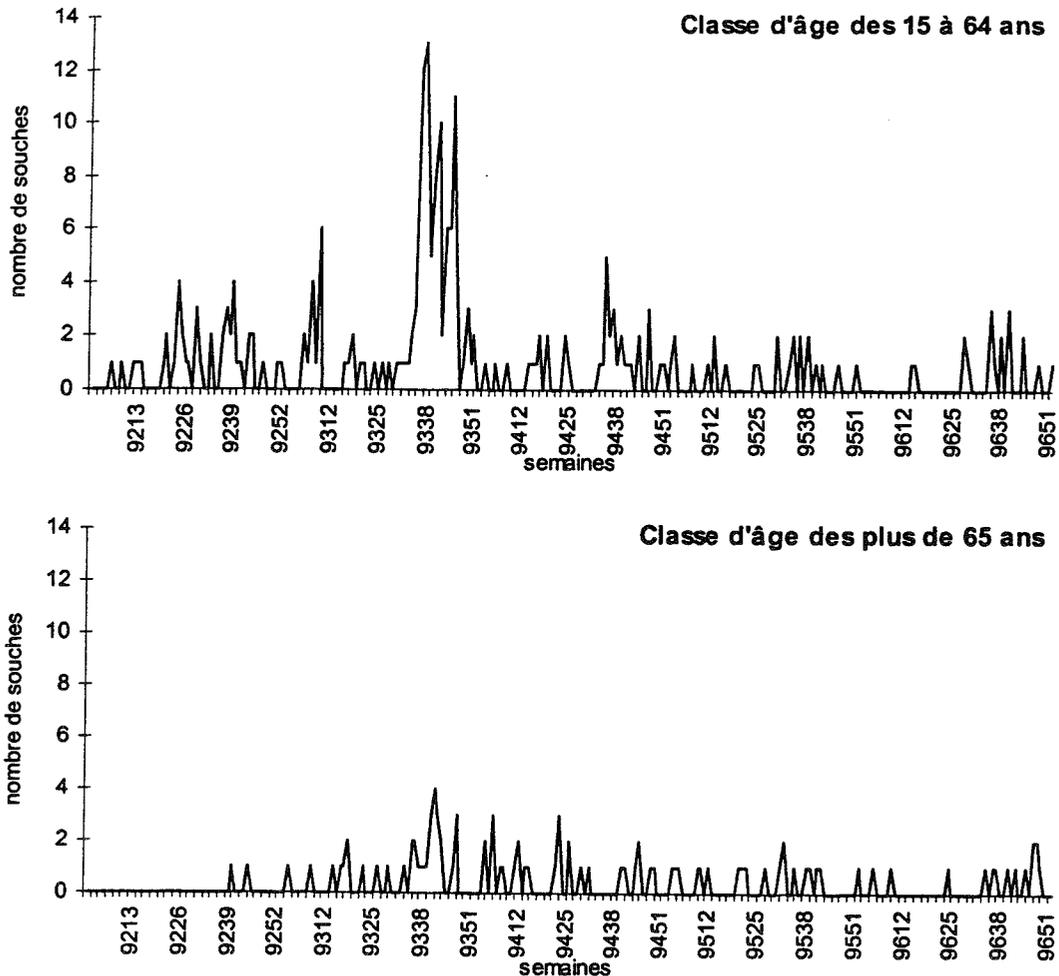


4-1-6-2 Analyse par classe d'âge des cas

Un accroissement de souches *S. Paratyphi B* apparaît de façon concomitante sur la période d'été 1993 (semaines 93-35 à 93-47) pour plusieurs classes d'âge dont en particulier les classes d'âge des 1-5 ans et des 15-64 ans (figure 24). Il correspond à la période définie comme épidémique qui a été investiguée. Un pic de *S. Paratyphi B* apparaît en été 1994 (semaines 94-35 à 94-48) dans la classe d'âge des moins de 5 ans. En comparaison avec les autres classes d'âge sur la même période, les souches reçues au CNR semble plus particulièrement concentrées dans cette classe d'âge. Un peu moins marqué que le pic épidémique de 1993 pour la même classe d'âge, il se détache nettement du niveau des autres années considérées comme normales.

Figure 24 : Séries hebdomadaires de souches *S. Paratyphi B* reçues au CNR par classes d'âge pour la période 1992-1996.





4-2 Résultats pour *S. Typhimurium*

4-2-1 Descriptif de la série originale et de l'épidémie reconnue et investiguée

La série originale hebdomadaire *S. Typhimurium* (STM) est une série avec une tendance croissante depuis 1994 et une saisonnalité d'environ 52 semaines (cf. 3.3, figure 2, p 7). La série présente une structure sinusoïdale annuelle apparente qui oscille entre 5 et 120 souches de salmonelles par semaine.

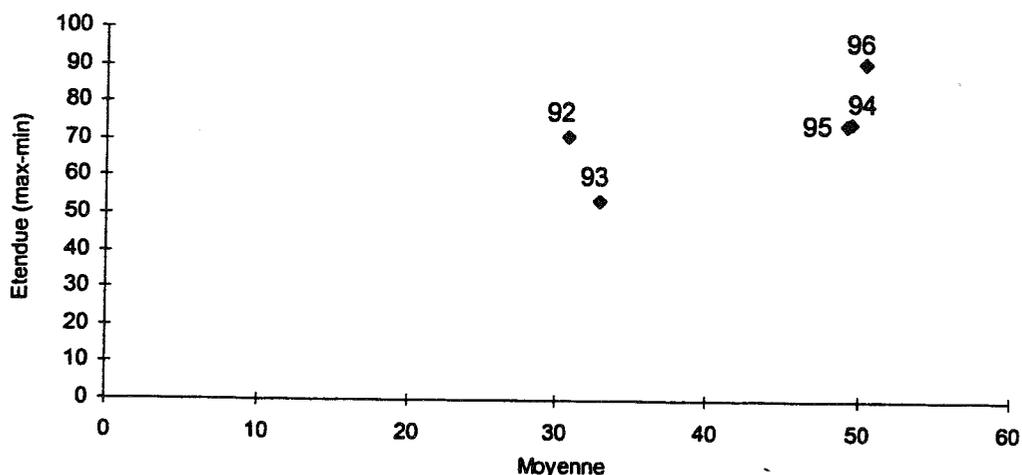
L'unique épidémie reconnue se situe au printemps été 1997. Elle a débuté la deuxième semaine de mai 1997. C'est sur signalement de la DDASS du Jura que l'augmentation du nombre de cas et leur concentration dans le sud du département ont pu être constatées et que le RNSP a été sollicité pour intervenir. Au total 113 cas ont été recensés, dont 109 confirmés et 4 probables, entre le 12 mai et le 4 juillet 1997 (semaines 97-20 à 97-27). Le sex ratio était de 1 et la classe d'âge des enfants de 1 à 5 ans était la plus touchée. Les résultats de l'investigation ont indiqué que cette épidémie communautaire était attribuable à la consommation d'un même fromage au lait cru (OR = 6,5 ; IC95%: 1,4-28,8) provenant d'un producteur unique et appartenant très probablement à un même lot. La surveillance par le CNR n'a pas permis de détecter cette épidémie avant ou au même moment que l'alerte soit donnée par les laboratoires auprès de la DDASS. Le CNR indiquait en effet que le nombre de souches *S. Typhimurium* reçues sur avril - mai 1997 pour le Jura et les départements avoisinants n'était pas augmenté. Cependant les laboratoires du Jura n'avaient pas envoyé leur souches au CNR (37).

4-2-2 Modélisation ARIMA pour S Typhimurium

4-2-2-1 Détection et remplacement des "points épidémiques"

Le graphique de la moyenne-étendue ne met pas en évidence d'année exceptionnelle (figure 25), malgré la présence de quelques pics apparents plus marqués sur la série d'origine. Aucune détection et aucun remplacement de point épidémique n'est de ce fait nécessaire.

Figure 25 : Graphique de la moyenne étendue de la série originale S. Typhimurium (STM) pour la période 1992-1996.



4-2-2-2 Identification du modèle ARIMA

L'homogénéité des variations annuelles ne nécessitent pas une transformation pour stabiliser la variance de la série.

L'inspection du graphique de la série STM (figure 26) et de sa fonction d'autocorrélation (figure 27) montrent une tendance linéaire et une saisonnalité de 52 semaines. En effet, l'ACF décroît vers 0 dans les vingt premiers intervalles de temps et remonte, telle une courbe sinusoïde, autour de 52 semaines.

Figure 26 : Série originale de S. Typhimurium (STM) pour les 261 semaines de la période 1992-1996 (261 décalages).

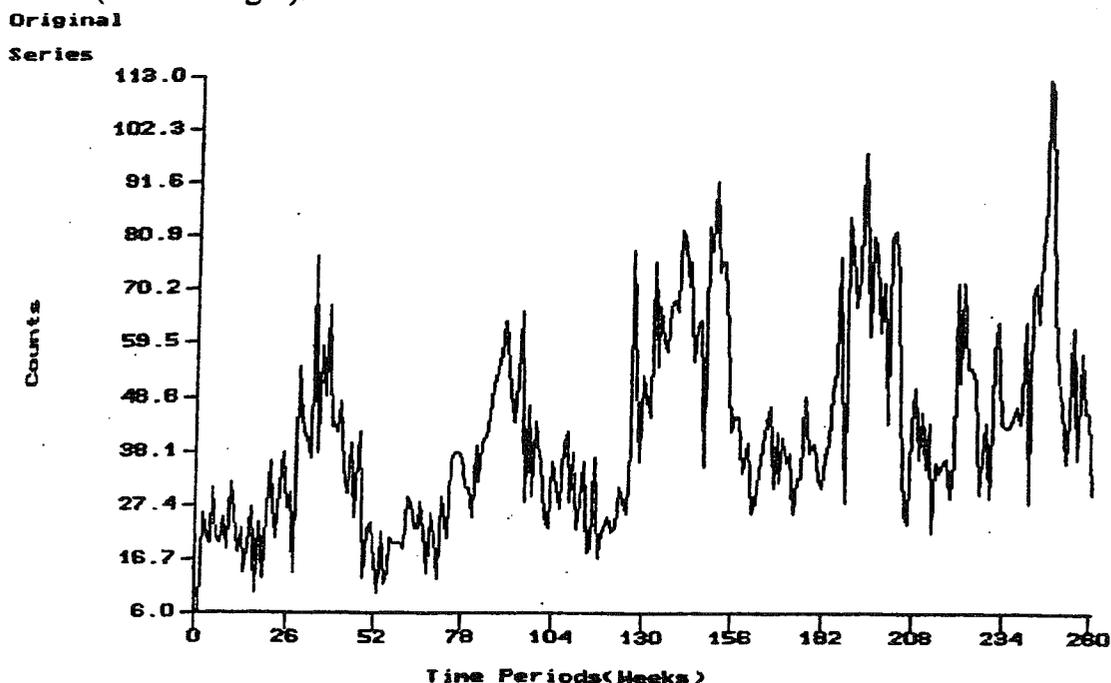
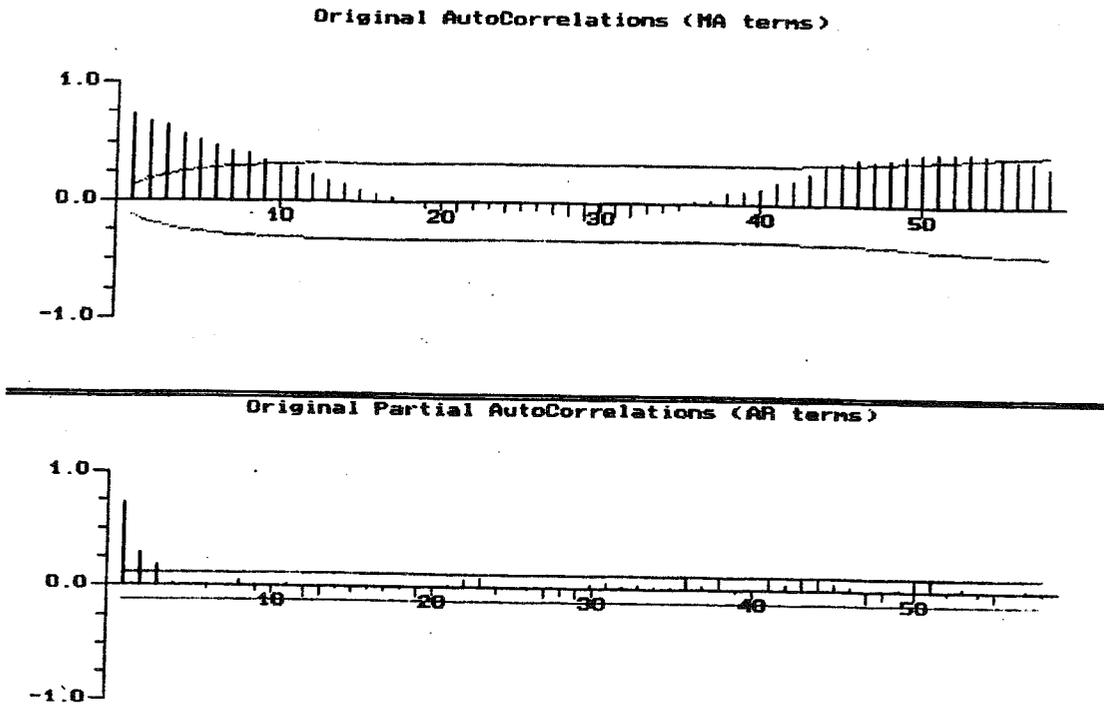


Figure 27 : Graphes de la fonction d'autocorrélation (ACF) et de la fonction d'autocorrélation partielle (PACF) de la série STM.



Dans un premier temps, il faut donc introduire une différenciation d'ordre 52 qui réalise une série Y_t transformée sur 52 semaines (figure 28), telle que :

$$\nabla^{52} X_t = Y_t = X_t - X_{t-52}.$$

Figure 28 : Série transformée Y_t .

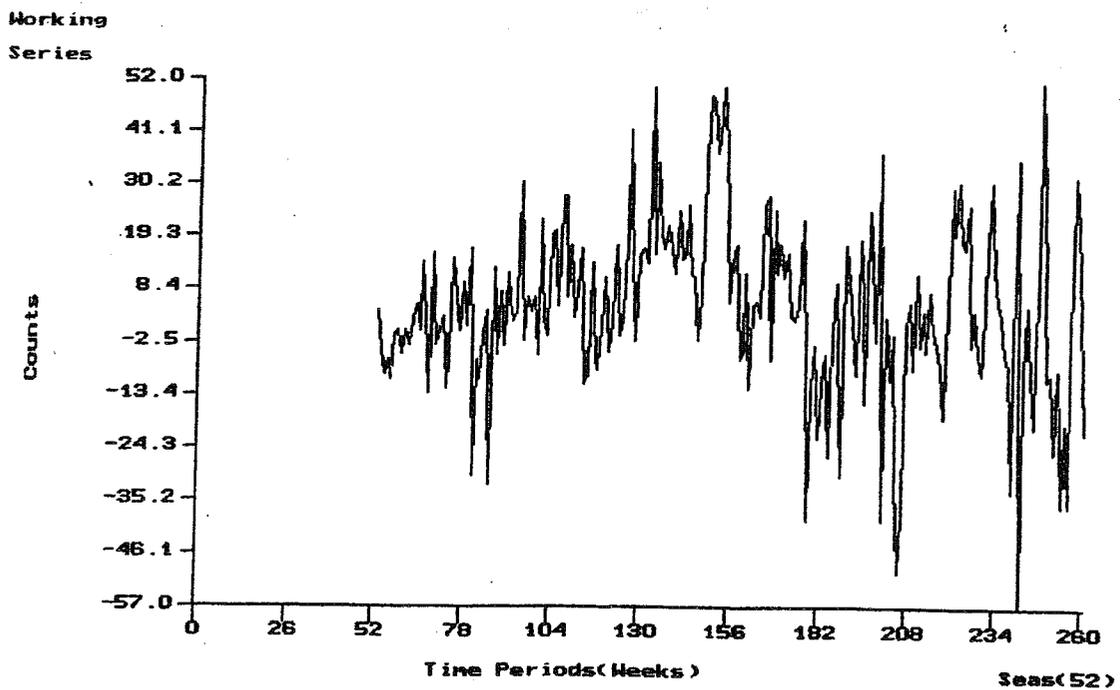
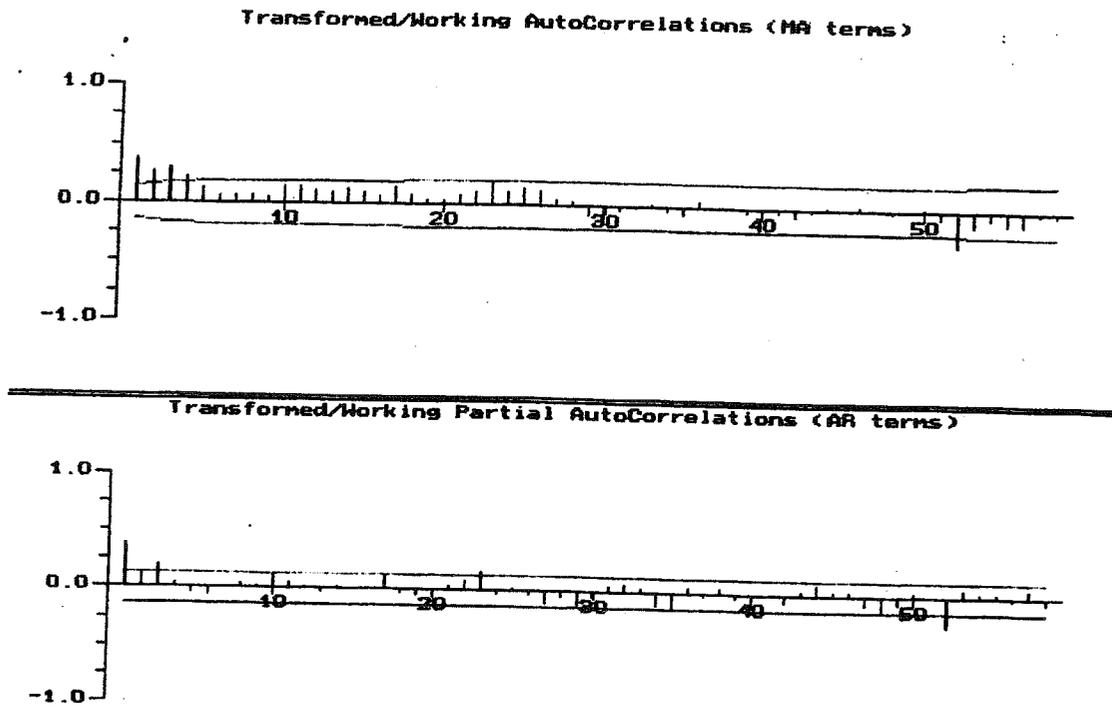


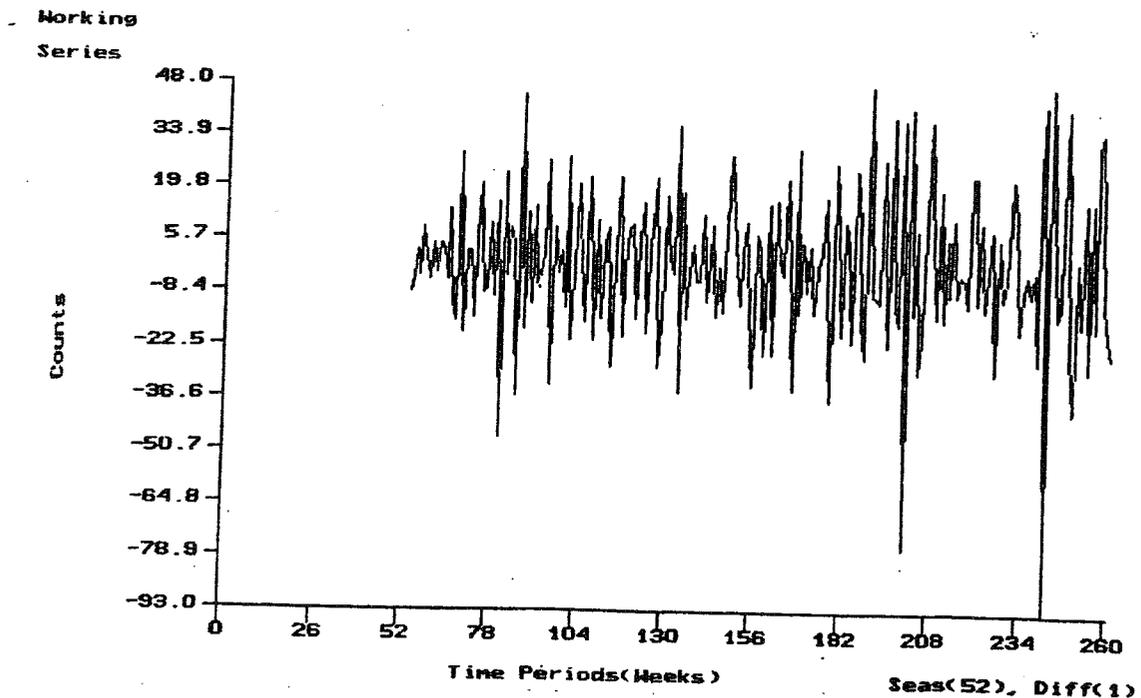
Figure 29 : Graphes des fonctions d'autocorrélations (ACF, PACF) de la série Y_t .



L'ACF de la série Y_t affiche qu'il persiste des pics significatifs dans les 10 premiers décalages, indiquant la nécessité d'ajouter une différenciation d'ordre 1, qui réalise une série Z_t (figure 30) telle que :

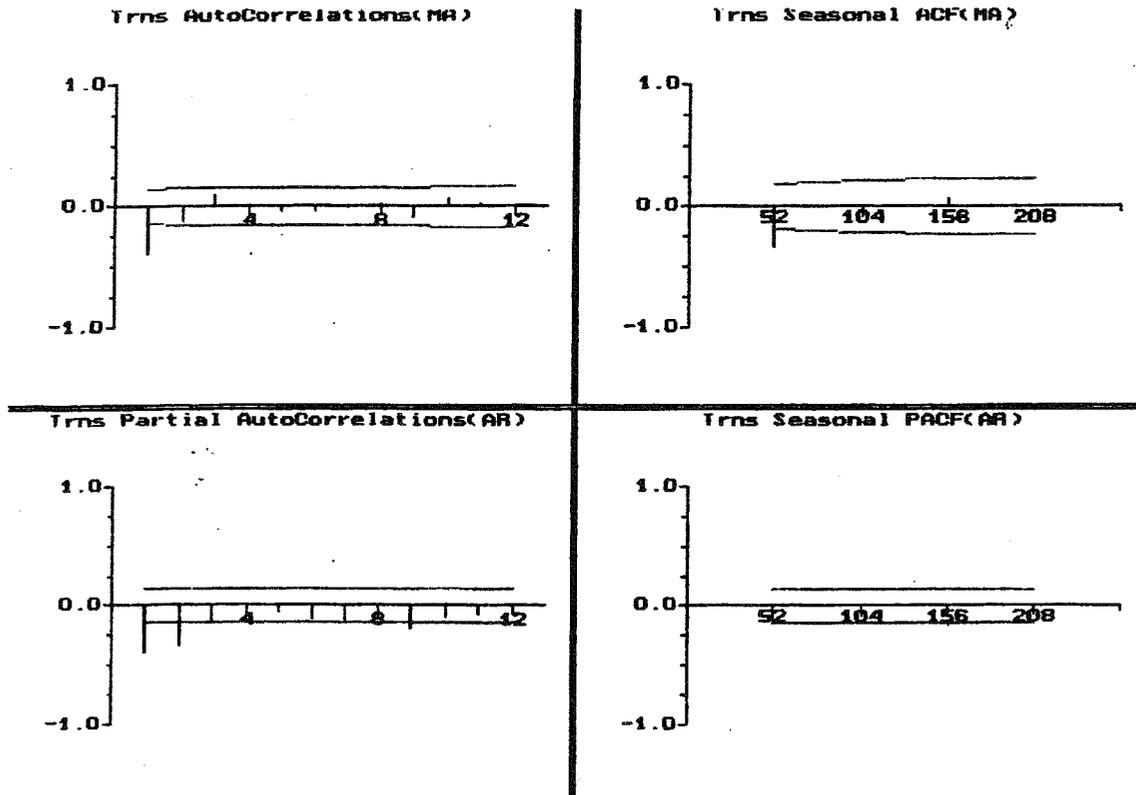
$$\nabla \nabla_{52} X_t = Z_t = Y_t - Y_{t-1} = (X_t - X_{t-52}) - (X_{t-1} - X_{t-53}).$$

Figure 30 : Série transformée Z_t .



La série Z_t résultante d'une double différenciation, d'ordre 52 et d'ordre 1, est centrée sur 0 et semble osciller aléatoirement autour de cette valeur. Le contrôle des ACF et PACF vérifie sa stabilisation (figure 31).

Figure 31 : Fonctions d'autocorrélation (ACF, PACF) de tendance (Trns) et de saison (Trns seasonal) de la série Z_t .



L'ACF montre une autocorrélation significative aux décalages 1 et 52 avec décroissance exponentielle sur les trois premiers décalages. La PACF indique une décroissance exponentielle sur les quatre premiers décalages et une autocorrélation significative au décalage 52. Ce corrélogramme confirme, en comparaison avec le corrélogramme initial, que la série Z_t est stationnaire. Les décroissances exponentielles de l'ACF et de la PACF suggèrent un processus AR d'ordre 1 et MA d'ordre 1. La persistance d'une autocorrélation sur le décalage 52 indique la nécessité de modéliser également la saisonnalité résiduelle.

Le modèle ainsi suggéré est un modèle SARIMA (1.1.1) (0.1.1)₅₂ qui s'écrit :

$$\nabla_1 \cdot \nabla_{52} \cdot \phi(B) \cdot X_t = \theta(B) \cdot \Theta(B^{52}) \cdot \varepsilon_t$$

4-2-2-3 Estimations des paramètres et analyse des résidus

L'estimation des paramètres et les tests statistiques sont présentés dans le tableau IX.

Tableau IX : Diagnostic des paramètres et matrice des coefficients de corrélation

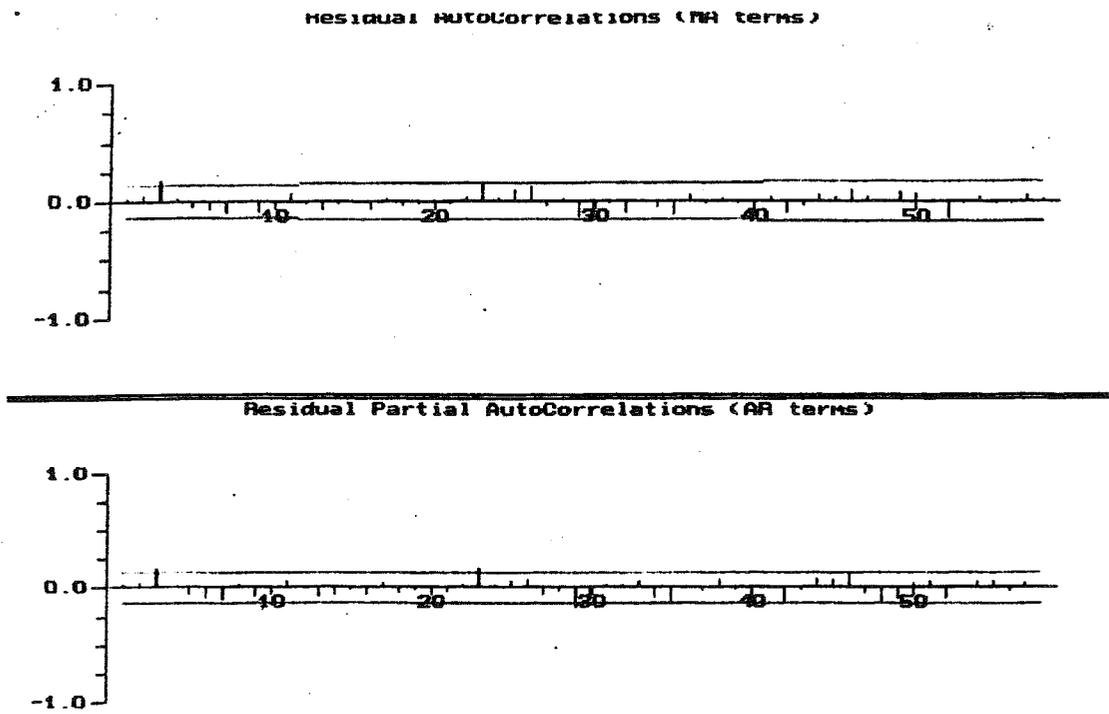
Type de paramètre (ordre)	Estimation	Ecart type	IC à 95%	t de Student	ddl	p
AR(1)	0.260	0.072	[0.118 - 0.402]	3.6	205	$< 10^{-3}$
MA(1)	0.920	0.030	[0.862 - 0.979]	30.9	205	$< 10^{-6}$
MA(52)	0.772	0.069	[0.636 - 0.908]	11.1	205	$< 10^{-6}$

IC à 95% : intervalle de confiance à 95%, ddl : degrés de libertés, p : degré de significativité.

Matrice	AR(1)	MA(1)	MA(52)
AR(1)	1		
MA(1)	4.98×10^{-1}	1	
MA(52)	-2.72×10^{-2}	-4.53×10^{-3}	1

Les trois paramètres sont significativement différent de 0 (t de Student > 1.96 et $0 \notin$ IC à 95%) et sont indépendants (coefficient de corrélation < 0.9). Par conséquent ils ont été conservés tous les trois dans le modèle. Le corrélogramme des valeurs résiduelles montre quelques pics à la limite de la significativité qui n'indiquent aucune corrélation interprétable (figure 32).

Figure 32 : Fonctions d'autocorrélation (ACF, PACF) des résidus.



Les résultats de l'analyse des résidus et les tests statistiques correspondants sont présentés dans les tableaux X et XI.

Tableau X : Diagnostic des résidus.

Moyenne résiduelle (μ_e)	-0.3082
Ecart réduit de la moyenne résiduelle	0.8847
t de Student de la moyenne résiduelle	-0.3483
Variance résiduelle	162.81
Déviatiion standard résiduelle (σ_e)	12.76
Nombre de résidus négatifs	110
Nombre de résidus positifs	98
Nombre de croisements de l'axe	102
Valeur résiduelle minimale	-48.40
Valeur résiduelle maximale	39.83
Etendue des résidus	88.34

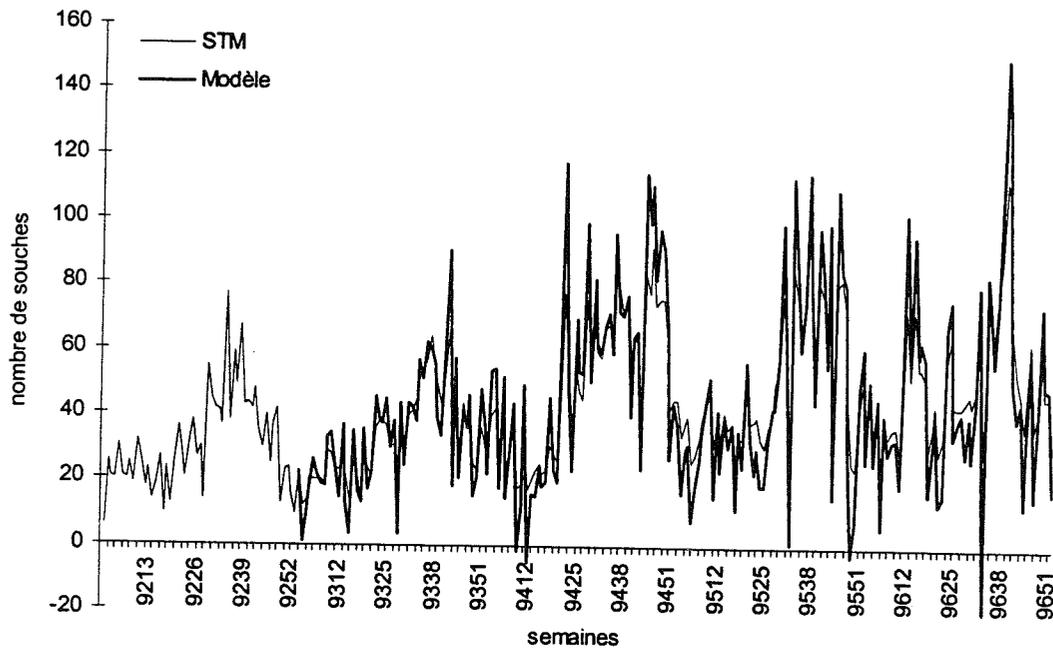
Tableau XI : Tests statistiques sur les résidus

Tests	décalages	statistique Q	X ² (5%)	ddl	p
Box-Pierce	6	6.88	7.81	4	0.075
	12	11.38	16.92	8	0.250
	18	13.11	24.99	16	0.593
	24	18.97	32.67	22	0.587
Ljung-Box	6	7.06	7.81	4	0.070
	12	11.80	16.91	8	0.224
	18	13.69	24.99	16	0.549
	24	20.33	32.67	22	0.500

ddl : degrés de liberté, p : degré de significativité.

Les X² sont tous non significatifs ($p > 0.05$) indiquant que les résidus ne sont pas corrélés. La figure 33 montre l'ajustement du modèle aux données.

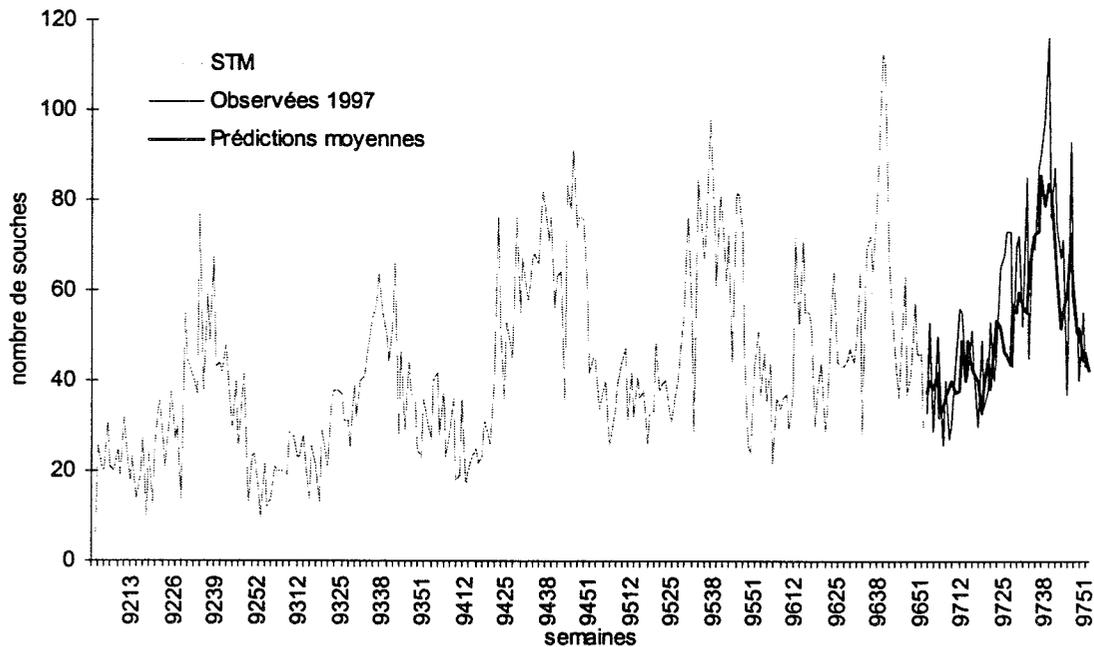
Figure 33 : Ajustement du modèle SARIMA (1.1.1)x(0.1.1)₅₂ à la série originale *S. Typhimurium* (STM).



4-2-2-4 Détermination du seuil d'alerte Box & Jenkins

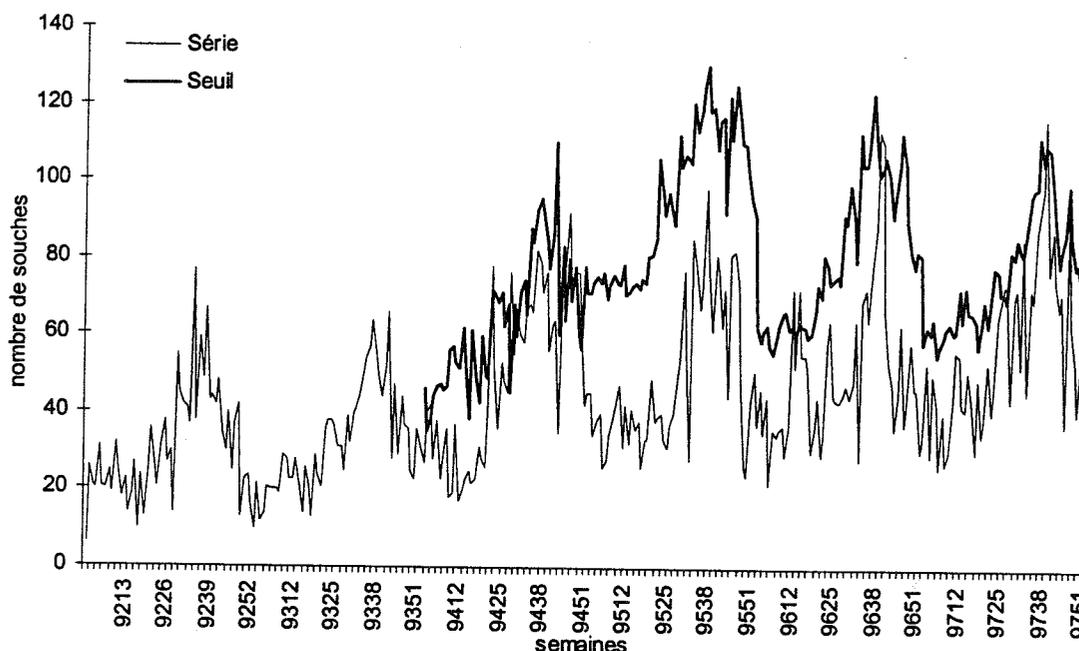
A partir du modèle SARIMA identifié sur la série observée (STM), Les valeurs ont été prédites pour l'année suivante 1997. La figure 34 montre l'ajustement des prédictions moyennes aux valeurs observées sur 1997.

Figure 34 : Comparaison de la série *S. Typhimurium* observée et des prédictions moyennes du modèle SARIMA pour l'année 1997.



Les valeurs prédites de chaque semaine $X_t(s)$ ont été calculées rétrospectivement pas à pas pour chaque année. A chaque pas, les prédictions sont basées sur les données disponibles des années antérieures. La figure 35 donne un aperçu visuel de la comparaison entre les données observées et les prédictions du modèle. La limite unilatérale à 95% représente le seuil d'alerte Box & Jenkins pour *S. Typhimurium* sur la période 1994-1997.

Figure 35 : Série originale et seuil Box & Jenkins à 95% pour *S. Typhimurium* sur la période 1994-1997.



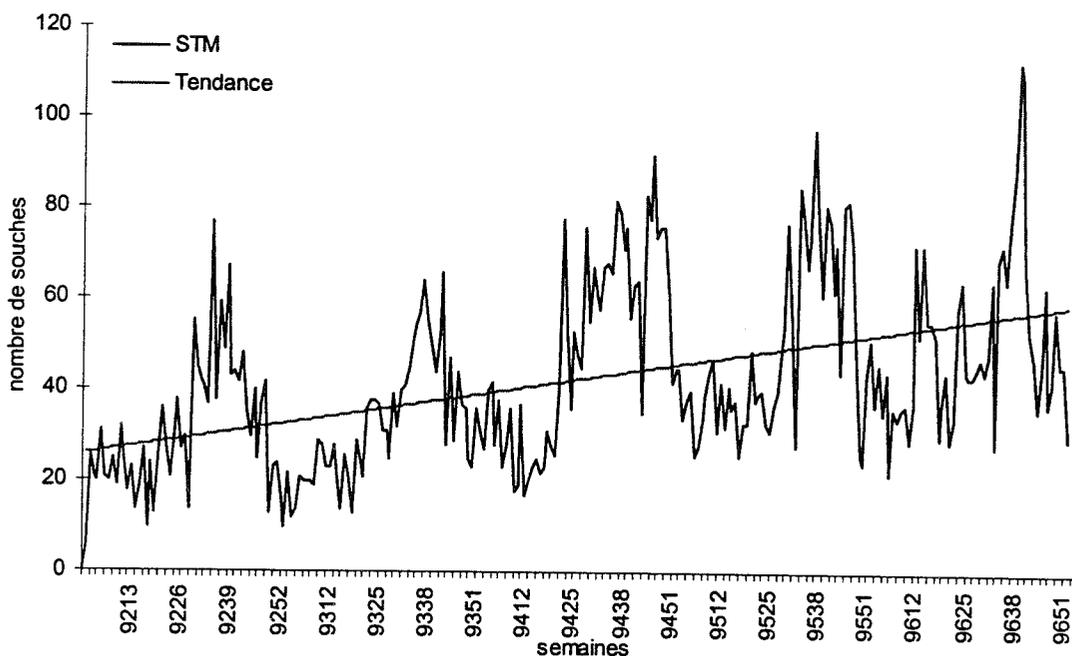
4-2-3 Modélisation spectrale pour *S. Typhimurium*

4-2-3-1 *Identification du modèle spectral*

La première étape est l'estimation de la tendance linéaire de la série initiale STM (figure 36). La droite de régression obtenue s'écrit : $Y_t = 25.93 + 0.126.t$

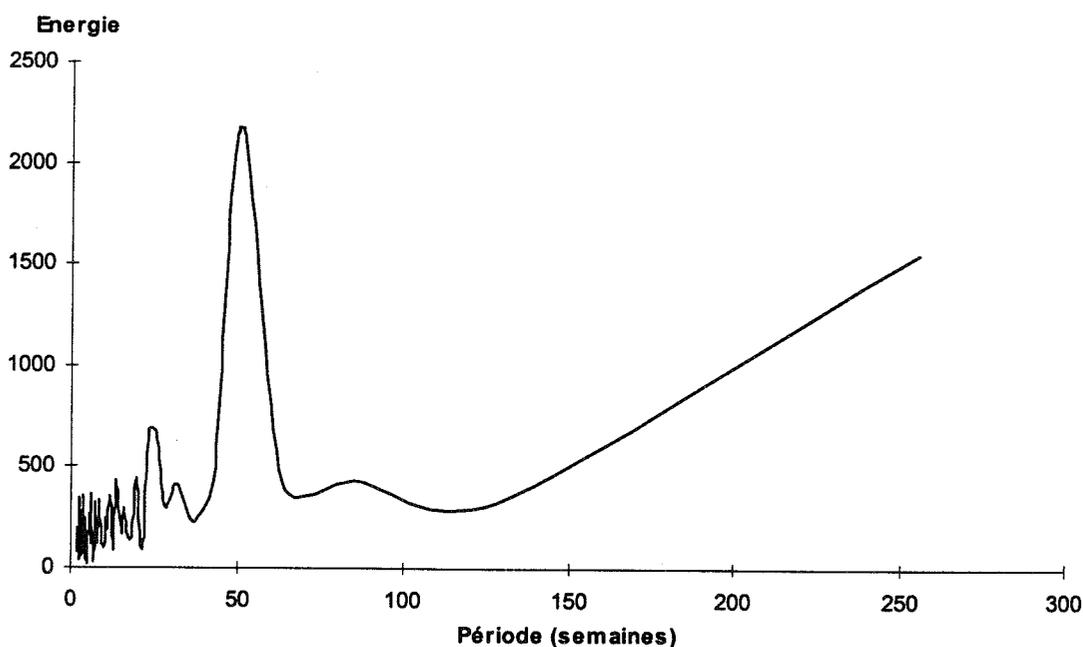
Le coefficient de régression R^2 est égal à 48.1%.

Figure 36 : Série originale *S. Typhimurium* (STM) et tendance linéaire estimée pour la période 1992-1996.



Le périodogramme de la série STM montre une périodicité de 52 semaines, une harmonique significative sur 26 semaines (figure 37).

Figure 37 : Périodogramme de la série STM.



Pour l'ajustement du modèle, nous avons retenu :

- les deux composantes cycliques 52 et 26, qui sont définies par leur amplitude (A1, A2) et leur phase (φ_1 , φ_2) respectives ;
- la tendance linéaire.

Le modèle s'écrit :

$$X_t = a + b.t + A_1.\cos(2\pi (t - \varphi_1)/52) + A_2.\cos(2\pi (t - \varphi_2)/26) + \varepsilon$$

La méthode des moindres carrés pour estimer les paramètres a donné les résultats regroupés dans le tableau XII.

Tableau XII : Paramètres de l'analyse spectrale pour la série *S. Typhimurium*

Tendance		Composantes cycliques Amplitude (A) / Phase (φ)	
a	b	52	26
28.71	0.11	-15.53 / 11.98	-5.71 / 0.03

Avec ces estimations de paramètres, 75.8% de la variance de la série initiale est prise en compte par le modèle.

L'étape d'optimisation des paramètres exclut les valeurs "épidémiques" ayant une probabilité de moins de 5% d'être observées sur l'ensemble de la série (valeur "épidémique" > moyenne + 1.64 déviation standardisée). La valeur seuil de détection a été calculée à 74.9. Les 24 valeurs "épidémiques" exclues ont été les valeurs des semaines :

- 34 de l'année 1992,
- 22, 28, 36, 37, 39, 44, 45, 46, 48 et 49 de l'année 1994,
- 30, 33, 34, 37, 38, 40, 41, 45 et 46 de l'année 1995,
- 37, 38, 39 et 40 de l'année 1996.

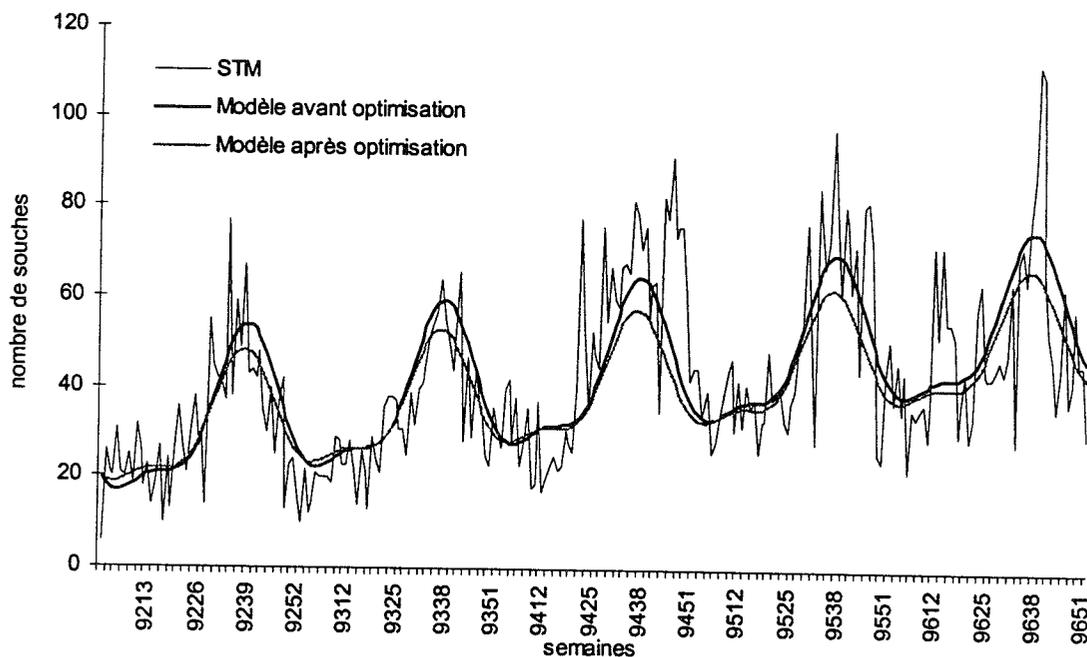
La méthode des moindres carrés appliquée à nouveau pour estimer les paramètres optimisés a donné les résultats regroupés dans le tableau XIII.

Tableau XIII : Paramètres optimisés pour l'analyse spectrale de la série *S. Typhimurium*.

Tendance		Composantes cycliques	
		Amplitude (A) / Phase (φ)	
a	b	52	26
27.91	0.09	-12.22 / 11.24	-4.80 / -0.96

Avec ces estimations de paramètres, 75.4% de la variance de la série initiale est prise en compte par le modèle final.

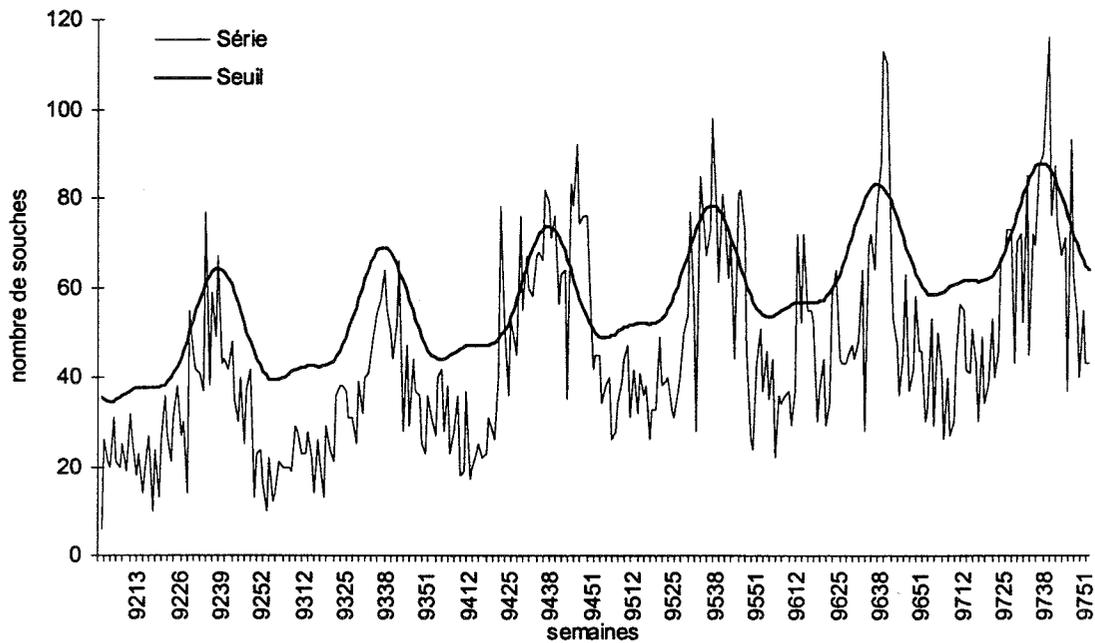
Figure 38 : Modèles estimés par analyse spectrale, avant et après retrait des points épidémiques de la série STM pour la période 1992-1996.



4-2-3-2 Détermination du seuil d'alerte Serfling

Le seuil d'alerte Serfling est déterminé par le calcul de la limite unilatérale à 95 % du modèle final obtenu après optimisation. La figure 39 donne le seuil rétrospectif sur les années 1992 à 1997.

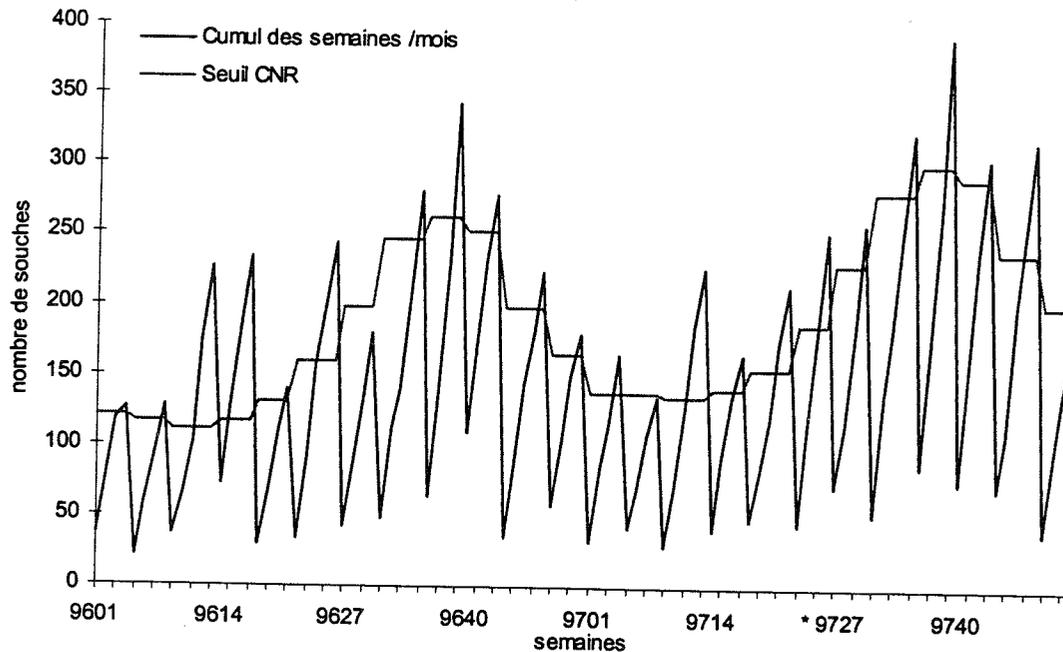
Figure 39 : Série originale *S. Typhimurium* et seuil Serfling à 95 % pour la période 1992-1997.



4-2-4 Détermination du seuil CNR par méthode des moyennes historiques

Le seuil CNR ($M5c + 20\%$) et les observations cumulées de *S. Typhimurium* ont été reconstruites pour la période de deux années à analyser 1996-1997 afin de les rendre comparables aux seuils hebdomadaires Box - Jenkins et Serfling (figure 40)

Figure 40 : Seuil CNR et nombre cumulé mensuel de souches *S. Typhimurium* par semaine pour la période 1996-1997.



4-2-5 Détection des semaines épidémiques par chaque seuil et détermination des critères de jugement

L'analyse des dépassements de seuil a porté sur les années 1996 et 1997 (annexe 4).

Le cumul des semaines réalisé pour chaque mois, dépasse le seuil CNR sur les semaines (figure 40) :

- 4, 8, 12 et 13, 15 à 17, 21, 25 et 26, 35, 39, 43, 48 et 52 de 1996 ;
- 4, 12 et 13, 17, 21 et 22, 26, 30, 35, 39, 43, 47 et 48 de 1997.

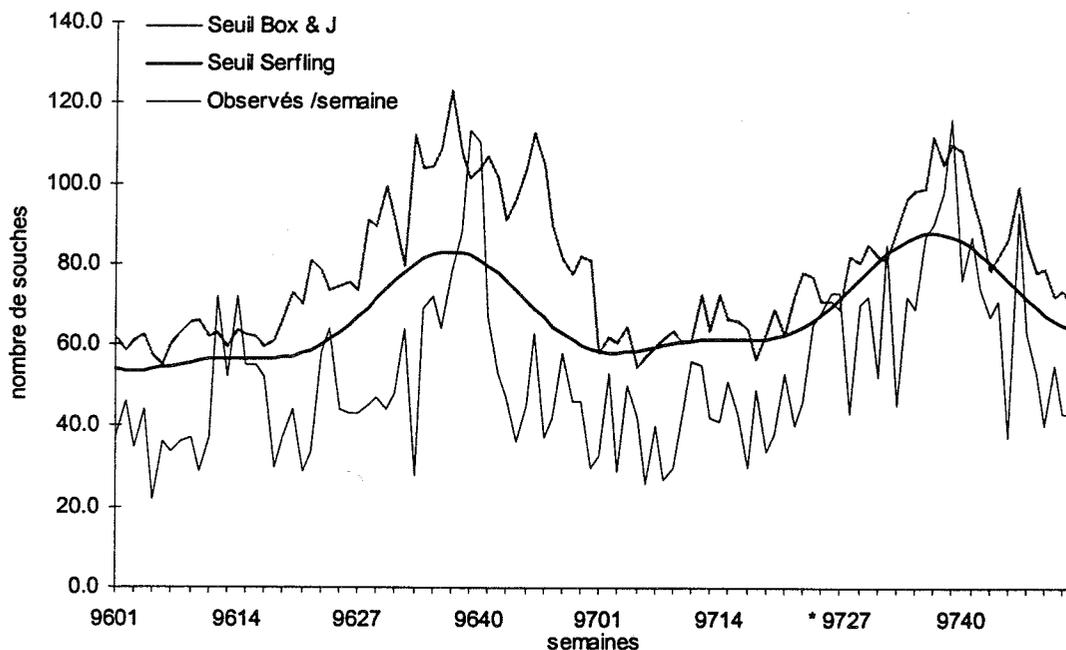
Le seuil Box & Jenkins oscille entre 55 et 113 (figure 41). Le seuil a été dépassé pour les semaines :

- 12, 14, 39 et 40 de 1996 ;
- 26 et 27, 32, 39 de 1997.

Le seuil Serfling oscille entre 54 et 88 (figure 41). Le seuil a été dépassé pour les semaines :

- 12, 14, 24, 38 à 40 de 1996 ;
- 25 à 27, 32, 37 à 39, 41, 46 de 1997.

Figure 41 : Seuil Box & Jenkins et seuil Serfling superposés au nombre hebdomadaire observé de souches *S. Typhimurium* pour la période 1996-1997.



Considérant les deux années qui couvrent la période janvier 1996 à décembre 1997 et conformément à notre référence qui définit la période de 8 semaines (semaines 97-20 à 97-27), comme période épidémique, les trois tableaux de contingence suivants ont été établis :

Pour le seuil CNR :

	Semaines "épidémiques"	Semaines "non épidémiques"	
Signal +	3	25	vpp = 0.11
Signal -	5	71	vpn = 0.93
	Se = 0.38	Sp = 0.74	Total = 104

Pour le seuil Box & Jenkins :

	Semaines "épidémiques"	Semaines "non épidémiques"	
Signal +	2	6	vpp = 0.25
Signal -	6	90	vpn = 0.94
	Se = 0.25	Sp = 0.94	Total = 104

Pour le seuil Serfling :

	Semaines "épidémiques"	Semaines "non épidémiques"	
Signal +	3	12	vpp = 0.20
Signal -	5	84	vpn = 0.94
	Se = 0.38	Sp = 0.88	Total = 104

Sur la période épidémique définie en 1997, semaines 97-20 à 97-27, le seuil CNR déclenche l'alerte sur les semaines 97-21 et 97-22, puis sur la semaine 97-26. Le seuil Box & Jenkins déclenche l'alerte sur les semaines 97-26 et 97-27. Le seuil Serfling déclenche l'alerte sur les semaines successives 97-25, 97-26 et 97-27.

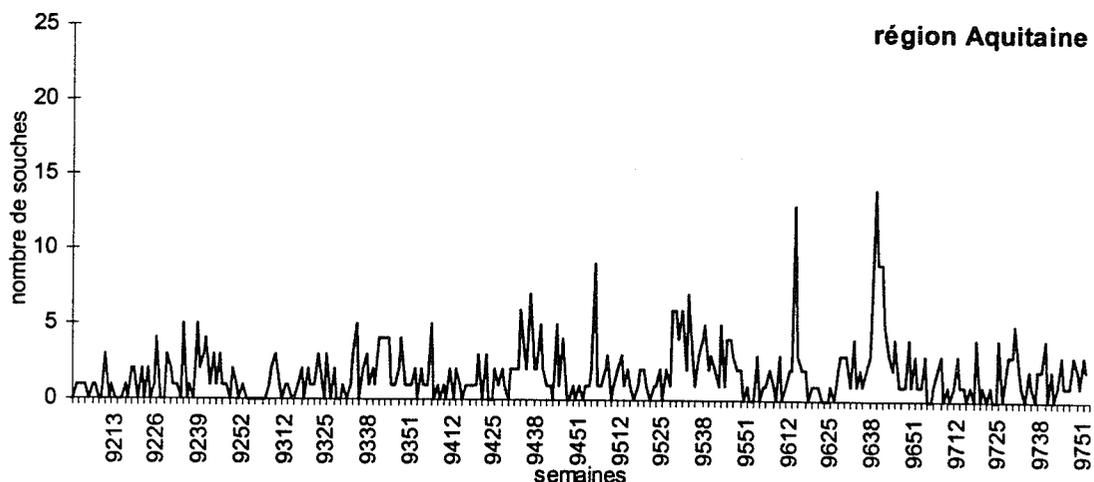
4-2-6 Recherche de critères complémentaires pour *S. Typhimurium*

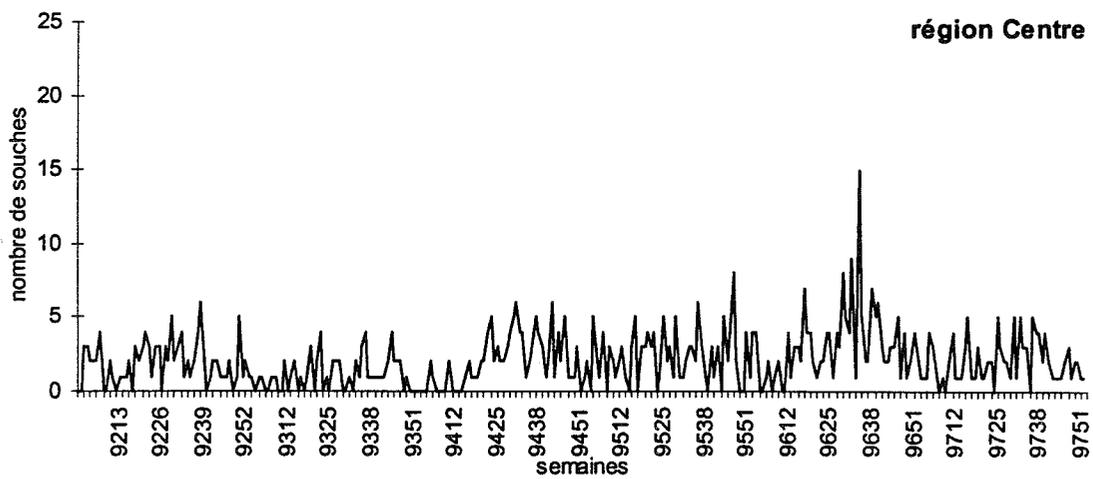
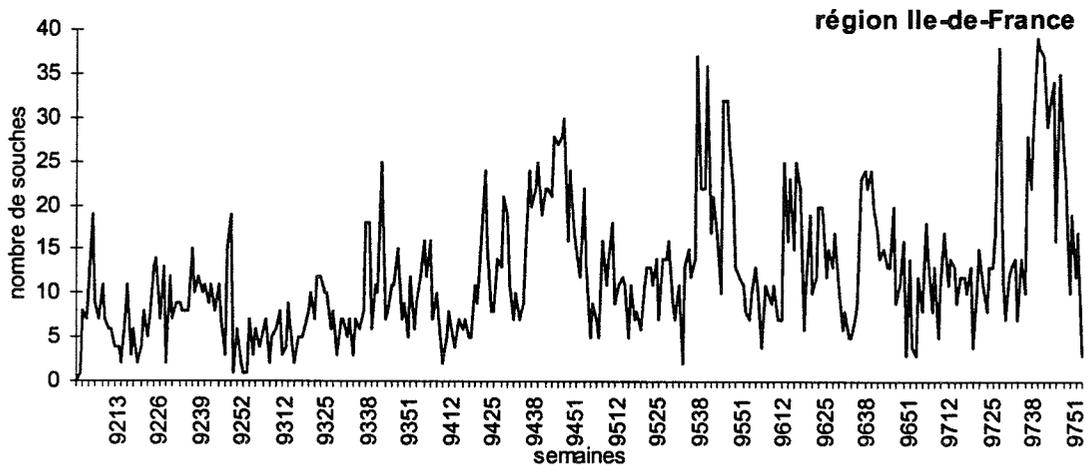
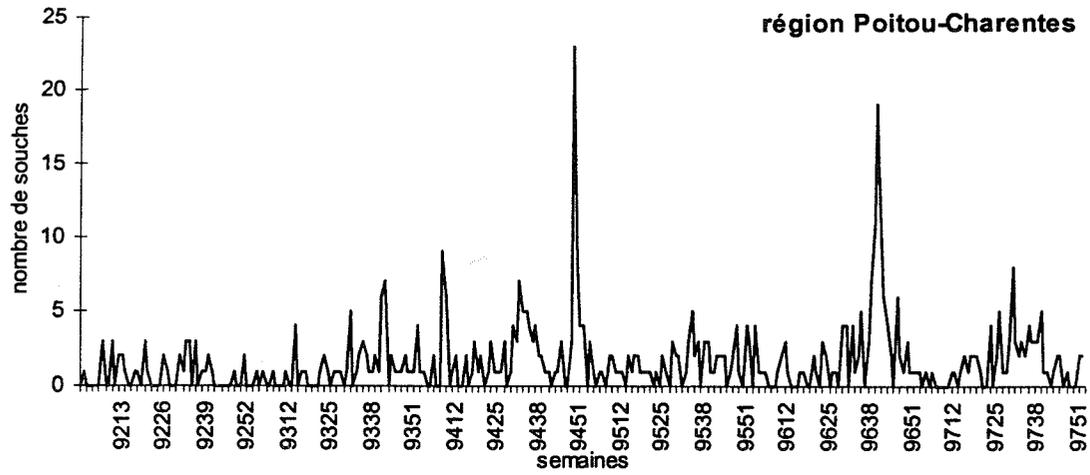
Pour les deux années dont le seuil a été analysé, plusieurs pics apparaissent suspects, en particulier sur les périodes où les seuils sont dépassés sur deux semaines successives. Il s'agit des périodes hebdomadaires 96-11 à 96-15, 96-38 à 96-41 et 97-37 à 97-40. Nous avons analysé en détail ces dépassements de seuil pour essayer de retrouver des caractéristiques susceptibles de confirmer le caractère épidémique ou non épidémique de ces observations.

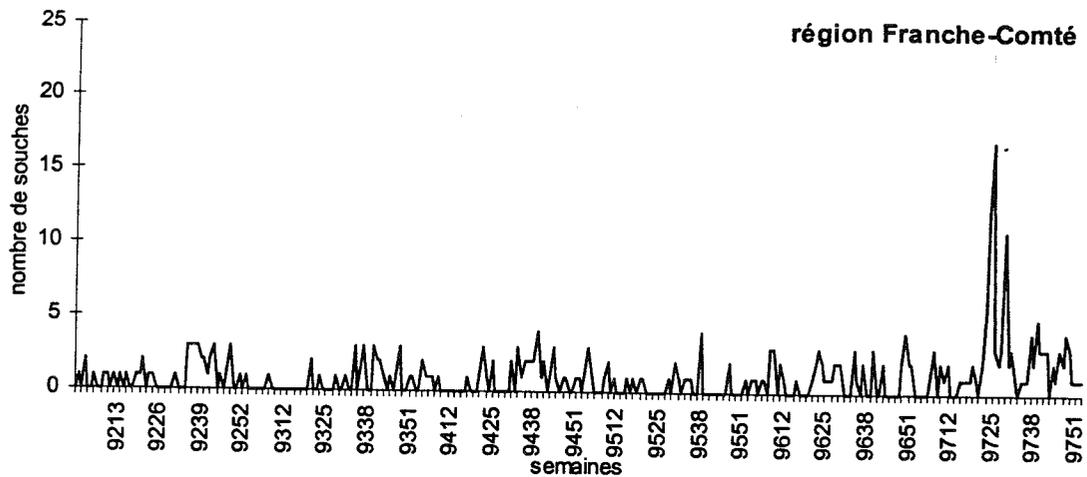
4-2-6-1 Analyse par zones géographiques

Sur les périodes correspondant aux dépassements de seuils les plus fréquents (96-12 à 96-17, 96-38 à 96-40, 97-37 à 97-39 et la période épidémique 97-20 à 97-27), l'analyse régionale par régions administratives a montré la prédominance des régions Aquitaine, Poitou-Charentes, Paris-Ile-de-France, Centre et Franche-Comté. Les séries hebdomadaires de la figure 42 montrent les pics pour chaque région et chaque période, évocateurs de concentration géographique de cas.

Figure 42 : Séries hebdomadaires de souches *S. Typhimurium* reçues au CNR pour les régions Aquitaine, Poitou-Charentes, Ile-de-France, Centre et Franche-Comté en 1992-1997.

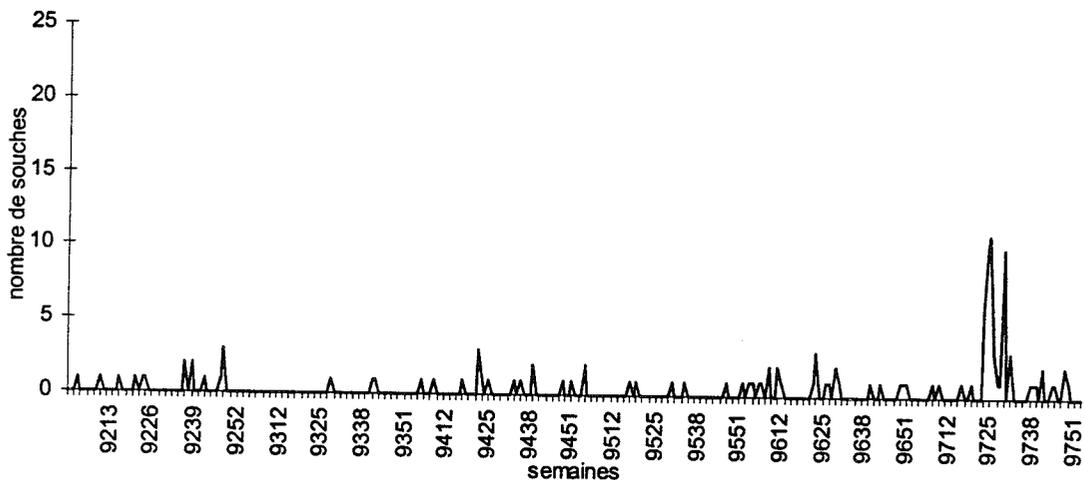






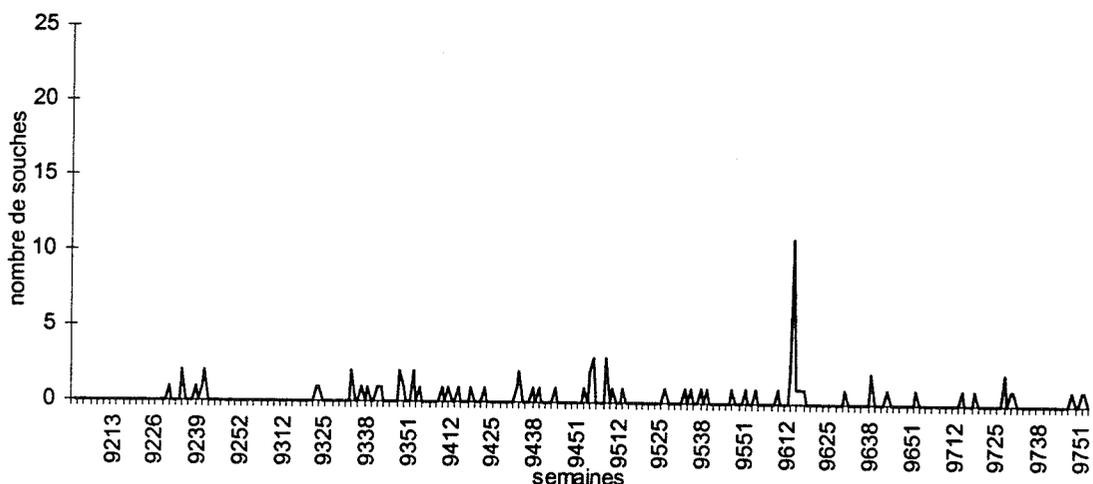
La région Franche-Comté fait apparaître le pic épidémique de 1997 correspondant à l'épidémie du département du Jura (figure 43).

Figure 43 : Série hebdomadaire de souches *S. Typhimurium* reçues au CNR pour le département du Jura (dep 39) en 1992-1997.



Pour les autres périodes, des concentrations départementales correspondants aux régions sus-citées ont été retrouvées comme dans le département du Lot-et-Garonne (dep 47) (figure 44).

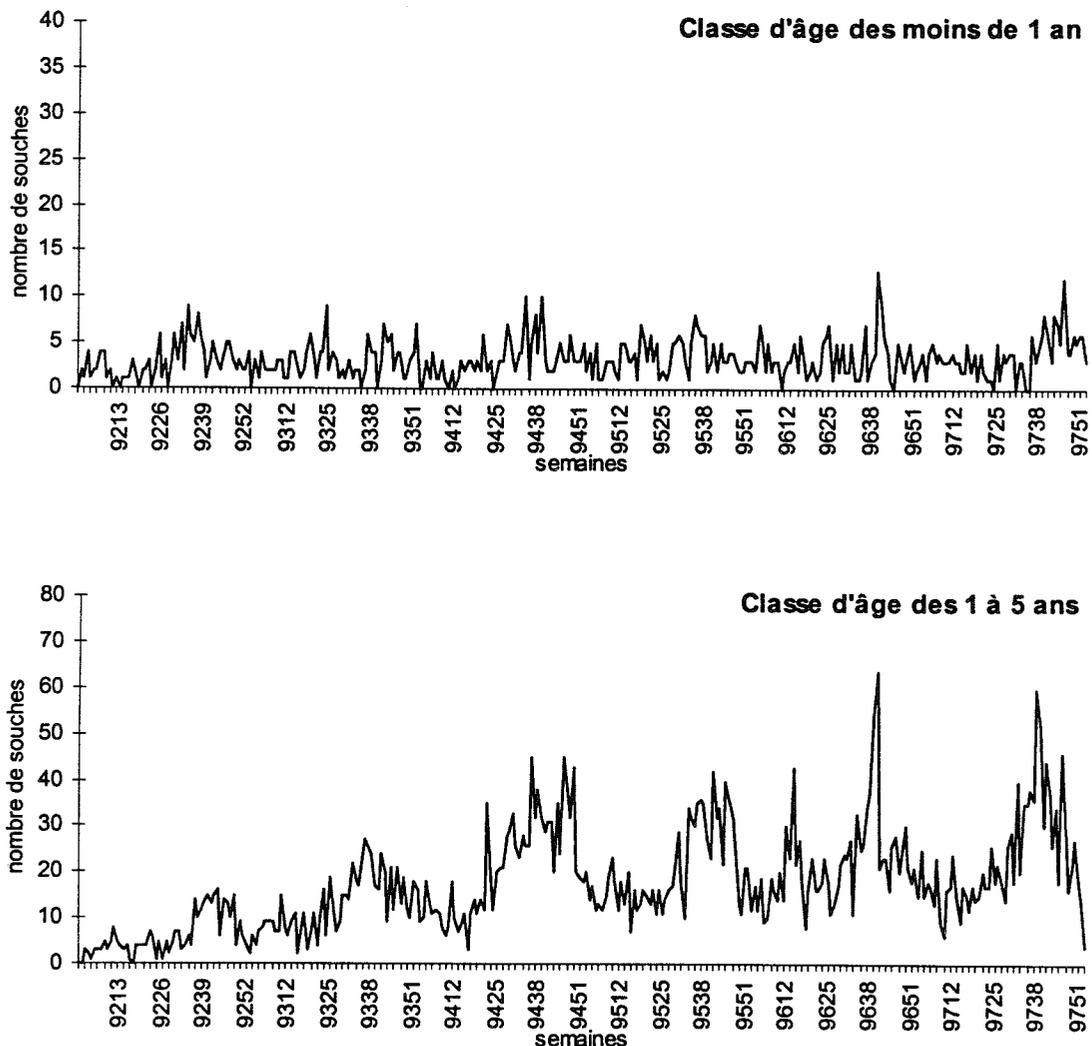
Figure 44 : Série hebdomadaire de souches *S. Typhimurium* reçues au CNR pour le département du Lot-et-Garonne (47) en 1992-1997.

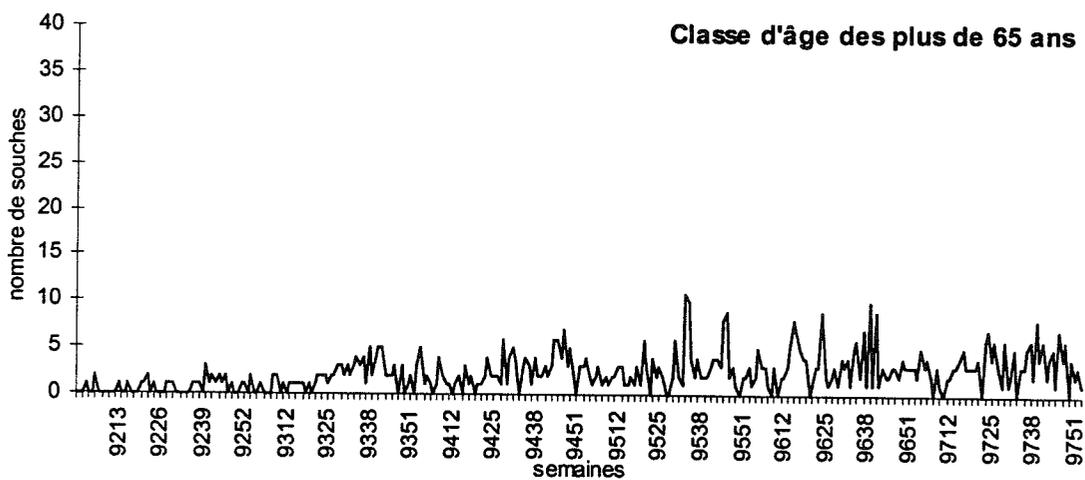
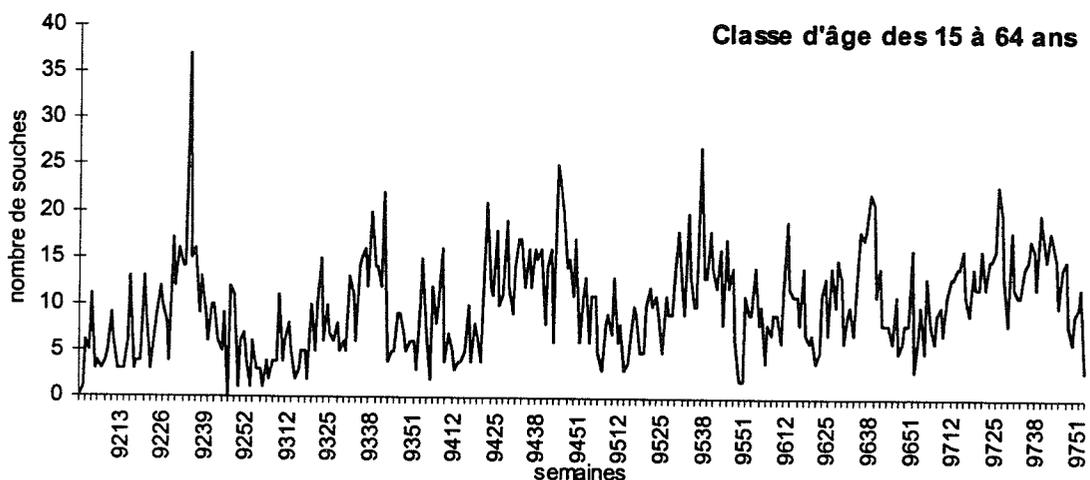
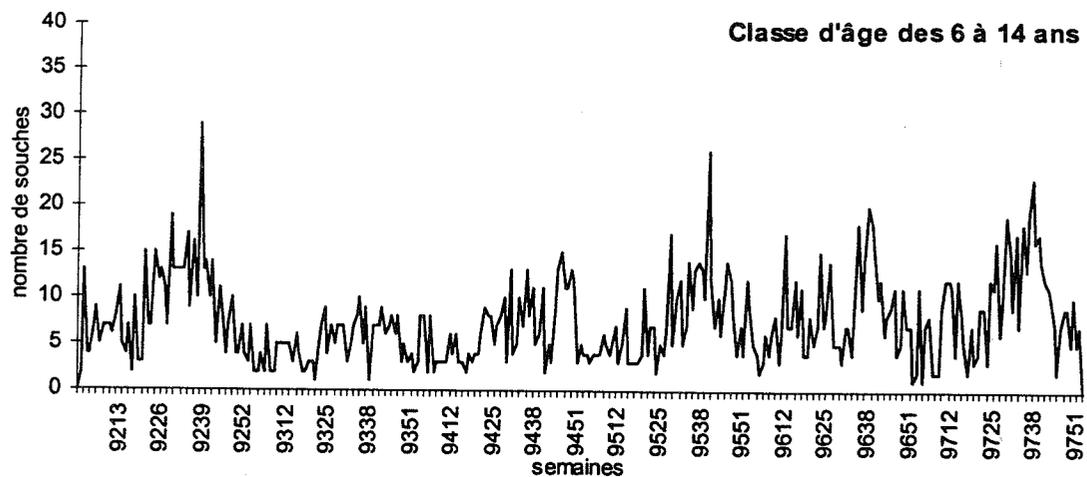


4-2-6-2 Age des cas

Les cas de salmonelloses analysés à un niveau national par classe d'âge ont donné les séries réunies sur la figure 45. Le nombre habituel de souches de salmonelles reçues au CNR est globalement plus important que pour d'autres sérotypes car *S. Typhimurium* est un sérotype prédominant. Ceci rend plus difficile le repérage de concentrations excessives de souches. Néanmoins par tranche d'âge, certains pics se détachent plus nettement du profil général de chaque série, par exemple les pics des périodes 96-12 à 96-17 et 96-37 à 96-41, en particulier dans la classe d'âge des 1 à 5 ans.

Figure 45 : Séries hebdomadaires de souches *S. Typhimurium* reçues au CNR par classes d'âge pour la période 1992-1997.





4-3 Généralisation des seuils pour les autres sérotypes

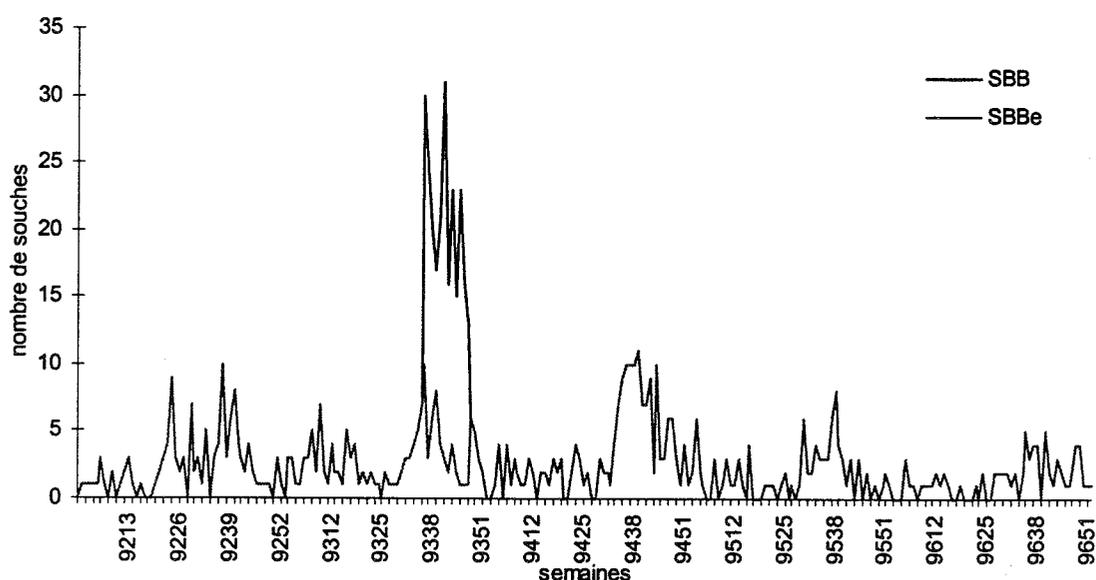
Selon les mêmes procédures, nous avons élaboré et comparé trois seuils pour chacun des sérotypes suivants : *S. Brandenburg*, *S. Goldcoast*, *S. Hadar*.

4-3-1 Série *S. Brandenburg* (SBB)

L'épidémie à *S. Brandenburg* a eu lieu de la semaine 94-23 à 94-31 (figure 47). Fin juin le CNR signalait une augmentation anormale d'isollements de *S. Brandenburg* dans l'Ouest de

la France. Il est décidé de ne pas mener d'enquête mais de renforcer la vigilance. Le 3 août l'analyse de la situation montrait que le nombre de souches reçues au CNR au cours des mois de juin et juillet était très supérieur au seuil d'alerte mensuel. Le RNSP a coordonné les investigations épidémiologiques à partir du 04/08/94. Le phénomène concernait principalement la région Pays-de-Loire en particulier le département de Loire-Atlantique (dep 44) et dans une moindre mesure les départements Maine-et-Loire (dep 49), Vendée (dep 85) et Morbihan (dep 56). Les enfants étaient les plus nombreux parmi les cas recensés. L'ensemble des investigations dont l'enquête cas-témoins a mis en évidence un rôti de porc saumuré produit dans une unité unique et le jambon blanc acheté à la coupe comme véhicule de l'épidémie (38).

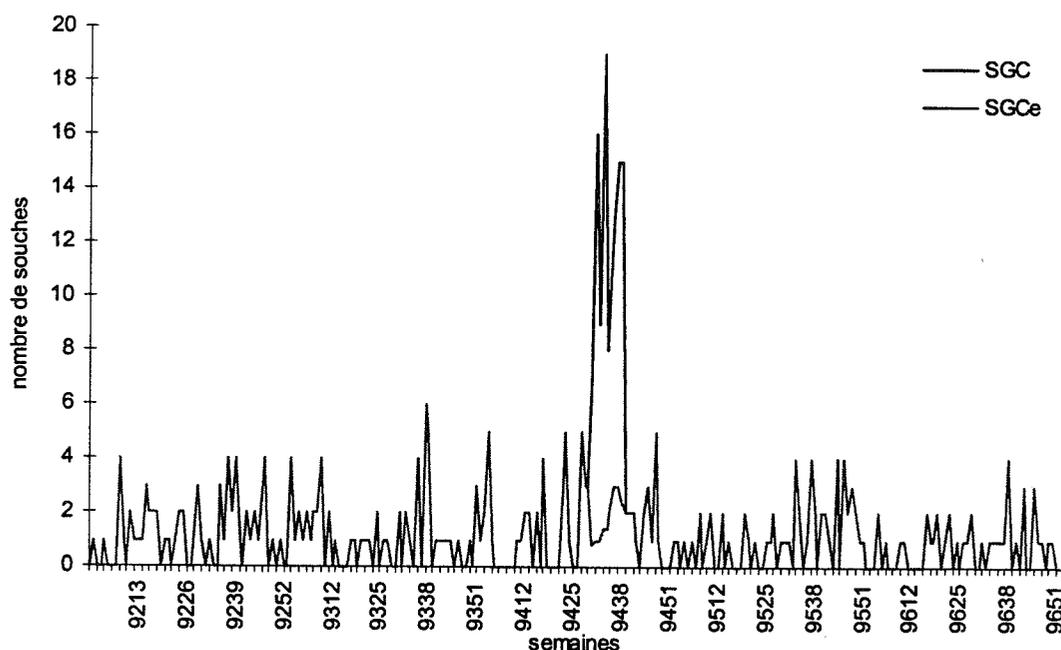
Figure 47 : Superposition de la série originale *S. Brandenburg* (SBB) et de la série échantillonnée (SBB_e) sur la période 1992-1996.



4-3-2 Série *S. Goldcoast* (SGC)

L'épidémie à *S. Goldcoast* a eu lieu de la semaine 94-26 à 94-39 (figure 48). Le 12/08/94 le CNR informait la DGS d'une augmentation du nombre de souches de *Salmonella* sérotype Goldcoast en juillet et août 1994. La plupart des souches provenaient du même laboratoire (Romans-sur-Isère, Drôme) et continuaient à être isolées, principalement chez des enfants. L'enquête cas-témoins a été réalisée le 17 août sous l'égide du RNSP. La distribution géographique indiquait une tendance à la concentration des cas dans les départements voisins de la Drôme, la majorité habitant à Romans-sur-Isère ou dans les environs immédiats (dep 40) (77%). Des cas ont été observés dans toutes les classes d'âge avec une incidence plus importante avant 5 ans et après 65 ans. Le risque d'infection à *S. Goldcoast* était significativement plus élevé chez les clients d'un supermarché, également plus important chez les consommateurs de jambon cuit acheté à la coupe (39).

Figure 48 : Superposition de la série originale *S. Goldcoast* (SGC) et de la série écrêtée (SGCe) sur la période 1992-1996.

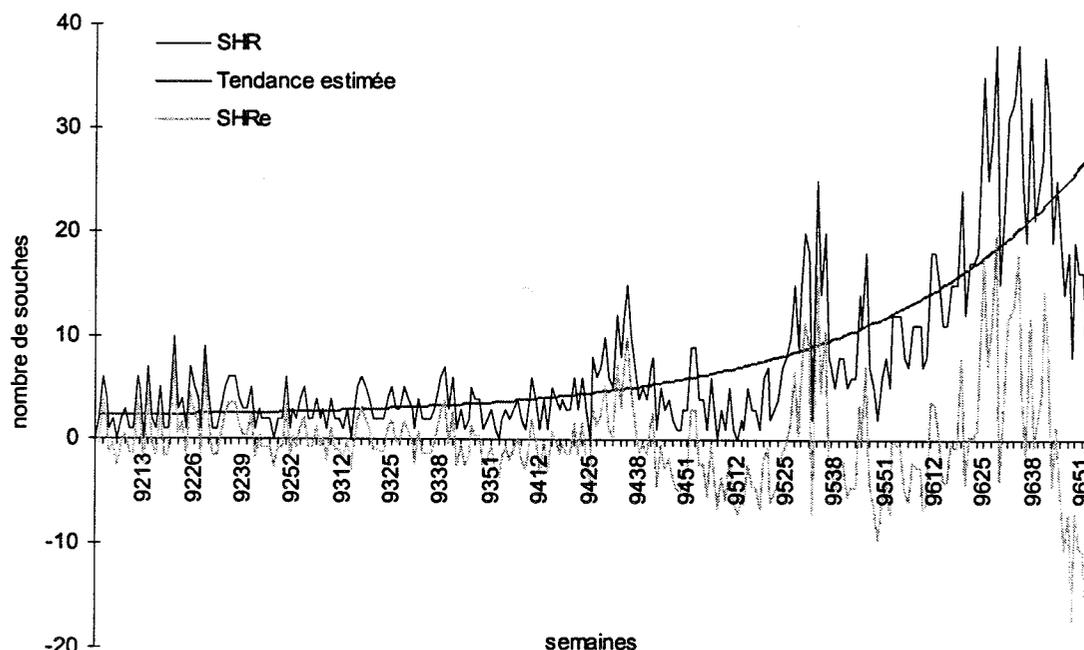


4-3-3 Série *S. Hadar* (SHR)

L'épidémie à *S. Hadar* investiguée a eu lieu de la semaine 95-23 à 95-37 sur l'ensemble de la France (figure 49). Le 04/08/95 le CNR informait la DGS d'un doublement du nombre des isolements humains de *Salmonella* Hadar au cours des mois de juin et juillet 1995 par rapport au nombre d'isolements réalisés pour les mêmes mois de l'année précédente. Le 07/08/95, le RNSP est chargé par la DGS de réaliser une enquête épidémiologique. Sur 1992-1993, 81% des souches de *S. Hadar* répertoriées par le CNEVA chez les animaux ou dans les denrées alimentaires, concernaient des volailles ou des denrées alimentaires à base de volailles. L'enquête cas-témoins réalisée à partir du 23/08/95 sur l'ensemble de la période épidémique, n'a pas mis en évidence d'aliment significativement plus consommé par les cas que par les témoins. Mais l'odds-ratio apparié concernant la consommation de poulet rosé était de 3,5 (IC 95% : 0,7-34,5 ; $p = 0,08$) (40).

La série originale (SHR) présente une tendance croissante exponentielle de type $(a + c \cdot \exp(b \cdot t))$. Cette tendance a été évaluée et soustraite de la série originale. C'est sur la série sans tendance (SHRe) que la procédure Box et Jenkins a été appliquée.

Figure 49 : Comparaison de la série originale *S. Hadar* (SHR) et de la série originale après retrait de la tendance (SHRe) sur la période 1992-1996.



4-3-4 Identification des modèles et construction des seuils

Aucun modèle satisfaisant n'a pu être identifié selon le processus Box et Jenkins. Les fonctions d'autocorrélation des séries SBB_e, SGC_e et SHR_e n'ont révélées aucune corrélation significative. Les séries apparaissent donc complètement aléatoires dans ce processus d'analyse. Par conséquent, aucun seuil n'a pu être construit par la méthode Box et Jenkins.

A l'opposé, la méthode Serfling a permis de repérer des cycles saisonniers sur le périodogramme, d'en déduire des modèles et de déterminer des seuils. Les seuils CNR ont été reconstruits sur la période de deux années 1994-1995 sous la forme de seuils hebdomadaires pour les rendre comparables au seuils Serfling. Les figures 50 à 52 présentent les seuils Serfling et CNR obtenus pour chaque sérotype.

4-3-4-1 Seuils pour *S. Brandenburg* (annexe 5)

Le modèle obtenu par analyse spectrale pour la série *S. Brandenburg* (SBB) prend en compte 42% de la variance et s'écrit :

$$X_t = 2.8 - 1.2 \cos(2\pi (t - 9.2)/52) + 0.2 \cos(2\pi (t - 0.3)/32) + 0.4 \cos(2\pi (t - 1.9)/16) + \varepsilon$$

Figure 50 : Périodogramme de la série *S. Brandenburg* (SBB) pour la période 1992-1996.

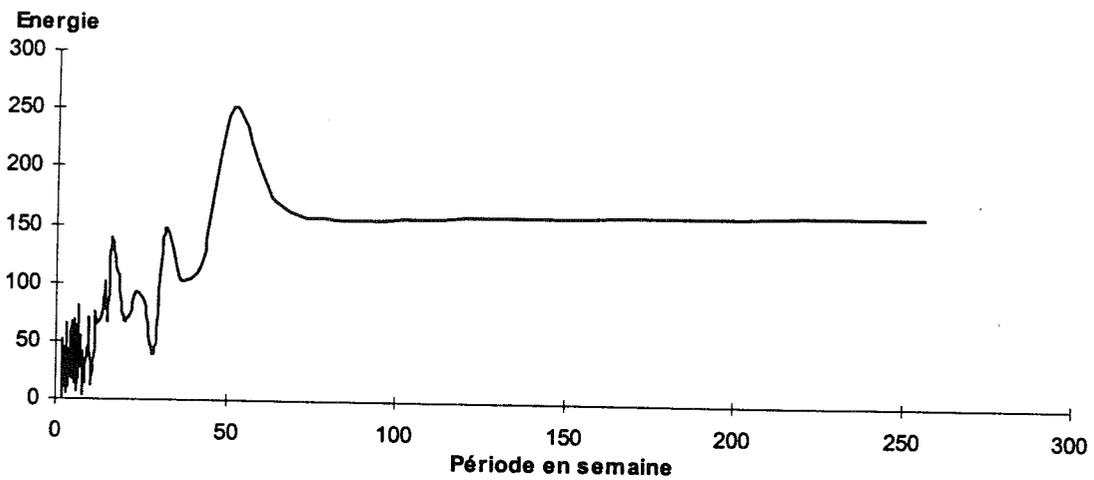


Figure 51 : Seuil Serfling à 95% pour la série *S. Brandenburg* sur la période 1994-1995.

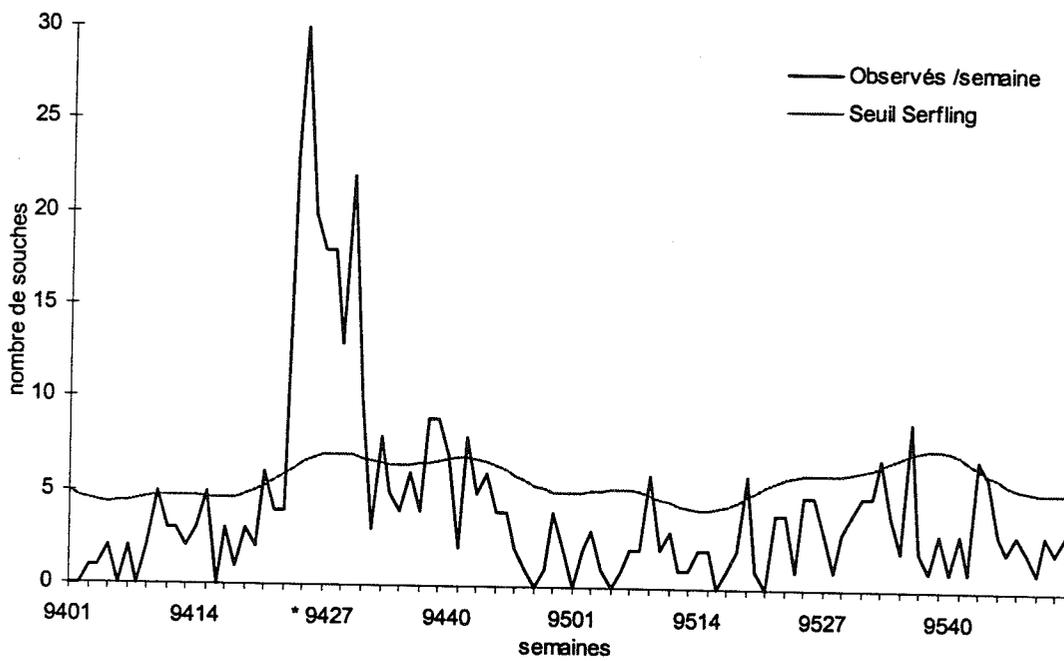
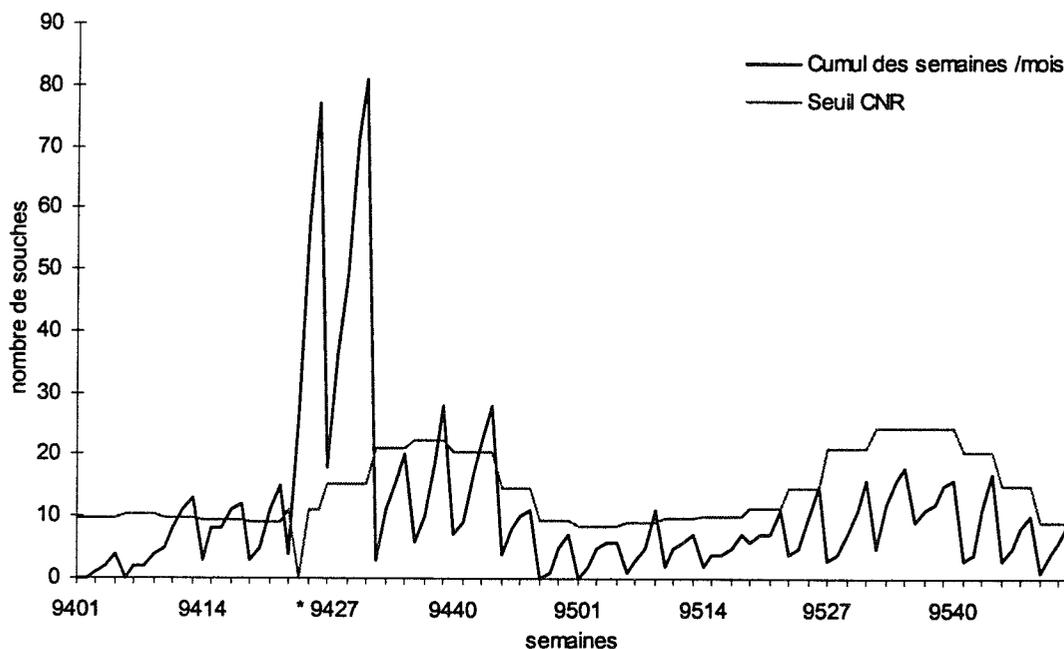


Figure 52 : Seuil CNR pour *S. Brandenburg* sur la période 1994-1995.



4-3-4-2 Seuils pour *S. Goldcoast* (annexe 6)

Le modèle obtenu par analyse spectrale pour la série *S. Goldcoast* (SGC) prend en compte 33% de la variance et s'écrit :

$$X_t = 1.36 + 0.37 \cos(2\pi (t + 12.7)/52) + 0.15 \cos(2\pi (t - 14.86)/26) - 0.2 \cos(2\pi (t - 1.5)/16) + \varepsilon$$

Figure 53 : Périodogramme de la série *S. Goldcoast* (SGC) pour la période 1992-1996.

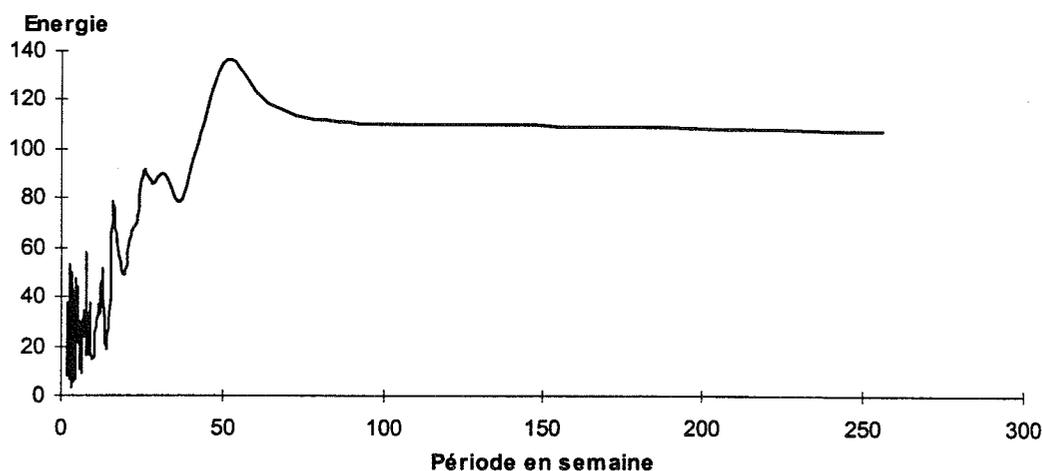


Figure 54 : Seuil Serfling à 95% pour la série *S. Goldcoast* sur la période 1994-1995.

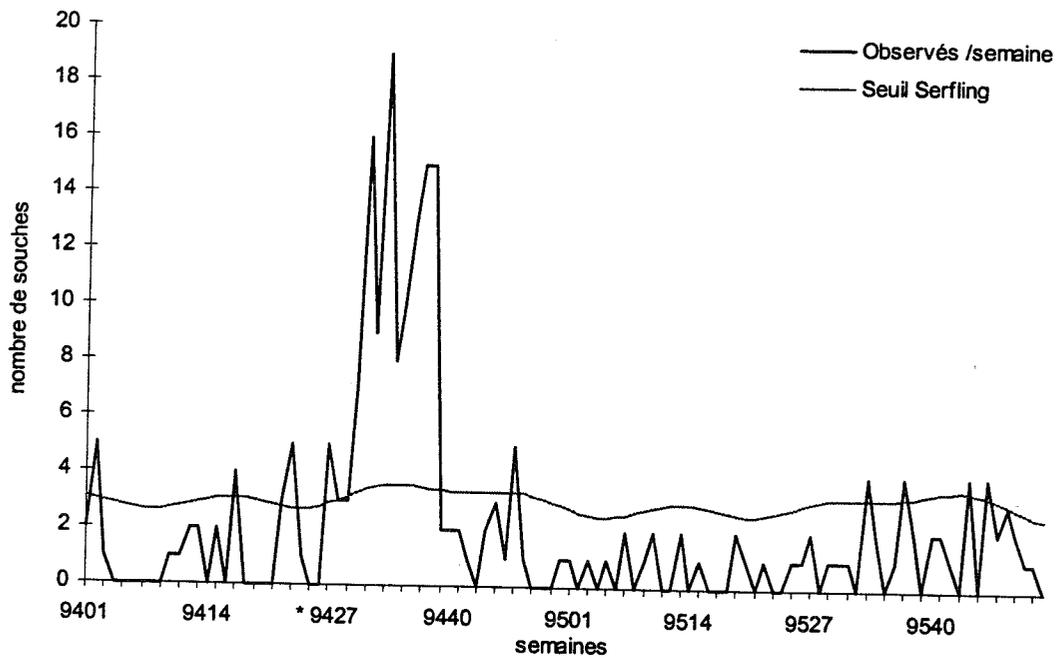
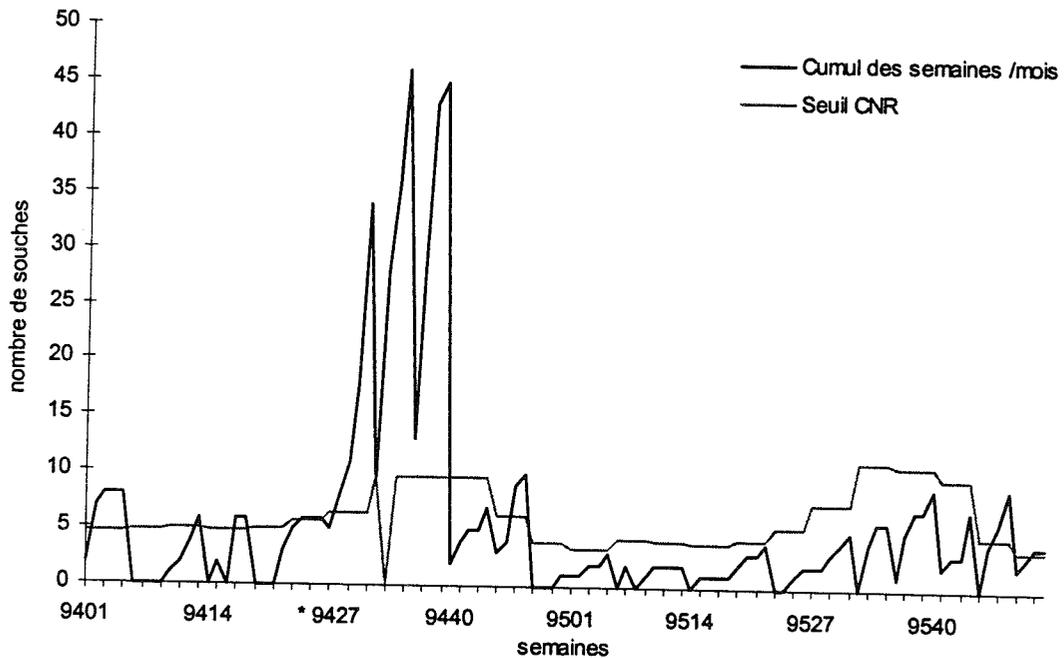


Figure 55 : Seuil CNR pour *S. Goldcoast* sur la période 1994-1995.



4-3-4-3 Seuils pour *S. Hadar* (annexe 7)

Le modèle obtenu par analyse spectrale pour la série *S. Hadar* (SHR) prend en compte 83.4% de la variance et s'écrit :

$$X_t = 2.44 + 0.26 \exp(0.02 t) - 1.68 \cos(2\pi (t - 5.18)/52) + 0.69 \cos(2\pi (t + 5.34)/28) + \varepsilon$$

Figure 56 : Périodogramme de la série *S. Hadar* (SHR) pour la période 1992-1996.

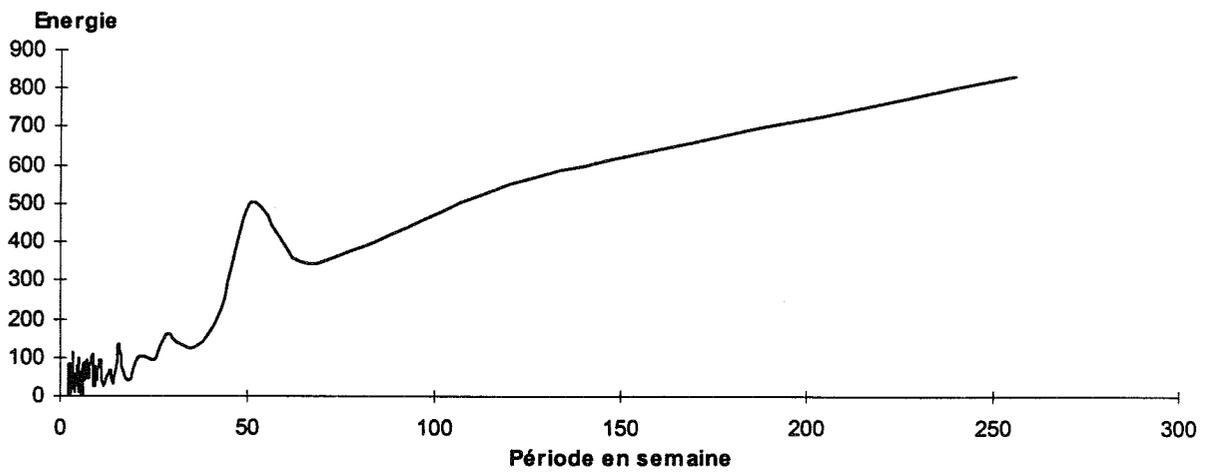


Figure 57 : Seuil Serfling à 95% pour la série *S. Hadar* sur la période 1994-1995.

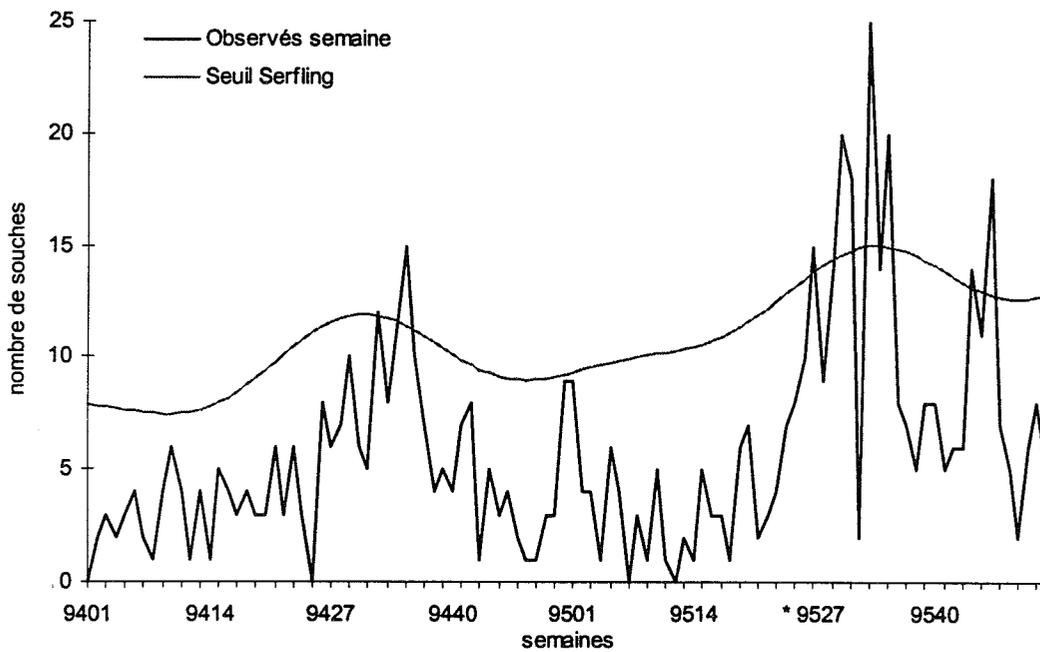
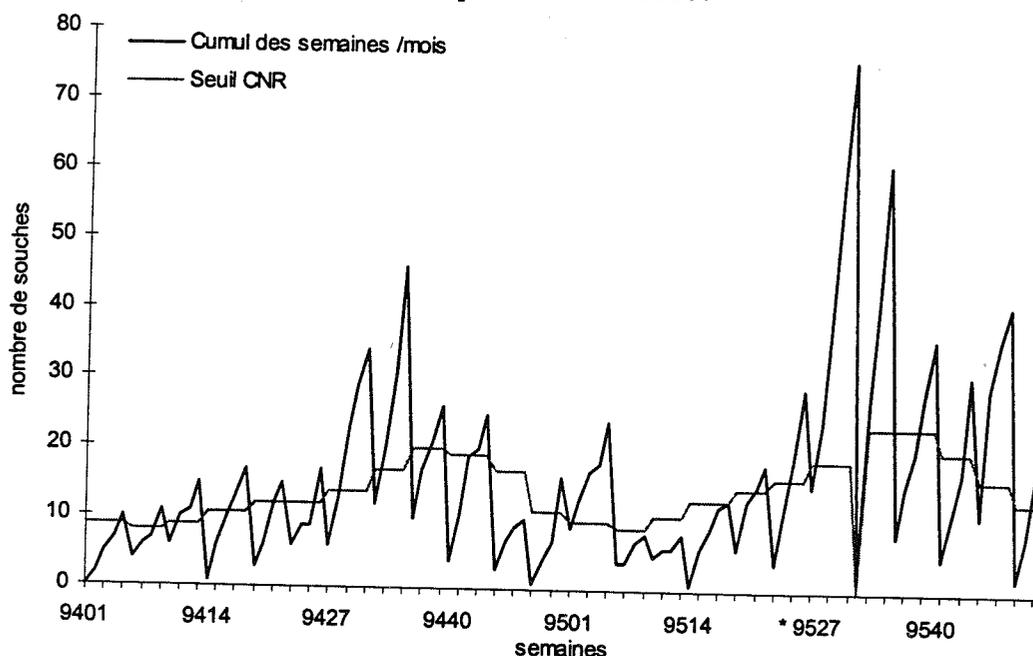


Figure 58 : Seuil CNR pour *S. Hadar* sur la période 1994-1995.



Les critères de validité ont été évalués sur les dépassements de seuils relativement aux épidémies de la période 1994 - 1995. A l'identique des évaluations de seuils pour *S. Paratyphi B* et *S. Typhimurium*, on a considéré que l'alerte est déclenchée dès la première semaine de dépassement du seuil. Le tableau XIV résume les résultats obtenus pour *S. Brandenburg*, *S. Goldcoast* et *S. Hadar*.

Tableau XIV : Résultats des critères de jugement pour les 3 sérotypes *S. Brandenburg*, *S. Goldcoast* et *S. Hadar*.

Sérotypes	Seuil CNR			Seuil Box & Jenkins			Seuil Serfling		
	Se	Sp	VPP	Se	Sp	VPP	Se	Sp	VPP
Brandenburg	0.89	0.88	0.42	/	/	/	0.89	0.86	0.38
Goldcoast	0.86	0.83	0.44	/	/	/	0.71	0.90	0.53
Hadar	0.60	0.61	0.21	/	/	/	0.33	0.96	0.56

4-4 Evaluation des critères de jugement en fonction de la référence "corrigée"

Les résultats des analyses complémentaires pour les séries *S. Paratyphi B* et *S. Typhimurium* ont fait apparaître, sur les périodes analysées, des arguments en faveur d'épisodes épidémiques non détectés. Des regroupements de cas ont été suggérés sur des critères géographiques ou de classes d'âge, sous la forme de pics "épidémiques" contrastant avec le reste de chaque série.

Pour *S. Paratyphi B*, une concentration géographique de cas a été repérée en 1994 dans trois régions, dont deux sont limitrophes (Poitou-Charentes, Pays-de-la-Loire, Ile-de-France). Cette concentration a été constatée sur plusieurs semaines successives (semaines 94-33 à 94-48) et prédominante pour la classe d'âge des 1 à 5 ans.

Pour *S. Typhimurium*, trois pics inhabituels ont été repérés. Les trois regroupements de cas ont été retrouvés :

- dans deux régions (Ile-de-France, Aquitaine) sur une période hivernale identique (semaines 96-12 à 96-17), concernant toutes les classes d'âge mais prédominant pour les 1 à 5 ans ;
- dans deux régions limitrophes (Aquitaine, Poitou-Charentes) sur une période identique (semaines 96-37 à 96-41) et concernant plus particulièrement les moins de 5 ans ;
- dans une région (Ile-de-France) sur la période des semaines 97-35 à 97-43, concernant deux classes d'âge les 1 à 5 ans et les 6 à 14 ans.

Les données des séries *S. Brandenburg* et *S. Goldcoast* ne suggèrent pas l'existence d'épidémies non détectées.

L'ensemble des observations est en faveur de plusieurs épidémies passées inaperçues et qui n'ont pas été prises en compte dans la référence initiale (référence "brute"). A partir de ces arguments descriptifs et subjectifs, les périodes 94-33 à 94-48 pour *S. Paratyphi B*, 96-12 à 96-17, 96-37 à 96-41 et 97-35 à 97-43 pour *S. Typhimurium* ont été considérées comme des épidémies vraies.

Sur la base de ces conclusions, une référence "corrigée" a été définie en semaines épidémiques vraies versus non épidémiques (référence "corrigée" = référence "brute" + semaines épidémiques définies a posteriori) (cf. 3.5).

En fonction de la référence "corrigée", les critères de sensibilité (Se) et de valeur prédictive positive (vpp) ont été réévalués pour chaque sérotype et pour les quatre sérotypes groupés. De plus l'évaluation a été réalisée pour les références "brute et "corrigée" en faisant varier le délai d'alerte de 1 à 3 semaines. Les résultats sont regroupés dans les tableaux XV et XVI.

Tableau XV : Sensibilité (Se) et valeur prédictive positive (vpp) estimées pour les seuils CNR, Box & Jenkins et Serfling en fonction de la référence "brute".

Sérotypes	Délai d'alerte	Seuil CNR		Seuil Box & J		Seuil Serfling	
		Se	vpp	Se	vpp	Se	vpp
Paratyphi B	1 sem	0.80	0.67	0.93	1.00	0.80	0.52
	2 sem	0.67	0.77	0.87	1.00	0.73	0.58
	3 sem	0.53	0.80	0.80	1.00	0.67	0.63
Brandenburg	1 sem	0.89	0.42	/	/	0.89	0.38
	2 sem	0.78	0.64	/	/	0.78	0.78
	3 sem	0.67	1.00	/	/	0.67	0.86
Goldcoast	1 sem	0.86	0.44	/	/	0.71	0.53
	2 sem	0.71	0.56	/	/	0.57	0.89
	3 sem	0.57	0.80	/	/	0.50	1.00
Typhimurium	1 sem	0.38	0.11	0.25	0.25	0.38	0.20
	2 sem	0.13	0.14	0.13	0.50	0.25	0.33
	3 sem	0	0	0	/	0.13	0.33
Groupés	1 sem	0.76	0.38	0.70	0.73	0.72	0.42
	2 sem	0.61	0.57	0.61	0.93	0.61	0.65
	3 sem	0.48	0.81	0.52	1.00	0.52	0.73

Tableau XVI : Sensibilité (Se) et valeur prédictive positive (vpp) estimées pour les seuils CNR, Box & Jenkins et Serfling en fonction de la référence "corrigée".

Sérotypes	Délai d'alerte	Seuil CNR		Seuil Box & J		Seuil Serfling	
		Se	vpp	Se	vpp	Se	vpp
Paratyphi B	1 sem	0.42	0.77	0.45	1.00	0.71	0.88
	2 sem	0.26	0.73	0.42	1.00	0.56	0.89
	3 sem	0.16	0.71	0.39	1.00	0.45	0.88
Typhimurium	1 sem	0.43	0.43	0.25	0.88	0.41	0.86
	2 sem	0.14	0.57	0.07	1.00	0.21	1.00
	3 sem	0.04	1.00	0	/	0.11	1.00
Groupés	1 sem	0.55	0.50	0.36	0.96	0.63	0.66
	2 sem	0.35	0.62	0.25	1.00	0.46	0.88
	3 sem	0.24	0.83	0.20	1.00	0.37	0.91

5 - DISCUSSION

5-1 Résultats principaux

Des seuils d'alerte hebdomadaires ont été élaborés rétrospectivement sur la période 1992-1997 pour cinq sérotypes de salmonelles selon trois approches méthodologiques : les moyennes des données historiques correspondant au système de surveillance actuel du CNR et deux modélisations de séries chronologiques, le Box & Jenkins et le Serfling. Cette étude a été réalisée sous l'hypothèse d'absence d'épidémie hors des périodes épidémiques prédéfinies par le RNSP et en considérant la semaine comme unité d'alerte.

Un modèle a été estimé par analyse spectrale permettant de construire un seuil Serfling pour les cinq séries de sérotypes (Paratyphi B, Brandenburg, Goldcoast, Typhimurium, Hadar). Pour trois séries (Brandenburg, Goldcoast, Hadar), un modèle ARIMA n'a pas pu être identifié selon la procédure Box & Jenkins. L'analyse a porté d'abord sur les performances purement quantitatives par le décompte à la semaine des dépassements de seuils.

A l'analyse des dépassements de seuil pour la série Paratyphi B, le seuil Box & Jenkins est apparu le plus réactif et avec les meilleurs qualités de détection en terme de sensibilité ($Se = 0.93$), spécificité ($Sp = 1.0$) et valeur prédictive positive ($VPP = 1.0$). Le dépassement du seuil Box & Jenkins (semaines 93-34 à 93-47) correspond au plus près à la période des semaines définies comme épidémiques (semaines 93-33 à 93-47) et a déterminé le mieux la fin de l'épidémie. Les seuils CNR et Serfling ont présenté des qualités très similaires. Le seuil CNR a détecté l'épidémie une semaine plus tôt que le seuil Serfling (93-35 versus 93-36) et l'alerte a été également prolongée avec une semaine d'écart (93-48 versus 93-49), au delà de la période épidémique prédéfinie. Les deux seuils ont détecté des semaines épidémiques supplémentaires, cinq pour le seuil CNR et onze pour le seuil Serfling, dont deux semaines (12.5%) détectées simultanément sur une période où le seuil Serfling a été dépassé sur six semaines consécutives (semaines 94-34 à 94-39).

Les seuils CNR et Serfling construits pour les séries aux profils similaires, *S. Brandenburg* et *S. Goldcoast*, ont montré des performances équivalentes à ce qui a été constaté pour la série *S. Paratyphi B*. Pour *S. Brandenburg*, 11 semaines pour le seuil CNR versus 13 semaines pour le seuil Serfling ont été détectées comme potentiellement épidémiques, réparties sur les deux années 1994-1995, dont 3 semaines avec dépassement simultané des deux seuils (12.5%). Le seuil CNR a détecté l'épidémie à *S. Goldcoast* (94-26 à 94-39) avec une meilleure prédiction du début et de la fin de l'épidémie que le seuil Serfling (94-26 à 94-39 versus 94-27 à 94-38). Treize semaines pour le seuil CNR versus 10 semaines pour le seuil Serfling ont été détectées comme potentiellement épidémiques, dont 3 semaines avec dépassement simultané des deux seuils (13%). Les dépassements supplémentaires du seuil CNR par le nombre cumulé des observations ont été tous retrouvés systématiquement en fin de mois (1 ou 2 dernières semaines du mois).

Concernant les séries avec tendance sur le long terme (*Typhimurium*, *Hadar*), les seuils avaient une sensibilité médiocre, estimée entre 25 et 60%, et une spécificité entre 61 et 96%. Pour *Typhimurium* les seuils CNR et Serfling avaient la même sensibilité, 38% versus 25% pour le seuil Box & Jenkins. Par contre le seuil Box & Jenkins a eu une meilleure spécificité, 94% versus 74% pour le seuil CNR et 88% pour le seuil Serfling. Le seuil CNR a détecté 25 semaines épidémiques supplémentaires dont 19 correspondent à la dernière semaine d'un mois (76%). Le seuil Box & Jenkins a détecté 6 semaines supplémentaires versus 12 semaines pour

le seuil Serfling, dont 3 étaient identiques au seuil CNR. L'épidémie à *S. Hadar* a été détectée plus précocement par le seuil CNR avec une alerte mieux ajustée sur la période épidémique définie. Par contre, hors période épidémique prédéfinie, le seuil CNR a été dépassé sur 35 semaines versus 4 semaines pour le seuil Serfling. Ceci a contribué à estimer une spécificité à 96% pour le seuil Serfling versus 61% pour le seuil CNR.

Les analyses complémentaires pour Paratyphi B et Typhimurium ont fait apparaître des concentrations de "cas" de salmonelloses par zones géographiques et par classes d'âge, sur des périodes non définies a priori comme épidémiques. Rétrospectivement il a été considéré qu'il s'agissait d'épidémies passées inaperçues.

5-2 Référence "brute" versus référence "corrigée"

5-2-1 Référence "brute"

Lorsqu'un modèle a été identifié, le seuil Box & Jenkins a montré la meilleure valeur prédictive positive. Il a donné un haut niveau de sensibilité pour Paratyphi B et est resté de bonne qualité avec l'augmentation du délai d'alerte. Pour les trois sérotypes aux profils similaires (Paratyphi B, Brandenburg, Goldcoast), il n'a pas été possible de conclure sur la comparaison seuil CNR et seuil Serfling. Le seuil CNR est meilleur que le seuil Serfling pour Paratyphi B, c'est l'inverse pour Goldcoast, et les deux ont été équivalents pour Brandenburg. Ces résultats ont été retrouvés dans les évaluations sérotypes regroupés.

Pour la série Typhimurium, les qualités de détection ont été aussi mauvaises pour les trois seuils, tant en sensibilité qu'en valeur prédictive positive. Typhimurium est un sérotype prédominant. Un excès inhabituel de cas peut être difficile à repérer au sein d'un nombre de cas importants en bruit de fond.

Quel que soit le sérotype, la sensibilité du seuil Serfling a été plus robuste que celle du seuil CNR. La sensibilité a moins diminué avec l'allongement du délai d'alerte. Pour deux sérotypes (Paratyphi B, Brandenburg), le gain en VPP parallèlement à l'allongement du délai d'alerte a été meilleur pour le seuil CNR que pour le seuil Serfling. Au contraire, pour deux autres sérotypes (Typhimurium, Goldcoast), le gain en VPP a été meilleur pour le seuil Serfling que pour le seuil CNR. Cette différence a été retrouvée dans l'évaluation en sérotypes regroupés, avec un meilleur gain en VPP pour le seuil CNR que pour le seuil Serfling. La part importante de Paratyphi B dans les résultats groupés peut expliquer ce résultat à l'avantage du seuil CNR.

Au vu des résultats groupés, les sensibilités des trois seuils ont été équivalentes quel que soit le délai d'alerte (0.70 à 0.76 pour 1 semaine, 0.48 à 0.52 pour 3 semaines). Le seuil Box & Jenkins a montré une meilleure VPP (0.73 à 1.00 pour 1 à 3 semaines). Les seuils Serfling et CNR ont montré des performances prédictives équivalentes. L'avantage est apparu pour le seuil Serfling sur le plan réactivité (délai d'alerte court de 1 à 2 semaines) par rapport au seuil CNR (délai d'alerte long de 3 semaines).

5-2-2 Référence "corrigée"

Le seuil Box & Jenkins est apparu avec des hauts niveaux de VPP (≥ 0.88). Le seuil Serfling est apparu de meilleure qualité que le seuil CNR pour trois sérotypes (Paratyphi B, Goldcoast, Typhimurium) et équivalent à celui-ci pour Brandenburg. Ces résultats ont été

retrouvés dans l'évaluation tous sérotypes confondus car les VPP du seuil CNR ont été plus fréquemment estimées à moins de 50% pour un délai d'alerte de 1 semaine.

Dans l'ensemble, la sensibilité des seuils Serfling et CNR ont été équivalentes. Cependant le seuil Serfling a gardé une sensibilité plus stable lorsque les délais d'alerte ont varié de 1 à 3 semaines.

5-2-3 De la référence "brute" à la référence "corrigée"

Les seuils Box & Jenkins et Serfling ont été les plus robustes en VPP. Le seuil Serfling a été le plus robuste sur la sensibilité. Ces résultats en faveur de la méthode Serfling s'expliquent essentiellement par la détection bien ajustée à la période épidémique à *S. Paratyphi B* retrouvée à l'analyse complémentaire sur l'année 1994 (semaines 94-33 à 94-48). Le seuil Serfling est apparu comme un bon compromis entre sensibilité et valeur prédictive.

La validité des deux références réside dans leur définition, indépendante des seuils d'alerte construits. La validité supplémentaire de la référence "corrigée", qui justifie son utilisation, est l'intégration de vraies épidémies repérées a posteriori sur des arguments descriptifs et subjectifs. Si ces événements avaient été signalés sur la base d'une augmentation importante, durable et associée à des particularités de lieux et de personnes, ils auraient été investigués.

5-3 Qualité des données

Les séries chronologiques ont été construites, pour chaque sérotype, sur la base des dates de réception des souches de salmonelles au CNR. En effet, le système de surveillance et d'alerte actuel du CNR ne tient compte que des souches reçues et sérotypées au CNR. Le délai entre la date d'isolement d'une souche en laboratoire et la date de réception au CNR a été estimé en moyenne à 8 jours sur les données de l'étude (de 1 à 239 jours; centile 75 = 9 jours). Compte tenu de ce délai qui peut être extrêmement variable et sensible à des aléas externes, il n'a pas été envisagé d'avoir recours à une surveillance sur la base des dates d'isolement. Au dernier recensement réalisé par le CNR, un à deux tiers des LABM français envoient leur souches de salmonelles au CNR, soit environ 2800 sur les 4400 LABM publiques et privés.

Les données utilisées dans l'étude ont donc été les données reçues et utilisées en routine par le CNR pour l'alerte. Ainsi il n'a pas été effectué de sélection préalable pour éliminer les éventuels doublons, distinguer les souches à l'origine de TIAC, de cas sporadiques ou d'épidémie communautaire. Ces données brutes ont été considérées comme les plus adéquates pour déterminer la ligne de base de ce qui est reçu habituellement par le CNR, sous l'hypothèse d'absence d'épidémie.

5-4 Choix des modèles, avantages / inconvénients

5-4-1 Méthode CNR

Le seuil CNR représente un intermédiaire entre les deux autres méthodes en effectuant un "lissage séquentiel" qui intègre les fluctuations historiques à échéance variable (fluctuations d'un intervalle de trois mois prises en compte sur cinq ans). L'intérêt du seuil CNR réside en son utilisation et son interprétation pratiques. La méthode a été mise au point

sur place au CNR par les utilisateurs habituels du système de surveillance. Depuis sa mise en place en 1992, il est à présent complètement intégré à une utilisation en routine, au minimum hebdomadaire. Des analyses complémentaires sont disponibles avec plusieurs indicateurs quantitatifs ainsi que des analyses géographiques par découpage administratif ou de contiguïté. Des signaux visuels ont été développés sous la forme de ratio pour aider à l'interprétation du positionnement par rapport au seuil. De plus, cette méthode ne fait pas intervenir d'étapes complexes de modélisation. L'utilisation de ce procédé simple, croisé avec des informations départementales sur les TIAC (les déclarations obligatoires etc.), et régionales est une approche pragmatique satisfaisante de la détection des épidémies.

Cependant le système reste opérateur dépendant. Les interprétations du seuil M5c sont parfois difficiles. Elles nécessitent d'être associés aux autres indicateurs et à des connaissances spécifiques à chaque sérotype pour pouvoir conclure à un événement inhabituel. La valeur observée qui sert de comparaison par rapport au seuil est une valeur cumulée qui manque de réactivité. Par ailleurs, la méthode ne prend pas bien en compte les changements récents, notamment en cas de changements de tendance et de niveaux en quelques saisons (Typhimurium, Hadar).

5-4-2 Méthode Box et Jenkins

Pour être utilisée de façon optimale pour des prédictions par analyse temporelle, une série de données doit satisfaire plusieurs conditions qui ne sont pas toujours possibles à réunir et dont les principales sont (41) : 1) des variations saisonnières claires, 2) des variations saisonnières de type additif, 3) des données continues sans rupture dans le temps, 4) un nombre suffisant de données (≥ 100) (26). Les séries doivent être suffisamment longues et stables nécessitant parfois une étape d'écrêtement préalable (17). Des variations importantes dans la distribution temporelle gênent l'identification d'un modèle adéquat (ex : *S. Hadar*).

La méthode Box et Jenkins identifie un modèle qui tend à s'ajuster au plus près aux données (42) et possède la propriété optimale de minimiser l'erreur de la prévision (43). La limite unilatérale à 95% du modèle qui définit le seuil est alors peu sensible aux fluctuations aléatoires car celles-ci sont intégrées à part entière dans le modèle. A priori le nombre de souches reçues au CNR chaque semaine est extrêmement aléatoire. Malgré cela, la méthode Box et Jenkins a été la plus satisfaisante pour déclencher une alerte adéquate pour Paratyphi B.

Au delà de la compréhension du concept d'analyse temporelle, la méthode Box et Jenkins est une modélisation qui nécessite un investissement important en ressources humaines compétentes et en temps (41). Plusieurs modèles sont parfois possibles, doivent être testés et comparés selon un processus itératif. Pour Chatfield, "practical experience is a must" (31). La part subjective dans le choix du modèle nécessite une grande expérience dans la décision d'une transformation éventuelle des données et dans l'interprétation des fonctions d'autocorrélation. Il n'a pas été nécessaire de réaliser une transformation préalable des données, ce qui a permis d'éviter une "sur-transformation" et son corollaire, le biais dans les prédictions (41, 44).

Un ajustement incorrect du modèle SARIMA à une série donne des estimations de paramètres imprécises et des prédictions moyennes avec de larges intervalles de confiance (22, 45). Le modèle SARIMA intègre les changements qui surviennent dans le passé récent. Les estimations des paramètres du modèle se modifient en conséquence, ainsi que les prédictions dont les erreurs sont minimisées de façon asymptotique (41). De plus, à chaque

saison, les données supplémentaires permettent un ajustement plus précis du modèle SARIMA à la série, comparativement à la saison précédente. Ainsi le seuil Box & Jenkins élaboré pour Paratyphi B par périodes de 52 semaines a fait apparaître une meilleure précision à chaque saison supplémentaire incluse dans les prédictions, permettant de construire un seuil plus stable et plus valide. En théorie seul le point immédiatement après la série modélisée peut être prédit avec confiance (31, 41).

5-4-3 Méthode Serfling

La méthode Serfling est plus rapide à appréhender et prend l'ensemble des données disponibles comme un tout. Le concept une fois acquis est d'application simple. Les étapes successives qui décomposent le processus de modélisation est de compréhension plus facile. L'analyse spectrale effectue un lissage qui prend en compte les caractéristiques essentielles de chaque série, y compris quand celles-ci présentent une saisonnalité peu marquée. A ce titre, cette méthode, par une analyse de type fréquentiel, permet une approche descriptive et explicative de la série qui pourrait être utilisée pour "préparer" la modélisation de la série selon la méthode Box et Jenkins (44).

Les méthodes de prévision fondées sur l'analyse spectrale sont "statiques" car les relations entre les observations sont supposées rester constantes dans le temps (43). Ainsi le modèle est peu ajusté aux changements récents puisque le même poids est donné à toutes les données antérieures (44). L'ajustement aux données n'est pas synonyme de précision pour les prévisions qui en résultent (42, 43). Les prédictions moyennes et l'intervalle de confiance créent une zone homogène au sein de laquelle les fluctuations hebdomadaires varient de façon aléatoire ce qui correspond mieux au concept de décision en santé publique qu'une prévision exacte à la semaine (33). Cependant le lissage général, qui sert de base pour les prédictions, rend le seuil très sensible aux fluctuations aléatoires qui surviennent d'une semaine à l'autre lorsqu'elles sont importantes.

5-5 Propositions

5-5-1 Notion de seuil

La notion de seuil fait intervenir la notion de niveau de "risque" à partir duquel l'alerte est déclenchée. Dans notre étude le risque d'erreur a été fixé a priori systématiquement à 5% dans une optique unilatérale. C'est le niveau choisi habituellement pour le seuil qui est fixé à 1.64 fois l'écart type de la prévision (33, 17, 45). Une étape supplémentaire serait une étude de sensibilité - spécificité de chaque seuil d'alerte déterminé, en faisant varier les niveaux d'intervalle de confiance sur le(s) modèle(s) retenu(s).

Au cours d'une saison non épidémique, un dépassement du seuil d'alerte peut survenir en moyenne sur deux à trois semaines, du seul fait du hasard (46). Si ce seuil est utilisé comme indicateur d'épidémie, un dépassement sur une seule semaine conduit à un risque d'erreur. Pour minimiser le risque d'une telle erreur, il est possible de fixer un critère temporel qui définit l'épidémie pour le dépassement de seuil sur un délai de n semaines successives (21, 33, 44).

5-5-2 Critères pour un algorithme décisionnel

Une méthode de seuil à développer serait une adaptation du seuil CNR pour déterminer pour chaque semaine un indicateur qui prend en compte une valeur moyenne et non plus une valeur cumulée. Sur le même principe que la méthode CNR, une moyenne mobile pourrait être calculée avec une fenêtre de cinq semaines, centrée sur la semaine en cours et déplacée pas à pas sur les cinq années précédentes (47, 48).

De plus, le délai entre l'isolement de la souche au LABM et la réception au CNR est en moyenne de 8 jours. Il est possible que la prise en compte de ce délai d'acheminement améliore la précocité de détection d'une épidémie. Un facteur correcteur d'une semaine sur le seuil pourrait être intégré dans le système (49). Le décalage entre observations et seuil reviendrait à comparer la valeur hebdomadaire observée au temps t à la valeur seuil attendue au temps $(t-1)$.

Dans une surveillance prospective hebdomadaire, où le seuil définit l'alerte, il est impossible de savoir si un dépassement de seuil pour une semaine constitue une alerte justifiée. En effet, il s'agit dans ce cas, soit d'un début d'alerte qui va se confirmer par un dépassement de seuil sur les semaines suivantes, soit d'une fausse alerte (faux positif) isolée sur une semaine et après laquelle les observations sont en dessous du seuil. Plusieurs niveaux d'alerte seraient donc à définir en fonction du nombre de semaines avec dépassement de seuil, en considérant différemment les alertes en semaines successives continues et en semaines éparées discontinues. Ainsi il serait possible de différencier un niveau de surveillance passive (niveau 1 = pas de dépassement de seuil), un niveau de sensibilisation (niveau 2 = première semaine de dépassement, analyses complémentaires géographiques et par classes d'âge), un niveau de sensibilisation passive - active (niveau 2+ = seuil dépassé plus de deux semaines, contacts directs avec les correspondants régionaux et départementaux, services vétérinaires etc.), un niveau d'alerte avérée (niveau 3 = enquête épidémiologique).

Il est apparu dans notre étude que le dépassement simultané des trois seuils correspondait de façon adéquate à la période prédéfinie épidémique. Ce résultat est en faveur d'une surveillance "multicritère" où la décision serait rapportée à plusieurs indicateurs, en particulier plusieurs seuils, (temporels, géographiques, caractéristiques de personnes) permettant d'accroître les performances de détection et de déclenchement de l'alerte (45, 50).

Dans le but de construire un algorithme d'aide à la décision, plusieurs indicateurs sont à considérer. En effet, il existe une difficulté à établir un critère absolu qui définit l'épidémie communautaire. En pratique, toute épidémie se définit en temps, lieu et personne (51). Dans notre étude, ces trois composantes sont apparues indispensables dans la détection d'une épidémie et du déclenchement de l'alerte. Envisager un algorithme décisionnel c'est intégrer au minimum les trois dimensions suivantes :

- un seuil chronologique qui servirait de signal principal et permettrait de repérer la survenue d'un événement inhabituel ;
- un indicateur d'analyse régionale et départementale qui indiquerait un agrégat géographique ;
- un indicateur de personne tel l'âge des sujets chez lesquels une souche de salmonelle est isolée, catégorisant ainsi une population plus particulièrement touchée.

L'association d'un seuil national par sérotype à un seuil régional tout sérotype confondu pourrait également améliorer les performances de détection d'épidémie en un lieu donné et focaliser la surveillance (52).

Le développement des réseaux informatiques associé à l'automatisation de l'analyse permettrait une réactivité plus rapide du système, depuis le recueil des données jusqu'au déclenchement de l'alerte (53-55).

6 - CONCLUSION

Assurer la sécurité sanitaire nécessite une vigilance constante et une capacité d'intervention appropriée. Aujourd'hui en France, pour parer au risque d'épidémies communautaires à salmonelles, le système de surveillance des salmonelloses du CNR et sa collaboration étroite avec le RNSP permettent de concilier de façon satisfaisante ces deux aspects.

L'expertise d'une situation inhabituelle suspecte d'être épidémique est un préalable indispensable à la prise de décision pour l'action et nécessite des indicateurs d'évaluation fiables. La notion de seuil épidémique, surtout de seuil de déclenchement de l'alerte et d'intervention est difficile à établir de façon absolue. Notre étude a montré l'intérêt de disposer de seuils de surveillance nationaux par sérotype, si ce n'est de seuils d'alerte, pour repérer les anomalies, les quantifier et guider les choix dans le type d'intervention à adopter pour y remédier. Des critères de deuxième ligne plus détaillés tels que l'analyse géographique ou l'âge des cas de salmonelloses sont apparus des compléments indispensables pour identifier une épidémie.

La décision doit donc reposer sur un processus "multicritère". Cette étude était une première étape dans ce sens. Elle a contribué à montrer des combinaisons potentiellement efficaces associant des critères quantitatifs et qualitatifs. D'autres études sont nécessaires pour évaluer les performances de telles combinaisons et pour l'élaboration d'un processus décisionnel systématique sous la forme d'un algorithme.

REFERENCES

1. Benenson A. *Control of communicable diseases in man*, Washington, DC: American Public Health Association, 1995. pp. 410-415.
2. Lepoutre A., Salomon J., Charley C., and Le Querrec F. Les toxi-infections alimentaires collectives en 1993. **Bull Epidemiol Hebd** 1994;52:245-7.
3. Tauxe R.V. *Salmonella*: a postmodern pathogen. **J of Food Protection** 1991;54(7):563-8.
4. Bouvet E. and Hubert B. Epidémiologie des salmonelloses mineures. **Rev Prat** 1992;42:2275-8.
5. Le Minor L. and Véron M. *Bactériologie Médicale*, Paris:Flammarion, 1990 pp. 411-425.
6. Bouvet P., Fayolle V., Guibert G., K'Ouas G, Lenormand P, Metz L., and Grimont P.A.D. La surveillance des salmonelloses en France : données récentes du centre de référence. **Salmonella and Salmonellosis Proceedings**, Ploufragan, France 1997, pp. 559-64.
7. Desenclos J.C., Bouvet P., Pierre V., Brisabois A., Fremy S., Lahellec C., et al. Epidémiologie des infections à *Salmonella* : tendances récentes en France et en Europe. **Bull Soc Fr Microbiol** 1996;11:209-15.
8. Anonyme An outbreak of *Salmonella* Agona due to contaminated snacks. **CDR** 1995;7:29-32.
9. Desenclos J.C., Haeghebaert S., Delarocque-Astagneau E., Vaillant V., Decludt B., Rebière I., et al. Epidémiologie des épidémies communautaires à *Salmonella* en France, 1993-1996. **Salmonella and Salmonellosis Proceedings**, Ploufragan, France 1997, pp. 555-8.
10. Desenclos J.C., Rebière I., Bouvet P., Benz-Lemoine E., Robain M., Bouvier N., et al. Bilan de l'investigation de 5 épidémies communautaires de salmonellose, France 1993-1994. **Bull Epidemiol Hebd** 1996;9:39-43.
11. Altekruse S., Koehler J., Hickman-Brenner F., Tauxe R.V., and Ferris K. A comparison of *Salmonella* enteritidis phage types from egg-associated outbreaks and implicated laying flocks. **Epidemiol Infect** 1993;110:17-22.
12. Mishu B., Koehler J., Lee L.A., Rodrigue D., Hickman-Brenner F., Blake P., and Tauxe R.V. Outbreaks of *Salmonella* enteritidis Infections in the United States, 1985-1991. **J Infect.Dis.** 1994;169:547-52.
13. Decludt B., Haeghebaert S., Bouvet P., and Grimont P.A.D. Epidémie de salmonellose à *Salmonella* sérotype Hadar (France, juin-septembre 1995). **Bull Epidemiol Hebd** 1996;32:140-1.
14. Bouvet P. and Grimont P.A.D. Augmentation de l'incidence des infections à *Salmonella* sérotype Hadar en France. **Bull Epidemiol Hebd** 1996;32:139-40.

15. Desenclos J.C., Bouvet P., Benz-Lemoine E., Grimont F., Desqueyroux H., Rebière I., and Grimont P. Large outbreak of *Salmonella enterica* serotype *paratyphi B* infection caused by a goats' milk cheese, France, 1993: a case finding and epidemiological study. **BMJ** 1996;312:91-4.
16. Galley A. Estimation du nombre de toxi-infections alimentaires collectives à salmonelles : à partir de trois systèmes de surveillance par la méthode capture-recapture. **Réseau National de Santé Publique**, Saint-Maurice, France. Rapport, septembre 1997, 64 pages.
17. Watier L., Richardson S., and Hubert B. A time series construction of an alert threshold with application to *S. Bovismorbificans* in France. **Stat Med.** 1991;10:1493-509.
18. Thacker S.B. and Berkelman R.L. Public Health Surveillance in the United States. **Epidemiologic Reviews** 1988;10:164-90.
19. Watier L. Analyse des séries temporelles des infections à salmonelles non typhiques. **Rev Epidem et Santé Publ** 1995;43:173-85.
20. Watier L. and Richardson S. Surveillance des toxi-infections alimentaires dues aux salmonelles : élaboration d'un seuil d'alerte épidémique. **Rapport de convention entre le Réseau National de Santé Publique et l'INSERM**, 1996, 23 pages.
21. Watier L. and Richardson S. Surveillance des toxi-infections alimentaires dues aux salmonelles : élaboration d'un seuil d'alerte épidémique. **Rapport de convention entre le Réseau National de Santé Publique et l'INSERM**, 1994, 44 pages.
22. Watier L. Revue méthodologique de quelques techniques spécifiques à l'analyse des séries temporelles en épidémiologie et santé publique. **Rev Epidem et Santé Publ** 1995;43:162-72.
23. Chatfield C. *Problem Solving : A statistic guide*, London:Chapman & Hall, 1993.
24. Armitage P. *Statistical Methods in Medical Research*, Oxford and Edinburgh:Blackwell Scientific Publications, 1987.
25. Box G.E.P. and Jenkins G.M. *Time Series Analysis. Forecasting and Control*, San Francisco:Holden-Day, 1976.
26. Helfenstein U. Box-Jenkins modelling of some viral infectious diseases. **Stat Med** 1986;5:34-47.
27. Coulombier D., Quénel P., Coignard B. Etude de cas : réalisation d'un seuil épidémique pour une série chronologique par la méthode de Box et Jenkins. Formation EPITER. **Réseau National de Santé Publique**. Saint-Maurice, France. Octobre 1997, 28 pages.
28. Ljung G. and Box G. On a measure lack of fit in times series models. **Biometrika** 1978;65:297-303.
29. Box G.E.P. and Pierce D.A. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. **J Am Stat Assoc.** 1970;65:1509-26.
30. Akaike H. On the likelihood of a time series model. **Statistician** 1978;27:215-35.

31. Chatfield C. *The Analysis of Time Series : An Introduction*, London:Chapman and Hall 1989.
32. Transnational College of LEX. *Who is Fourier ? A Mathematical Adventure*. Language Research Foundation 1995.
33. Serfling R.E. Methods for current statistical analysis of excess pneumonia and influenza deaths. **Public Health Report** 1963;78:494-506.
34. Housworth J. and Langmuir A.D. Excess mortality from epidemic influenza, 1957-1966. **Am J Epidemiol** 1974;100:40-8.
35. Centers for Disease Control Guidelines for evaluating surveillance systems. **MMWR** 1988;37(S-5):1-18.
36. Réseau National de Santé Publique. Le point sur les épidémies de Salmonellose, paratyphi B et virchow. **Réseau National de Santé Publique** .Rapport. Décembre 1993.
37. Delarocque Astagneau, E., De Valk, H. Epidémie de salmonellose à *Salmonella enterica* sérotype Typhimurium, Jura, mai-juillet 1997. **Réseau National de Santé Publique**. Saint-Maurice, France. Rapport.
38. Réseau National de Santé Publique. Epidémie de salmonellose à *Salmonella* sérovar Brandenburg, France, juin-août 1994. Rapport d'investigation. **Réseau National de Santé Publique**, Saint-Maurice, France. Décembre 1994.
39. Réseau National de Santé Publique. Une épidémie communautaire de salmonellose sérotype Goldcoast survenue dans le département de la Drôme en juillet et août 1994. Rapport d'investigation. **Réseau National de Santé Publique**, Saint-Maurice, France. Mai 1995.
40. Réseau National de Santé Publique. Epidémie de salmonelloses à *Salmonella* sérotype Hadar, France, juin-septembre 1995. Rapport d'investigation. **Réseau National de Santé Publique**, Saint-Maurice, France. Décembre 1995.
41. Chatfield C. and Prothero D.L. Box-Jenkins Seasonal Forecasting: Problems in a Case-study. **J R Statist Soc A** 1973;136:295-336.
42. Choi K. and Thacker S.B. An evaluation of influenza mortality surveillance, 1962-1979. II. Percentage of Pneumonia and Influenza Deaths as an Indicator of Influenza Activity. **Am J Epidemiol** 1981;113:227-35.
43. Choi K. and Thacker S.B. An evaluation of influenza mortality surveillance, 1962-1979. I. Time Series Forecasts of Expected Pneumonia and Influenza Deaths. **Am J Epidemiol** 1981;113:215-26.
44. Quénel P. La surveillance épidémiologique des épidémies de grippe. Validation d'un critère d'alerte épidémique. Université Paris XI., 1993. **Doctorat**.
45. Quénel P. and Dab W. Influenza A and B epidemic criteria based on time-series analysis of health services surveillance data. **Europ J Epidemiol** 1998;14:275-85.

46. Quénel P., Dab W., Hannoun C., et al. Sensitivity, specificity and predictive values of health service based indicators for the surveillance of influenza A epidemics. **Intern J Epidemiol** 1994;23:849-55.
47. MMWR. Proposed Changes in Format for Presentation of Notifiable Disease Report Data. **MMWR** 1989;38(47): 805-9.
48. Farrington C.P. and Beale A.D. Computer-aided detection of temporal clusters of organisms reported to the Communicable Disease Surveillance Centre. **CDR** 1993;3(R6):78-82.
49. Watier L. Etude des variations temporelles des infections à salmonelles non typhiques : modélisation et prévision. Université de Paris XI., 1993. **Doctorat**.
50. Noah N.D. Transmissible agents. *Oxford Textbook of Public Health*. **Oxford Medical Publications**, 1991.pp. 417-34.
51. Dabis F., Drucker J., and Moren A. *Epidémiologie d'intervention*. **Paris:Arnette**, 1992, pp. 69-102.
52. Helfenstein U. The use of Transfer Function Models, Intervention Analysis and Related Time Series Methods in Epidemiology. **Intern J Epidemiol** 1991;20(3):808-15.
53. Grant A.D. and Eke B. Application of information technology to the laboratory reporting of communicable disease in England and Wales. **CDR** 1993;3(R6):75-8.
54. Valleron A. and Garnerin P. Computerised surveillance of communicable diseases in France. **CDR** 1993;3(R6):82-7.
55. Fisher I.S.T., Rowe B., Bartlett C.L.R., and Gill O.N. Salm-Net - laboratory based surveillance of human salmonella infections in Europe. **PHLS Microbiol Dig** 1994;11:181-2.

ANNEXES

Centre de Référence des Salmonella et Shigella
 Institut Pasteur
 28 rue du Docteur Roux
 75724 PARIS CEDEX 15



Institut Pasteur - Paris

Unité des Entérobactéries (Professeur P. A. D. Grimont)
 Centre National des Salmonella et Shigella (adjoint : P. Bouvet)
 Centre Nat. de Typage Moléculaire Entérique (adj.: F. Grimont)
 TEL 0145688339 FAX : 0145688837

* Fiche de renseignement devant accompagner chaque souche *

Laboratoire Adresse complète et lisible du laboratoire expéditeur

Nom complet :

N° et rue :

Ville

Examen demandé pour la souche jointe

Cocher les cases choisies

Aucune demande (Information seulement, aucune souche envoyée).

Sérotype trouvé :

- Identification antigénique complète d'une Salmonella
 - Agglutinations O minimales effectuées par le laboratoire expéditeur
 (Résultat +/-)
 4,5 6,7,8 9 VI Autres:
- Agglutinations minimales non effectuées (facturation B40)
- Identification complète d'une Shigella (gratuite)
- Identification d'une autre Entérobactérie (facturation B60)

Identification d'un autre bacille à Gram négatif (facturation B60)
 Groupe bactérien suspecté:

Lysotypie de Salmonella sérotypes Typhi / Paratyphi A /
 Paratyphi B / Dublin (et uniquement en cas d'épidémie, Typhimurium /
 Enteritidis)

Typage moléculaire (entente préalable Tél. 0145688344)

Souche d'origine non humaine facturée (voir ci-dessous)

Demande particulière : entente préalable Tél: 0145688339 / Fax: 0145688837
 Joindre copie des résultats déjà obtenus.

Renseignements épidémiologiques essentiels

• PRELEVEMENT HUMAIN

Nom du patient ou Réf.

Age: <1 1-5 6-14 15-64 ≥65

Sexe : F / M Statut: Malade Porteur inconnu

Origine: Sang Selles Urines Bile

Pus LCR Autre ?.....

Cas: isolé Voyage récent (pays, date)

épidémie familiale épid. hospitalière T.I.A.C.

Infect. collective crèche colonie vacances école

Nombre de cas: Aliment en cause

Date d'isolement: et précisions :

• PRELEVEMENT NON-HUMAIN (Facturé. Info: Tél 0145688339)

Référence de la souche:

Nature exacte du prélèvement

Vétérinaire.....

Alimentaire.....

Environnement

Origine géographique de l'animal ou du produit:

Nous vous remercions pour votre collaboration à la surveillance
 épidémiologique des infections dues aux Entérobactéries

RELEVÉ DES SOUCHES D'ORIGINE HUMAINE ETUDIÉES AU CNRS : COMPARAISON AVEC LES DONNÉES HISTORIQUES

98-4-provisoire (20/04/98 - 13:25:25) RELEVÉ PAR SÉROTYPE

- La moyenne historique calculée sur 5 ans (MS) représente la moyenne des nombres de souches enregistrées le même mois et les mois précédents et suivants, durant les 5 dernières années.
- Un sérotype est considéré uniquement si le nombre mensuel de souches est supérieur ou égal à : 5
- POUR UN MOIS DONNÉ, LES VALEURS ANORMALEMENT ÉLEVÉES DANS LA SÉRIE SONT DÉTECTÉES ET MINORÉES DE LA MANIÈRE SUIVANTE :
 Si valeur supérieure à : Moyenne du mois + 1 x Écart-type
 remplacée par : Moyenne du mois + 0 x Écart-type
- SEUIL D'ALERTE CHOISI: dépassement de la MSC de 20 pour cent
- ATTENTION : SEULES LES SÉROTYPES DÉPASSANT LE SEUIL D'ALERTE SÉLECTIONNÉS SONT IMPRIMÉS

CENTRE DE RÉFÉRENCE DES SALMONELLA ET SHIGELLA
 Unité des Entérobactéries
 INSTITUT PASTEUR, Paris

TEL : 01 45 68 8339 et 8345
 FAX : 01 45 68 88 37
 e-mail : ppriscnt@pasteur.fr / phouvet@pasteur.fr

INDICATEURS D'ALERTE (*)			NOMBRE (%) 98-4	mens à mai sur 5 ans					avr. sur 5 ans			Cumul mois 1 à 4				mens et avr. (***)		mai à avr. (***)	
MSC	2MOIS*3	12MOIS*3		Min	Max	Méd	MS	MSC	IC99	Moy	Méd	Scm	moy 3 ans	97	98	Indic. (*)	98	Moy*3ans	98

.....	Typhi	7 (2.7)	1	8	2	2	1	2	3	2	15	9	9	10		8	3.3	56	42
		Anatum	6 (2.3)	0	9	2	3	2	3	1	2	8	9	15	10		6	4.6	57	59.6
		Goldcoast	5 (1.9)	1	7	3	3	2	3	2	2	12	10	11	16	14	5.6	54	87.3

(*) L'INDICATEUR D'ALERTE permet de visualiser rapidement un dépassement d'une des valeurs examinées

L'indicateur ooooo indique que la moyenne (mensuelle 3mois/5 ans - somme de 2 mois/3 ans - somme de 12 mois/3 ans) est < 1.

INDICATEUR D'ALERTE	INDICATEURS ACTIVES UNIQUEMENT SI
DÉPASSEMENT (EN %) DE :							
- la moyenne historique (MScum)	1-49	50-99	100-149	150-199	200-249	>= 250	(Total mois - MSCcum) >= 2
- la moyenne (2MOIS*3)	1-9	10-19	20-29	30-39	>= 40		(Total 2 mois - 2 MOIS*3ans) >= 10
- la moyenne (12MOIS*3)	1-9	10-19	20-29	30-39	>= 40		(Total 12 mois - 12 MOIS*3ans) >= 10
- le cumul mois 1- 4	1-9	10-19	20-29	30-39	>= 40		(Total - MOY 3ans) >= 5

(***) Les moyennes historiques sont calculées à partir des mêmes périodes de 2 mois ou 12 mois au cours des 3 années précédentes.
 (1 min)

Mois	Année - Semaine	lag	Observés	Cumul des semaines /mois	Seuil CNR	Seuil Box & J	Seuil Serfling
Juillet	9327	80	1	1	12	5.6	4.4
	9328	81	1	2	12	3.7	4.7
	9329	82	1	3	12	4.8	5
	9330	83	2	5	12	4.7	5.4
Août	9331	84	3	3	18	5.2	5.9
	9332	85	3	6	18	4.4	6.5
	* 9333	86	4	10	18	5.6	7.2
	* 9334	87	5	15	18	5 **	7.8
	* 9335	88	7	23	18 **	4.9 **	8.3
Septembre	* 9336	89	30	30	24 **	3.7 **	8.7 **
	* 9337	90	25	55	24 **	5.3 **	9 **
	* 9338	91	20	75	24 **	4.7 **	9.1 **
	* 9339	92	17	92	24 **	4.4 **	9 **
Octobre	* 9340	93	21	21	33	5.3 **	8.8 **
	* 9341	94	31	52	33 **	5.6 **	8.4 **
	* 9342	95	16	68	! 33 **	5.9 **	7.9 **
	* 9343	96	23	91	33 **	5.5 **	7.3 **
Novembre	* 9344	97	15	15	30 **	6 **	6.7 **
	* 9345	98	23	38	30 **	6.3 **	6.2 **
	* 9346	99	16	54	30 **	6.3 **	5.6 **
	* 9347	100	13	67	30 **	6.4 **	5.2 **
	9348	101	6	73	30 **	6.5	4.9 **
Décembre	9349	102	5	5	24	6.7	4.6 **
	9350	103	3	8	24	6.2	4.5
	9351	104	2	10	24	6.7	4.4
	9352	105	0	10	24	7	4.4
Janvier	9401	106	0	0	21	6	4.4
	9402	107	1	1	21	6.1	4.5
	9403	108	4	5	21	5.8	4.5
	9404	109	0	5	21	5.9	4.5
Février	9405	110	4	4	12	6.5	4.5
	9406	111	1	5	12	6.8	4.5
	9407	112	3	8	12	7.6	4.4
	9408	113	2	10	12	6.8	4.4
Mars	9409	114	1	1	9	8.5	4.3
	9410	115	1	2	9	7.1	4.3
	9411	116	3	5	9	6.9	4.4
	9412	117	2	7	9	7.8	4.4
	9413	118	0	7	9	7.5	4.6
Avril	9414	119	2	2	9	7.7	4.7
	9415	120	2	4	9	7.5	4.9
	9416	121	1	5	9	8.8	5
	9417	122	3	8	9	8.4	5.1
Mai	9418	123	2	2	7.2	8.7	5.2
	9419	124	3	5	7.2	7.9	5.2
	9420	125	0	5	7.2	8.3	5.1
	9421	126	0	5	7.2	8	5
Juin	9422	127	2	2	9	8.3	4.8
	9423	128	4	6	9	7.8	4.6
	9424	129	3	9	9 **	8.3	4.4
	9425	130	1	10	9 **	8.2	4.2
	9426	131	2	12	9 **	8.8	4.1

* : semaines épidémiques définies par l'investigation

** : semaines épidémiques détectées par le seuil

! : alerte déclenchée par le CNR

(*) : semaines épidémiques supplémentaires en référence "corrigée"

Mois	Année - Semaine	lag	Observés	Cumul des semaines /mois	Seuil CNR	Seuil Box & J	Seuil Serfling
Juillet	9427	132	0	0	12	8.9	4.1
	9428	133	0	0	12	8.4	4.3
	9429	134	3	3	12	8.9	4.6
	9430	135	2	5	12	9.2	4.9
Août	9431	136	2	2	19.2	9.7	5.4
	9432	137	1	3	19.2	9.5	5.9
	(*) 9433	138	4	7	19.2	10.3	6.5
	(*) 9434	139	7	14	19.2	10.4	7 **
	(*) 9435	140	9	23	19.2 **	11	7.4 **
Septembre	(*) 9436	141	10	10	31.2	11.4	7.8 **
	(*) 9437	142	10	20	31.2	10.1	8 **
	(*) 9438	143	10	30	31.2	10.8	8.1 **
	(*) 9439	144	11	41	31.2 **	11.3	8 **
Octobre	(*) 9440	145	7	7	43.2	10.6	7.8
	(*) 9441	146	7	14	43.2	10.4	7.5
	(*) 9442	147	9	23	43.2	10.3	7.1 **
	(*) 9443	148	2	25	43.2	10.8	6.7
Novembre	(*) 9444	149	10	10	39	10.5	6.3 **
	(*) 9445	150	3	13	39	10.4	5.9
	(*) 9446	151	3	16	39	10.4	5.6
	(*) 9447	152	6	22	39	10.5	5.3 **
	(*) 9448	153	6	28	39	12	5.2 **
Décembre	9449	154	3	3	27	11.8	5.1
	9450	155	1	4	27	11.2	5
	9451	156	4	8	27	11.1	5
	9452	157	1	9	27	10.7	5
Janvier	9501	158	2	2	15	6.4	5
	9502	159	6	8	15	7.1	4.9 **
	9503	160	2	10	15	8.2	4.8
	9504	161	1	11	15	6.3	4.6
	9505	162	0	11	15	9.4	4.4
Février	9506	163	0	0	12	7.5	4.2
	9507	164	3	3	12	8.9	4.1
	9508	165	0	3	12	8.2	3.9
	9509	166	1	4	12	8.6	3.8
Mars	9510	167	3	3	9.6	7.7	3.7
	9511	168	1	4	9.6	8.8	3.7
	9512	169	1	5	9.6	9.4	3.7
	9513	170	3	8	9.6	7.5	3.8
Avril	9514	171	1	1	7.2	8.3	4
	9515	172	0	1	7.2	8.4	4.1
	9516	173	4	5	7.2	8.8	4.2
	9517	174	0	5	7.2	9.4	4.3
Mai	9518	175	0	0	9.6	9.2	4.3
	9519	176	0	0	9.6	9.3	4.3
	9520	177	1	1	9.6	8.3	4.2
	9521	178	1	2	9.6	8.4	4.1
	9522	179	1	3	9.6	10	4
Juin	9523	180	0	0	12	12	3.9
	9524	181	1	1	12	10.2	3.8
	9525	182	2	3	12	8.7	3.8
	9526	183	0	3	12	10.1	3.9

* : semaines épidémiques définies par l'investigation

** : semaines épidémiques détectées par le seuil

! : alerte déclenchée par le CNR

(*) : semaines épidémiques supplémentaires en référence "corrigée"

Mois	Année - Semaine	lag	Observés	Cumul des semaines /mois	Seuil CNR	Seuil Box & J	Seuil Serfling
Janvier	9601	210	37	37	121.0	62.3	53.9
	9602	211	46	83	121.0	58.3	53.6
	9603	212	35	118	121.0	61.0	53.5
	9604	213	44	127	121.0 **	62.5	53.7
Février	9605	214	22	22	116.2	57.7	54.0
	9606	215	36	58	116.2	55.3	54.4
	9607	216	34	92	116.2	60.1	54.8
	9608	217	36	128	116.2 **	63.0	55.3
Mars	9609	218	37	37	111.2	65.7	55.8
	9610	219	29	66	111.2	66.2	56.2
	9611	220	37	103	111.2	61.8	56.4
	9612	221	72	175	111.2 **	63.0 **	56.6 **
	9613	222	52	227	111.2 **	59.4	56.7
Avril	9614	223	72	72	116.2	63.7 **	56.7 **
	9615	224	55	127	116.2 **	62.5	56.7
	9616	225	55	182	116.2 **	62.2	56.6
	9617	226	52	234	116.2 **	59.6	56.6
Mai	9618	227	30	30	130.7	60.9	56.7
	9619	228	37	67	130.7	66.2	56.9
	9620	229	44	111	130.7	73.2	57.3
	9621	230	29	140	130.7 **	70.1	57.9
Juin	9622	231	34	34	159.7	81.1	58.8
	9623	232	58	92	159.7	78.3	59.9
	9624	233	64	156	159.7	73.6	61.3 **
	9625	234	44	200	159.7 **	74.7	63.0
	9626	235	43	243	159.7 **	75.6	64.9
Juillet	9627	236	43	43	198.4	73.4	67.0
	9628	237	45	88	198.4	91.0	69.2
	9629	238	47	135	198.4	89.1	71.5
	9630	239	44	179	198.4	99.2	73.8
Août	9631	240	48	48	246.7	90.7	76.0
	9632	241	64	112	246.7	79.4	78.0
	9633	242	28	140	246.7	112.3	79.8
	9634	243	69	209	246.7	104.0	81.2
	9635	244	72	281	246.7 **	104.0	82.3
Septembre	9636	245	64	64	261.2	109.0	82.9
	9637	246	78	142	261.2	123.1	83.1
	9638	247	88	230	261.2	107.5	82.9 **
	9639	248	113	343	261.2 **	101.4 **	82.2 **
Octobre	9640	249	110	110	251.6	103.5 **	81.0 **
	9641	250	67	177	251.6	106.5	79.6
	9642	251	53	230	251.6	101.5	77.8
	9643	252	47	277	251.6 **	90.7	75.7
Novembre	9644	253	36	36	198.4	96.3	73.5
	9645	254	45	81	198.4	103.6	71.2
	9646	255	63	144	198.4	112.6	69.0
	9647	256	37	181	198.4	104.9	66.8
	9648	257	42	223	198.4 **	89.4	64.8
Décembre	9649	258	58	58	164.5	81.5	63.0
	9650	259	46	104	164.5	77.6	61.4
	9651	260	46	150	164.5	81.9	60.2
	9652	261	30	180	164.5 **	80.8	59.3

* : semaines épidémiques définies par l'investigation

** : semaines épidémiques détectées par le seuil

! : alerte déclenchée par le CNR

(*) : semaines épidémiques supplémentaires en référence "corrigée"

Mois	Année - Semaine	lag	Observés	Cumul des semaines /mois	Seuil CNR	Seuil Box & J	Seuil Serfling
Janvier	9701	262	33	33	137.9	58.2	58.6
	9702	263	53	86	137.9	62.1	58.3
	9703	264	29	115	137.9	60.6	58.2
	9704	265	50	165	137.9 **	64.5	58.4
Février	9705	266	42	42	137.9	54.8	58.7
	9706	267	26	68	137.9	57.1	59.1
	9707	268	40	108	137.9	59.4	59.6
	9708	269	27	135	137.9	61.6	60.1
Mars	9709	270	30	30	135.5	63.3	60.5
	9710	271	44	74	135.5	60.6	60.9
	9711	272	56	130	135.5	61.2	61.2
	9712	273	55	185	135.5 **	72.2	61.4
	9713	274	42	227	135.5 **	63.8	61.4
Avril	9714	275	41	41	140.3	72.6	61.5
	9715	276	51	92	140.3	66.4	61.4
	9716	277	43	135	140.3	65.8	61.4
	9717	278	30	165	140.3 **	63.9	61.3
Mai	9718	279	49	49	154.8	56.8	61.4
	9719	280	34	83	154.8	63.0	61.6
	* 9720	281	38	121	154.8	69.0	62.0
	* 9721	282	53	174	154.8 **	62.7	62.6
	* 9722	283	40	214	154.8 **	72.6	63.5
Juin	* 9723	284	46	46	186.2	78.0	64.6
	* 9724	285	65	111	186.2	76.8	66.0
	* 9725	286	68	179	186.2	71.2	67.7 **
	* 9726	287	73	252	186.2 **	70.9 **	69.6 **
Juillet	* 9727	288	73	73	229.8	68.8 **	71.7 **
	9728	289	43	116	229.8	81.9	74.0
	9729	290	70	186	229.8	80.6	76.3
	9730	291	72	258	229.8 **	85.0	78.5
Août	9731	292	52	52	280.7	81.7	80.7
	9732	293	85	137	280.7	80.8 **	82.8 **
	9733	294	45	182	280.7	88.7	84.5
	9734	295	72	254	280.7	96.2	86.0
	9735	296	69	323	280.7 **	98.2	87.0
Septembre	9736	297	87	87	300.0	98.7	87.7
	9737	298	90	177	300.0	111.6	87.9 **
	9738	299	98	275	300.0	104.7	87.6 **
	9739	300	116	391	300.0 **	109.7 **	86.9 **
Octobre	9740	301	76	76	290.3	108.2	85.8
	9741	302	87	163	290.3	96.8	84.3 **
	9742	303	74	237	290.3	89.7	82.5
	9743	304	67	304	290.3 **	78.3	80.4
Novembre	9744	305	71	71	237.1	81.8	78.2
	9745	306	37	108	237.1	86.6	76.0
	9746	307	93	201	237.1	99.2	73.7 **
	9747	308	63	264	237.1 **	85.5	71.5
	9748	309	53	317	237.1 **	77.8	69.5
Décembre	9749	310	40	40	200.8	78.7	67.7
	9750	311	55	95	200.8	72.0	66.2
	9751	312	43	138	200.8	73.6	64.9
	9752	313	43	181	200.8	69.9	64.0

* : semaines épidémiques définies par l'investigation

** : semaines épidémiques détectées par le seuil

! : alerte déclenchée par le CNR

(*) : semaines épidémiques supplémentaires en référence "corrigée"

Mois	Année - Semaine	lag	Observés	Cumul des semaines /mois	Seuil CNR	Seuil Box & J	Seuil Serfling
Janvier	9401 106		0	0	9.7		4.9
	9402 107		0	0	9.7		4.7
	9403 108		1	1	9.7		4.5
	9404 109		1	2	9.7		4.4
	9405 110		2	4	9.7		4.3
Février	9406 111		0	0	10.3		4.4
	9407 112		2	2	10.3		4.5
	9408 113		0	2	10.3		4.6
	9409 114		2	4	10.3		4.7
Mars	9410 115		5	5	9.7		4.8 **
	9411 116		3	8	9.7		4.8
	9412 117		3	11	9.7 **		4.8
	9413 118		2	13	9.7 **		4.8
Avril	9414 119		3	3	9.4		4.7
	9415 120		5	8	9.4		4.6 **
	9416 121		0	8	9.4		4.6
	9417 122		3	11	9.4 **		4.6
	9418 123		1	12	9.4 **		4.7
Mai	9419 124		3	3	9.0		4.8
	9420 125		2	5	9.0		5.0
	9421 126		6	11	9.0 **		5.3 **
	9422 127		4	15	9.0 **		5.6
Juin	* 9423 128		4	4	11.0		6.0
	* 9424 129		23	27	! 11.0		6.3 **
	* 9425 130		30	57	11.0 **		6.6 **
	* 9426 131		20	77	11.0 **		6.8 **
Juillet	* 9427 132		18	18	15.2 **		7.0 **
	* 9428 133		18	36	15.2 **		7.0 **
	* 9429 134		13	49	15.2 **		7.0 **
	* 9430 135		22	71	15.2 **		7.0 **
	* 9431 136		10	81	15.2 **		6.9 **
Août	9432 137		3	3	21.0		6.7
	9433 138		8	11	21.0		6.6 **
	9434 139		5	16	21.0		6.5
	9435 140		4	20	21.0		6.5
Septembre	9436 141		6	6	22.6		6.5
	9437 142		4	10	22.6		6.5
	9438 143		9	19	22.6		6.6 **
	9439 144		9	28	22.6 **		6.7 **
Octobre	9440 145		7	7	20.3		6.8 **
	9441 146		2	9	20.3		6.9
	9442 147		8	17	20.3		6.9 **
	9443 148		5	22	20.3 **		6.8
	9444 149		6	28	20.3 **		6.7
Novembre	9445 150		4	4	14.5		6.5
	9446 151		4	8	14.5		6.2
	9447 152		2	10	14.5		6.0
	9448 153		1	11	14.5		5.7
Décembre	9449 154		0	0	9.4		5.4
	9450 155		1	1	9.4		5.2
	9451 156		4	5	9.4		5.1
	9452 157		2	7	9.4		5.0

* : semaines épidémiques définies par l'investigation

** : semaines épidémiques détectées par le seuil

! : alerte déclenchée par le CNR

(*) : semaines épidémiques supplémentaires en référence "corrigée"

Mois	Année - Semaine	lag	Observés	Cumul des semaines /mois	Seuil CNR	Seuil Box & J	Seuil Serfling
Janvier	9501	158	0	0	8.4		5.0
	9502	159	2	2	8.4		5.1
	9503	160	3	5	8.4		5.1
	9504	161	1	6	8.4		5.2
	9505	162	0	6	8.4		5.3
Février	9506	163	1	1	9.0		5.3
	9507	164	2	3	9.0		5.2
	9508	165	2	5	9.0		5.1
	9509	166	6	11	9.0 **		5.0 **
Mars	9510	167	2	2	9.7		4.8
	9511	168	3	5	9.7		4.6
	9512	169	1	6	9.7		4.4
	9513	170	1	7	9.7		4.3
Avril	9514	171	2	2	10.0		4.2
	9515	172	2	4	10.0		4.2
	9516	173	0	4	10.0		4.3
	9517	174	1	5	10.0		4.4
	9518	175	2	7	10.0		4.7
Mai	9519	176	6	6	11.3		4.9 **
	9520	177	1	7	11.3		5.2
	9521	178	0	7	11.3		5.5
	9522	179	4	11	11.3		5.7
Juin	9523	180	4	4	14.5		5.9
	9524	181	1	5	14.5		6.1
	9525	182	5	10	14.5		6.1
	9526	183	5	15	14.5 **		6.2
Juillet	9527	184	3	3	21.0		6.2
	9528	185	1	4	21.0		6.2
	9529	186	3	7	21.0		6.2
	9530	187	4	11	21.0		6.2
	9531	188	5	16	21.0		6.3
Août	9532	189	5	5	24.5		6.5
	9533	190	7	12	24.5		6.7 **
	9534	191	4	16	24.5		6.9
	9535	192	2	18	24.5		7.1
Septembre	9536	193	9	9	24.5		7.3 **
	9537	194	2	11	24.5		7.5
	9538	195	1	12	24.5		7.5
	9539	196	3	15	24.5		7.5
	9540	197	1	16	24.5		7.4
Octobre	9541	198	3	3	20.3		7.2
	9542	199	1	4	20.3		6.9
	9543	200	7	11	20.3		6.6 **
	9544	201	6	17	20.3		6.3
Novembre	9545	202	3	3	14.8		6.0
	9546	203	2	5	14.8		5.7
	9547	204	3	8	14.8		5.5
	9548	205	2	10	14.8		5.4
Décembre	9549	206	1	1	9.0		5.3
	9550	207	3	4	9.0		5.2
	9551	208	2	6	9.0		5.2
	9552	209	3	9	9.0		5.2

* : semaines épidémiques définies par l'investigation

** : semaines épidémiques détectées par le seuil

! : alerte déclenchée par le CNR

(*) : semaines épidémiques supplémentaires en référence "corrigée"

Mois	Année - Semaine	lag	Observés	Cumul des semaines /mois	Seuil CNR	Seuil Box & J	Seuil Serfling
Janvier	9401	106	2	2	4.6		3.1
	9402	107	5	7	4.6 **		3.0 **
	9403	108	1	8	4.6 **		2.9
	9404	109	0	8	4.6 **		2.9
	9405	110	0	8	4.6 **		2.8
Février	9406	111	0	0	4.8		2.7
	9407	112	0	0	4.8		2.6
	9408	113	0	0	4.8		2.6
	9409	114	0	0	4.8		2.6
Mars	9410	115	1	1	5.1		2.7
	9411	116	1	2	5.1		2.8
	9412	117	2	4	5.1		2.8
	9413	118	2	6	5.1 **		2.9
Avril	9414	119	0	0	4.8		3.0
	9415	120	2	2	4.8		3.1
	9416	121	0	0	4.8		3.1
	9417	122	4	6	4.8 **		3.1 **
	9418	123	0	6	4.8 **		3.1
Mai	9419	124	0	0	5.1		3.0
	9420	125	0	0	5.1		2.9
	9421	126	0	0	5.1		2.8
	9422	127	3	3	5.1		2.8 **
Juin	9423	128	5	5	5.8		2.7 **
	9424	129	1	6	5.8 **		2.7
	9425	130	0	6	5.8 **		2.7
	* 9426	131	0	6	5.8 **		2.8
Juillet	* 9427	132	5	5	6.5		2.9 **
	* 9428	133	3	8	6.5 **		3.0
	* 9429	134	3	11	6.5 **		3.1
	* 9430	135	7	18	6.5 **		3.3 **
	* 9431	136	16	34	6.5 **		3.4 **
Août	* 9432	137	9	9	9.7		3.5 **
	* 9433	138	19	28	! 9.7		3.6 **
	* 9434	139	8	36	9.7 **		3.6 **
	* 9435	140	10	46	9.7 **		3.6 **
Septembre	* 9436	141	13	13	9.7 **		3.5 **
	* 9437	142	15	28	9.7 **		3.5 **
	* 9438	143	15	43	9.7 **		3.5 **
	* 9439	144	2	45	9.7 **		3.4
Octobre	9440	145	2	2	9.7		3.4
	9441	146	2	4	9.7		3.4
	9442	147	1	5	9.7		3.4
	9443	148	0	5	9.7		3.4
	9444	149	2	7	9.7		3.4
Novembre	9445	150	3	3	6.3		3.4
	9446	151	1	4	6.3		3.4
	9447	152	5	9	6.3 **		3.4 **
	9448	153	1	10	6.3 **		3.3
Décembre	9449	154	0	0	3.9		3.2
	9450	155	0	0	3.9		3.1
	9451	156	0	0	3.9		3.0
	9452	157	1	1	3.9		2.9

* : semaines épidémiques définies par l'investigation

** : semaines épidémiques détectées par le seuil

! : alerte déclenchée par le CNR

(*) : semaines épidémiques supplémentaires en référence "corrigée"

Mois	Année - Semaine	lag	Observés	Cumul des semaines /mois	Seuil CNR	Seuil Box & J	Seuil Serfling
Janvier	9501	158	1	1	3.4		2.8
	9502	159	0	1	3.4		2.7
	9503	160	1	2	3.4		2.6
	9504	161	0	2	3.4		2.5
	9505	162	1	3	3.4		2.5
Février	9506	163	0	0	4.4		2.5
	9507	164	2	2	4.4		2.6
	9508	165	0	0	4.4		2.7
	9509	166	1	1	4.4		2.8
Mars	9510	167	2	2	4.1		2.9
	9511	168	0	2	4.1		2.9
	9512	169	0	2	4.1		3.0
	9513	170	2	2	4.1		3.0
Avril	9514	171	0	0	3.9		3.0
	9515	172	1	1	3.9		2.9
	9516	173	0	1	3.9		2.8
	9517	174	0	1	3.9		2.8
	9518	175	0	1	3.9		2.7
Mai	9519	176	2	2	4.4		2.6
	9520	177	1	3	4.4		2.6
	9521	178	0	3	4.4		2.6
	9522	179	1	4	4.4		2.6
Juin	9523	180	0	0	5.3		2.7
	9524	181	0	0	5.3		2.8
	9525	182	1	1	5.3		2.9
	9526	183	1	2	5.3		3.0
Juillet	9527	184	2	2	7.5		3.1
	9528	185	0	2	7.5		3.2
	9529	186	1	3	7.5		3.2
	9530	187	1	4	7.5		3.2
	9531	188	1	5	7.5		3.2
Août	9532	189	0	0	11.4		3.2
	9533	190	4	4	11.4		3.2 **
	9534	191	2	6	11.4		3.2
	9535	192	0	6	11.4		3.2
Septembre	9536	193	1	1	10.9		3.2
	9537	194	4	5	10.9		3.3 **
	9538	195	2	7	10.9		3.3
	9539	196	0	7	10.9		3.4
	9540	197	2	9	10.9		3.4
Octobre	9541	198	2	2	9.9		3.5
	9542	199	1	3	9.9		3.5
	9543	200	0	3	9.9		3.5
	9544	201	4	7	9.9		3.5 **
Novembre	9545	202	0	0	4.6		3.5
	9546	203	4	4	4.6		3.4 **
	9547	204	2	6	4.6 **		3.2
	9548	205	3	9	4.6 **		3.1
Décembre	9549	206	2	2	3.6		2.9
	9550	207	1	3	3.6		2.8
	9551	208	1	4	3.6 **		2.7
	9552	209	0	4	3.6 **		2.6

* : semaines épidémiques définies par l'investigation

** : semaines épidémiques détectées par le seuil

! : alerte déclenchée par le CNR

(*) : semaines épidémiques supplémentaires en référence "corrigée"

Mois	Année - Semaine	lag	Observés	Cumul des semaines /mois	Seuil CNR	Seuil Box & J	Seuil Serfling
Janvier	9401	106	0	0	8.9		7.8
	9402	107	2	2	8.9		7.8
	9403	108	3	5	8.9		7.8
	9404	109	2	7	8.9		7.7
	9405	110	3	10	8.9 **		7.6
Février	9406	111	4	4	8.1		7.6
	9407	112	2	6	8.1		7.5
	9408	113	1	7	8.1		7.5
	9409	114	4	11	8.1 **		7.5
Mars	9410	115	6	6	8.9		7.5
	9411	116	4	10	8.9 **		7.5
	9412	117	1	11	8.9 **		7.5
	9413	118	4	15	8.9 **		7.6
Avril	9414	119	1	1	10.5		7.8
	9415	120	5	6	10.5		8.0
	9416	121	4	10	10.5		8.2
	9417	122	3	13	10.5 **		8.4
	9418	123	4	17	10.5 **		8.7
Mai	9419	124	3	3	12.1		9.1
	9420	125	3	6	12.1		9.4
	9421	126	6	12	12.1		9.8
	9422	127	3	15	12.1 **		10.1
Juin	9423	128	6	6	12.1		10.5
	9424	129	3	9	12.1		10.8
	9425	130	0	9	12.1		11.1
	9426	131	8	17	12.1 **		11.4
Juillet	9427	132	6	6	13.7		11.6
	9428	133	7	13	13.7		11.8
	9429	134	10	23	13.7 **		11.9
	9430	135	6	29	13.7 **		11.9
	9431	136	5	34	13.7 **		11.9
Août	9432	137	12	12	16.9		11.9 **
	9433	138	8	20	16.9 **		11.8
	9434	139	11	31	16.9 **		11.6
	9435	140	15	46	16.9 **		11.4 **
Septembre	9436	141	10	10	20.2		11.2
	9437	142	7	17	20.2		10.9
	9438	143	4	21	20.2 **		10.7
	9439	144	5	26	20.2 **		10.4
Octobre	9440	145	4	4	19.4		10.1
	9441	146	7	11	19.4		9.9
	9442	147	8	19	19.4		9.6
	9443	148	1	20	19.4 **		9.4
	9444	149	5	25	19.4 **		9.3
Novembre	9445	150	3	3	16.9		9.1
	9446	151	4	7	16.9		9.1
	9447	152	2	9	16.9		9.0
	9448	153	1	10	16.9		9.0
Décembre	9449	154	1	1	11.3		9.0
	9450	155	3	4	11.3		9.1
	9451	156	3	7	11.3		9.2
	9452	157	9	16	11.3 **		9.2

* : semaines épidémiques définies par l'investigation

** : semaines épidémiques détectées par le seuil

! : alerte déclenchée par le CNR

(*) : semaines épidémiques supplémentaires en référence "corrigée"

Mois	Année - Semaine	lag	Observés	Cumul des semaines /mois	Seuil CNR	Seuil Box & J	Seuil Serfling
Janvier	9501	158	9	9	9.7		9.4
	9502	159	4	13	9.7 **		9.5
	9503	160	4	17	9.7 **		9.6
	9504	161	1	18	9.7 **		9.7
	9505	162	6	24	9.7 **		9.8
Février	9506	163	4	4	8.9		9.9
	9507	164	0	4	8.9		10.0
	9508	165	3	7	8.9		10.0
	9509	166	1	8	8.9		10.1
Mars	9510	167	5	5	10.5		10.2
	9511	168	1	6	10.5		10.2
	9512	169	0	6	10.5		10.3
	9513	170	2	8	10.5		10.4
Avril	9514	171	1	1	12.9		10.5
	9515	172	5	6	12.9		10.6
	9516	173	3	9	12.9		10.7
	9517	174	3	12	12.9		10.9
	9518	175	1	13	12.9 **		11.1
Mai	9519	176	6	6	14.5		11.3
	9520	177	7	13	14.5		11.6
	9521	178	2	15	14.5 **		11.9
	9522	179	3	18	14.5 **		12.2
Juin	* 9523	180	4	4	16.1		12.5
	* 9524	181	7	11	16.1		12.9
	* 9525	182	8	19	16.1 **		13.2
	* 9526	183	10	29	16.1 **		13.5
Juillet	* 9527	184	15	15	18.6		13.9 **
	* 9528	185	9	24	18.6 **		14.2
	* 9529	186	14	38	18.6 **		14.4
	* 9530	187	20	58	18.6 **		14.6 **
	* 9531	188	18	76	18.6 **		14.8 **
Août	* 9532	189	2	2	! 23.39 **		14.9
	* 9533	190	25	27	23.4 **		15.0 **
	* 9534	191	14	41	23.4 **		15.0
	* 9535	192	20	61	23.4 **		15.0 **
Septembre	* 9536	193	8	8	23.4		14.9
	* 9537	194	7	15	23.4		14.8
	9538	195	5	20	23.4		14.6
	9539	196	8	28	23.4 **		14.4
	9540	197	8	36	23.4 **		14.1
Octobre	9541	198	5	5	20.2		13.9
	9542	199	6	11	20.2		13.6
	9543	200	6	17	20.2		13.4
	9544	201	14	31	20.2 **		13.2 **
Novembre	9545	202	11	11	16.1		13.0
	9546	203	18	29	16.1 **		12.8 **
	9547	204	7	36	16.1 **		12.7
	9548	205	5	41	16.1 **		12.6
Décembre	9549	206	2	2	12.9		12.6
	9550	207	6	8	12.9		12.6
	9551	208	8	16	12.9 **		12.7
	9552	209	5	21	12.9 **		12.8

* : semaines épidémiques définies par l'investigation

** : semaines épidémiques détectées par le seuil

! : alerte déclenchée par le CNR

(*) : semaines épidémiques supplémentaires en référence "corrigée"

