

# Place des outils d'analyse des séries temporelles dans la surveillance épidémiologique pour la détection des épidémies et leur analyse

Thèse présentée pour obtenir le grade de docteur  
d'Aix-Marseille Université

AMU, Site de la Timone – 17 décembre 2018

Gabriel Bédubourg, CESP/SESSTIM

Sous la direction du Pr Jean-Baptiste Meynard  
et du Dr Aurélie Mayet

# Plan

1. Introduction générale
2. Objectifs de la thèse
3. Méthode de simulation des épidémies
4. Evaluation des méthodes de détection des épidémies
5. Combinaison de méthodes statistiques pour la détection
6. Analyse étiologique des épidémies
7. Perspectives et conclusion

# ○ — Surveillance épidémiologique

" ...enregistrement continu de l'incidence d'événements de santé dans une population, l'analyse de ces données et leur diffusion..."

## Objectifs

- Décrire l'importance épidémiologique & tendances
- **Détecter les événements anormaux** ⇒ action
- Evaluer les actions de santé publique
- Définir des axes de recherche

# ○ — Données de surveillance : nature

## Complexité variable

- Dénombrement de cas ou d'événements (MEH)
- Description clinico-épidémiologique des cas (DO, FSD)



# Données de surveillance : propriétés

## Données indexées par le temps

- Représentation sous forme de séries temporelles
- Les analyses nécessitent des outils adaptés
- Souvent limitées à des objectifs descriptifs

# ○ — Evolution de la surveillance

## Auparavant

- Nombre limité d'événements surveillés
- Systèmes humains
- Analyse reposant sur l'épidémiologiste

## Maintenant : multiplication des sources

- Origine multiple des données
- Bases de données de grande taille
- Automatisation des analyses

# ○ — Détection des événements anormaux

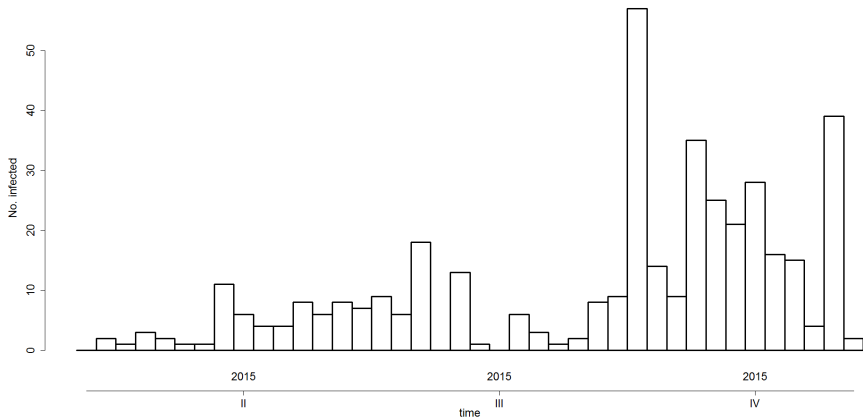
Evénement anormal = épidémie

**"...nombre de cas observé > nombre de cas attendu..."**



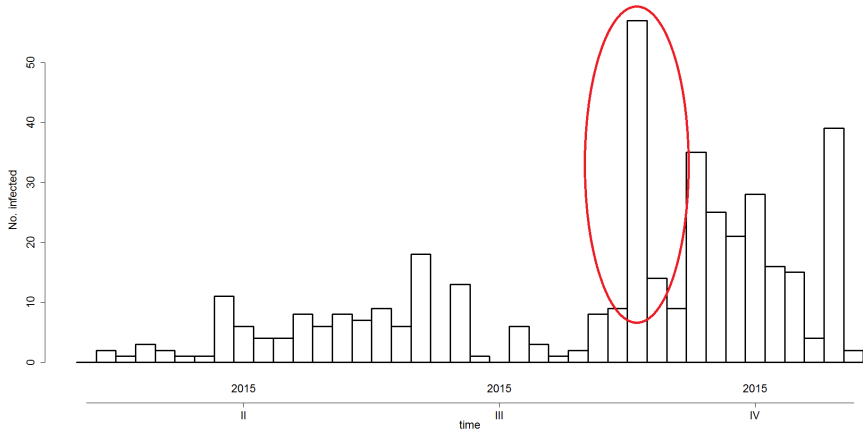
# Série temporelle

Surveillance des diarrhées, Unité Barkhane - semaines 8 à 48/2015



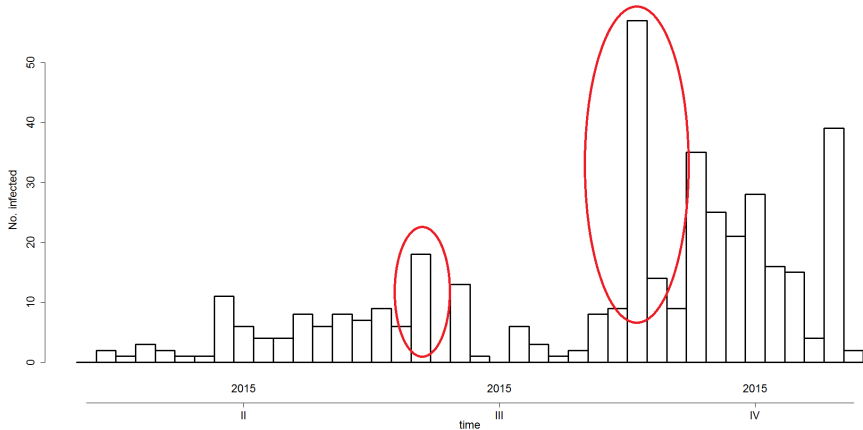
# Série temporelle

Surveillance des diarrhées, Unité Barkhane - semaines 8 à 48/2015



# Série temporelle

Surveillance des diarrhées, Unité Barkhane - semaines 8 à 48/2015





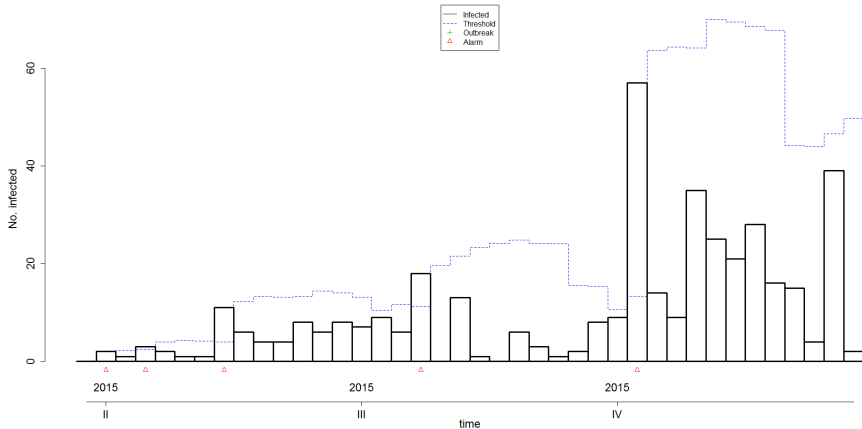
# ○ — Analyse des données de surveillance

## Détection des épidémies

- Analyse graphique
- Le plus souvent : utilisation d'outils statistiques
- Pas de réelle évaluation des outils statistiques

# Série temporelle

Surveillance des diarrhées, Unité Barkhane - semaines 8 à 48/2015



# ○ — Analyse des données de surveillance

## Une fois l'épidémie détectée → contre-mesures

- Le plus souvent, guidées par une enquête épidémiologique
- Pas ou peu d'exploitation des données déjà collectées
- Données de surveillance peuvent-elles orienter l'action ?

# Plan

1. Introduction générale
2. Objectifs de la thèse
3. Méthode de simulation des épidémies
4. Evaluation des méthodes de détection des épidémies
5. Combinaison de méthodes statistiques pour la détection
6. Analyse étiologique des épidémies
7. Perspectives et conclusion



# Objectifs de la thèse

## Axe 1 – Détection des épidémies

- Evaluer les principales méthodes statistiques de détection des épidémies publiées
  - Proposer une nouvelle approche reposant sur la combinaison optimale de méthodes de détection
- Etudes de simulation

## Axe 2 – Analyse étiologique des épidémies

- Développer une nouvelle méthode statistique d'analyse étiologique d'une épidémie à partir des données de surveillance épidémiologique

# Plan

1. Introduction générale
2. Objectifs de la thèse
- 3. Méthode de simulation des épidémies**
4. Evaluation des méthodes de détection des épidémies
5. Combinaison de méthodes statistiques pour la détection
6. Analyse étiologique des épidémies
7. Perspectives et conclusion

## Research Article

Received 16 May 2012,

Accepted 6 August 2012

Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5595

# An improved algorithm for outbreak detection in multiple surveillance systems<sup>‡</sup>

Angela Noufaily,<sup>a\*†</sup> Doyo G. Enki,<sup>a</sup> Paddy Farrington,<sup>a</sup>  
Paul Garthwaite,<sup>a</sup> Nick Andrews<sup>b</sup> and André Charlett<sup>b</sup>

# Structure des données simulées

## 4200 séries : 42 scénarios simulés 100 fois

Selon un modèle binomial négatif de moy.  $\mu$  et de var.  $\phi\mu$

$$\mu_t = \exp \left( \theta + \beta t + \sum_{j=0}^2 \left\{ \gamma_1 \cos \left( \frac{2\pi jt}{52} \right) + \gamma_2 \sin \left( \frac{2\pi jt}{52} \right) \right\} \right)$$

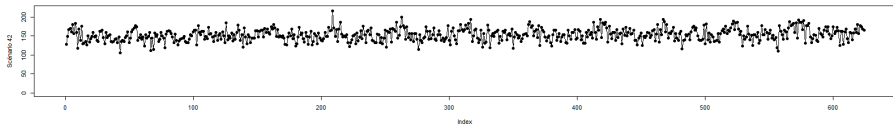
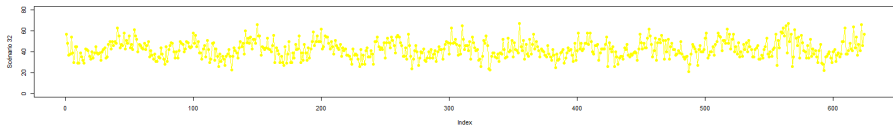
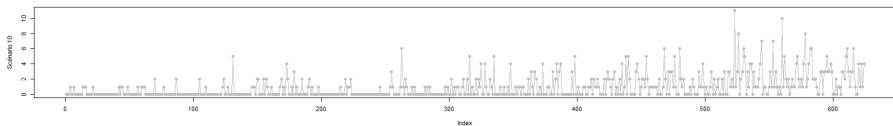
## Séries de 624 semaines

- Sem. 0 à 312 : sem. d'exercice (*training weeks*)
- Sem. 313 à 575 : sem. de référence (*baseline weeks*)
- Sem. 576 à 624 : sem. courantes (*current weeks*)

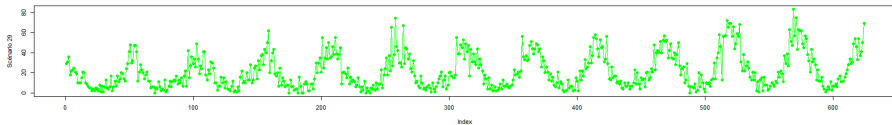
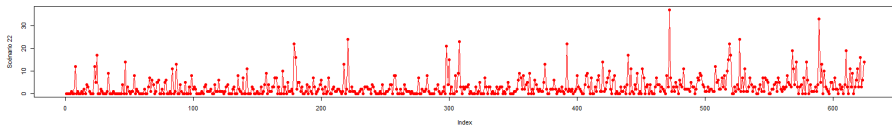
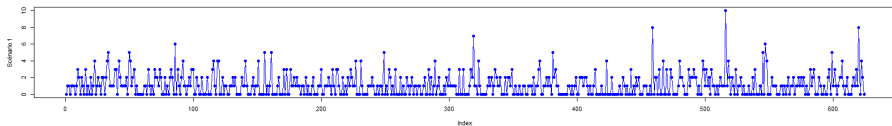
**Table I.** Parameters and criteria used to generate the 42 scenarios.

Scenario	$\theta$	$\beta$	$\gamma_1$	$\gamma_2$	$\phi$	$m$	Trend
1	0.1	0	0	0	1.5	0	0
2	0.1	0	0.6	0.6	1.5	1	0
3	0.1	0	0.6	0.6	1.5	2	0
4	0.1	0.0025	0	0	1.5	0	1
5	0.1	0.0025	0.6	0.6	1.5	1	1
6	0.1	0.0025	0.6	0.6	1.5	2	1
7	-2	0	0	0	2	0	0
8	-2	0	0.1	0.3	2	1	0
9	-2	0	0.1	0.3	2	2	0
10	-2	0.005	0	0	2	0	1
11	-2	0.005	0.1	0.3	2	1	1
12	-2	0.005	0.1	0.3	2	2	1
13	1.5	0	0	0	1	0	0
14	1.5	0	0.2	-0.4	1	1	0
15	1.5	0	0.2	-0.4	1	2	0
16	1.5	0.003	0	0	1	0	1
17	1.5	0.003	0.2	-0.4	1	1	1
18	1.5	0.003	0.2	-0.4	1	2	1
19	0.5	0	0	0	5	0	0
20	0.5	0	0.5	0.5	5	1	0
21	0.5	0	0.5	0.5	5	2	0
22	0.5	0.002	0	0	5	0	1
23	0.5	0.002	0.5	0.5	5	1	1
24	0.5	0.002	0.5	0.5	5	2	1
25	2.5	0	0	0	3	0	0
26	2.5	0	1	0.1	3	1	0
27	2.5	0	1	0.1	3	2	0
28	2.5	0.001	0	0	3	0	1
29	2.5	0.001	1	0.1	3	1	1
30	2.5	0.001	1	0.1	3	2	1
31	3.75	0	0	0	1.1	0	0
32	3.75	0	0.1	-0.1	1.1	1	0
33	3.75	0	0.1	-0.1	1.1	2	0
34	3.75	0.001	0	0	1.1	0	1
35	3.75	0.001	0.1	-0.1	1.1	1	1
36	3.75	0.001	0.1	-0.1	1.1	2	1
37	5	0	0	0	1.2	0	0
38	5	0	0.05	0.01	1.2	1	0
39	5	0	0.05	0.01	1.2	2	0
40	5	0.0001	0	0	1.2	0	1
41	5	0.0001	0.05	0.01	1.2	1	1
42	5	0.0001	0.05	0.01	1.2	2	1

# Grande variété de séries temporelles



# Proches des données réelles



# ○ — Simulation des épidémies

## 3 épidémies durant les *baseline weeks*

- Influence des épidémies du passé
- Importance croissante  $k_1 = 2, 3, 5$  ou  $10$
- Sans épidémies dans les semaines courantes


⇒ 5 jeux de 4200 séries

## Puis, 1 épidémie sur les *current weeks*

- Amplitudes croissantes  $k_2 = 1$  à  $10$
- Pour chaque valeur de  $k_1$

⇒ 5 jeux de 42000 séries





S0 à S312

*Training Weeks*

Pas d'épidémie

S313 à S575

*Baseline Weeks*

Epidémies d'amplitudes  
 $k_1 = 0, 2, 3, 5$  ou  $10$

=> 5 jeux de 4200 séries



624 semaines

S0 à S312

*Training Weeks*

Pas d'épidémie

S313 à S575

*Baseline Weeks*

Epidémies d'amplitudes  
 $k_1 = 0, 2, 3, 5$  ou  $10$

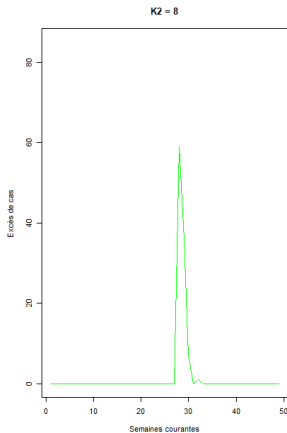
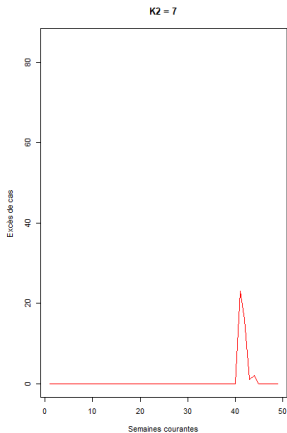
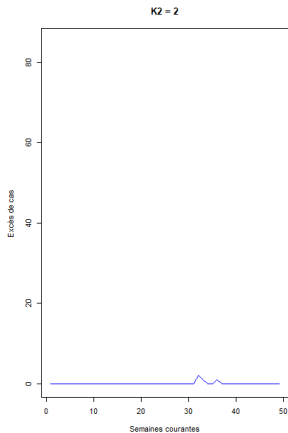
S576 à S624

*Current Weeks*

Epidémies d'amplitudes  
 $k_2 = 1$  à  $10$

624 semaines

# Exemples d'épidémies simulées



# Plan

1. Introduction générale
2. Objectifs de la thèse
3. Méthode de simulation des épidémies
- 4. Evaluation des méthodes de détection des épidémies**
5. Combinaison de méthodes statistiques pour la détection
6. Analyse étiologique des épidémies
7. Perspectives et conclusion



# Introduction

# Contexte

## Mutations de la surveillance

- Augmentation des items monitorés

## Automatisation

- Surveillance assistée par ordinateur
- Analyse automatisée des données
- **Nécessité de standardisation des outils de détection**

# ○ — Quelles performances attendues ?

## Couple sensibilité - spécificité élevé

- Trop peu sensible : "oublie" des épidémies
- Trop peu spécifique : trop de fausses alarmes

## Quel cadre pour l'évaluation des méthodes ?

- Représentatif de la diversité des données
- Reproductible
- Disponible
- → Etudes de simulation

# Objectifs

## Réaliser une évaluation objective des méthodes disponibles

- A partir de données simulées
- Dans un contexte opérationnel, *i. e.* sans paramétrage complexe

⇒ Déterminer quelle méthode est la plus adaptée aux objectifs de la surveillance

⇒ Séries temporelles hebdomadaires avec données historiques





# Matériel & méthodes

# Principes de l'étude

## Sur la base du logiciel R

- Package Surveillance
- Nombreuses méthodes de détection déployées
- **Objectif opérationnel**  $\Rightarrow$  "réglage" unique
- Prendre en compte les épidémies du passé

$\Rightarrow$  **Application à des données simulées** ( $\alpha = 0.01, 0.05$  et  $0.001$ )

# ○ — Méthodes testées 1

## Limites historiques

- *Current past graph* (méthode CDC)
- RKI 1 à 3
- Bayes 1 à 3

## Les Cusum

- Cusum, Cusum Rossi, Cusum GLM, Cusum GLM Rossi
- EARS C1 à C3

# ○ — Méthodes testées 2

## Modélisation

- Farrington "amélioré" et "historique"
- OutbreakP
- GLM périodique Poisson et NB (méthodes "maison")
- GLR *Negative Binomial*
- GLR Poisson

# Indicateurs de performance

## Article *princeps*

- FPR : taux de faux positifs (séries sans épidémie)
- POD : Probabilité de détection ( $\geq 1$  alarme/épidémie)

## Mais aussi

- Proba. de détection en 1<sup>ère</sup> semaine (POD1week)
- Sensibilité et Spécificité des méthodes
- Valeurs prédictives positives et négatives
- $F_1$  – score



# Résultats

# Performances des méthodes

Method	FPR	Specificity	POD	POD1week	Sensitivity	NPV	PPV	$F_1$ -measure
Improved Farrington	1.0%	99.0%	43.3%	34.0%	20.5%	95.0%	58.3%	0.30
Original Farrington	2.3%	97.7%	56.9%	45.5%	29.0%	95.4%	45.0%	0.35
Periodic Poisson GLM	3.3%	96.8%	67.8%	56.6%	35.6%	95.8%	42.3%	0.39
Periodic NegBin GLM	0.7%	99.4%	44.8%	36.3%	20.7%	95.0%	68.4%	0.32
CDC	3.6%	95.5%	45.0%	18.7%	34.2%	95.6%	33.2%	0.34
CUSUM	44.0%	52.7%	80.5%	70.5%	75.4%	97.0%	9.5%	0.17
CUSUM Rossi	39.5%	57.6%	77.0%	65.9%	71.8%	96.9%	10.1%	0.18
CUSUM GLM	44.2%	52.0%	84.4%	73.8%	79.5%	97.5%	9.9%	0.18
CUSUM GLM Rossi	39.9%	56.8%	81.1%	69.5%	76.1%	97.3%	10.4%	0.18
Bayes 1 ( $\alpha = 0.05$ )	10.1%	90.5%	76.2%	66.2%	39.1%	95.7%	21.4%	0.28
Bayes 2 ( $\alpha = 0.05$ )	9.4%	91.0%	80.8%	69.4%	45.7%	96.2%	25.0%	0.32
Bayes 3 ( $\alpha = 0.05$ )	11.1%	88.9%	83.4%	71.9%	51.8%	96.5%	23.6%	0.32
RKI 1	8.3%	92.3%	67.8%	58.9%	30.4%	95.3%	20.6%	0.25
RKI 2	5.5%	94.7%	67.8%	57.8%	34.5%	95.6%	30.0%	0.32
RKI 3	7.0%	93.0%	71.3%	60.6%	41.8%	96.0%	28.3%	0.34
GLR NegBin	4.3%	95.7%	50.8%	29.8%	21.6%	94.9%	24.9%	0.23
GLR Poisson	15.5%	84.5%	75.5%	60.3%	45.9%	95.9%	16.4%	0.24
EARS C1	6.9%	93.7%	66.3%	57.4%	25.6%	95.0%	21.2%	0.23
EARS C2	8.5%	92.4%	68.0%	57.1%	38.8%	95.8%	25.1%	0.31
EARS C3	7.4%	92.9%	54.2%	8.5%	35.3%	95.6%	24.7%	0.29
OutbreakP	59.9%	37.4%	70.4%	67.9%	66.1%	94.4%	6.5%	0.12

# Performances des méthodes

Method	FPR	Specificity	POD	POD1week	Sensitivity	NPV	PPV	$F_1$ -measure
Improved Farrington	1.0%	99.0%	43.3%	34.0%	20.5%	95.0%	58.3%	0.30
Original Farrington	2.3%	97.7%	56.9%	45.5%	29.0%	95.4%	45.0%	0.35
Periodic Poisson GLM	3.3%	96.8%	67.8%	56.6%	35.6%	95.8%	42.3%	0.39
Periodic NegBin GLM	0.7%	99.4%	44.8%	36.3%	20.7%	95.0%	68.4%	0.32
<b>CDC</b>	<b>3.6%</b>	<b>95.5%</b>	<b>45.0%</b>	<b>18.7%</b>	<b>34.2%</b>	<b>95.6%</b>	<b>33.2%</b>	<b>0.34</b>
CUSUM	44.0%	52.7%	80.5%	70.5%	75.4%	97.0%	9.5%	0.17
CUSUM Rossi	39.5%	57.6%	77.0%	65.9%	71.8%	96.9%	10.1%	0.18
CUSUM GLM	44.2%	52.0%	84.4%	73.8%	79.5%	97.5%	9.9%	0.18
CUSUM GLM Rossi	39.9%	56.8%	81.1%	69.5%	76.1%	97.3%	10.4%	0.18
Bayes 1 ( $\alpha = 0.05$ )	10.1%	90.5%	76.2%	66.2%	39.1%	95.7%	21.4%	0.28
Bayes 2 ( $\alpha = 0.05$ )	9.4%	91.0%	80.8%	69.4%	45.7%	96.2%	25.0%	0.32
Bayes 3 ( $\alpha = 0.05$ )	11.1%	88.9%	83.4%	71.9%	51.8%	96.5%	23.6%	0.32
RKI 1	8.3%	92.3%	67.8%	58.9%	30.4%	95.3%	20.6%	0.25
RKI 2	5.5%	94.7%	67.8%	57.8%	34.5%	95.6%	30.0%	0.32
RKI 3	7.0%	93.0%	71.3%	60.6%	41.8%	96.0%	28.3%	0.34
GLR NegBin	4.3%	95.7%	50.8%	29.8%	21.6%	94.9%	24.9%	0.23
GLR Poisson	15.5%	84.5%	75.5%	60.3%	45.9%	95.9%	16.4%	0.24
EARS C1	6.9%	93.7%	66.3%	57.4%	25.6%	95.0%	21.2%	0.23
EARS C2	8.5%	92.4%	68.0%	57.1%	38.8%	95.8%	25.1%	0.31
EARS C3	7.4%	92.9%	54.2%	8.5%	35.3%	95.6%	24.7%	0.29
OutbreakP	59.9%	37.4%	70.4%	67.9%	66.1%	94.4%	6.5%	0.12



# Performances des méthodes

Method	FPR	Specificity	POD	POD1week	Sensitivity	NPV	PPV	$F_1$ -measure
Improved Farrington	1.0%	99.0%	43.3%	34.0%	20.5%	95.0%	58.3%	0.30
Original Farrington	2.3%	97.7%	56.9%	45.5%	29.0%	95.4%	45.0%	0.35
Periodic Poisson GLM	3.3%	96.8%	67.8%	56.6%	35.6%	95.8%	42.3%	0.39
Periodic NegBin GLM	0.7%	99.4%	44.8%	36.3%	20.7%	95.0%	68.4%	0.32
CDC	3.6%	95.5%	45.0%	18.7%	34.2%	95.6%	33.2%	0.34
<b>CUSUM</b>	<b>44.0%</b>	<b>52.7%</b>	<b>80.5%</b>	<b>70.5%</b>	<b>75.4%</b>	<b>97.0%</b>	<b>9.5%</b>	<b>0.17</b>
CUSUM Rossi	39.5%	57.6%	77.0%	65.9%	71.8%	96.9%	10.1%	0.18
CUSUM GLM	44.2%	52.0%	84.4%	73.8%	79.5%	97.5%	9.9%	0.18
CUSUM GLM Rossi	39.9%	56.8%	81.1%	69.5%	76.1%	97.3%	10.4%	0.18
Bayes 1 ( $\alpha = 0.05$ )	10.1%	90.5%	76.2%	66.2%	39.1%	95.7%	21.4%	0.28
Bayes 2 ( $\alpha = 0.05$ )	9.4%	91.0%	80.8%	69.4%	45.7%	96.2%	25.0%	0.32
Bayes 3 ( $\alpha = 0.05$ )	11.1%	88.9%	83.4%	71.9%	51.8%	96.5%	23.6%	0.32
RKI 1	8.3%	92.3%	67.8%	58.9%	30.4%	95.3%	20.6%	0.25
RKI 2	5.5%	94.7%	67.8%	57.8%	34.5%	95.6%	30.0%	0.32
RKI 3	7.0%	93.0%	71.3%	60.6%	41.8%	96.0%	28.3%	0.34
GLR NegBin	4.3%	95.7%	50.8%	29.8%	21.6%	94.9%	24.9%	0.23
GLR Poisson	15.5%	84.5%	75.5%	60.3%	45.9%	95.9%	16.4%	0.24
EARS C1	6.9%	93.7%	66.3%	57.4%	25.6%	95.0%	21.2%	0.23
EARS C2	8.5%	92.4%	68.0%	57.1%	38.8%	95.8%	25.1%	0.31
EARS C3	7.4%	92.9%	54.2%	8.5%	35.3%	95.6%	24.7%	0.29
<b>OutbreakP</b>	<b>59.9%</b>	<b>37.4%</b>	<b>70.4%</b>	<b>67.9%</b>	<b>66.1%</b>	<b>94.4%</b>	<b>6.5%</b>	<b>0.12</b>

# Performances des méthodes

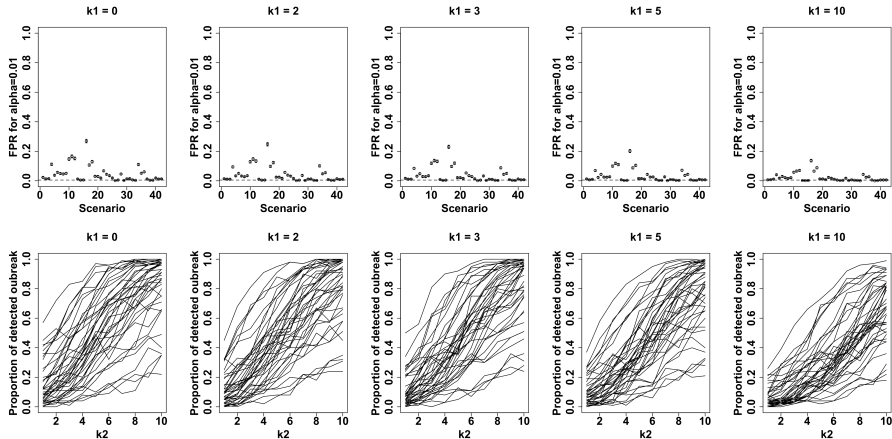
Method	FPR	Specificity	POD	POD1week	Sensitivity	NPV	PPV	$F_1$ -measure
Improved Farrington	1.0%	99.0%	43.3%	34.0%	20.5%	95.0%	58.3%	0.30
Original Farrington	2.3%	97.7%	56.9%	45.5%	29.0%	95.4%	45.0%	0.35
Periodic Poisson GLM	3.3%	96.8%	67.8%	56.6%	35.6%	95.8%	42.3%	0.39
Periodic NegBin GLM	0.7%	99.4%	44.8%	36.3%	20.7%	95.0%	68.4%	0.32
CDC	3.6%	95.5%	45.0%	18.7%	34.2%	95.6%	33.2%	0.34
CUSUM	44.0%	52.7%	80.5%	70.5%	75.4%	97.0%	9.5%	0.17
CUSUM Rossi	39.5%	57.6%	77.0%	65.9%	71.8%	96.9%	10.1%	0.18
CUSUM GLM	44.2%	52.0%	84.4%	73.8%	79.5%	97.5%	9.9%	0.18
CUSUM GLM Rossi	39.9%	56.8%	81.1%	69.5%	76.1%	97.3%	10.4%	0.18
Bayes 1 ( $\alpha = 0.05$ )	10.1%	90.5%	76.2%	66.2%	39.1%	95.7%	21.4%	0.28
Bayes 2 ( $\alpha = 0.05$ )	9.4%	91.0%	80.8%	69.4%	45.7%	96.2%	25.0%	0.32
Bayes 3 ( $\alpha = 0.05$ )	11.1%	88.9%	83.4%	71.9%	51.8%	96.5%	23.6%	0.32
RKI 1	8.3%	92.3%	67.8%	58.9%	30.4%	95.3%	20.6%	0.25
RKI 2	5.5%	94.7%	67.8%	57.8%	34.5%	95.6%	30.0%	0.32
RKI 3	7.0%	93.0%	71.3%	60.6%	41.8%	96.0%	28.3%	0.34
GLR NegBin	4.3%	95.7%	50.8%	29.8%	21.6%	94.9%	24.9%	0.23
GLR Poisson	15.5%	84.5%	75.5%	60.3%	45.9%	95.9%	16.4%	0.24
EARS C1	6.9%	93.7%	66.3%	57.4%	25.6%	95.0%	21.2%	0.23
<b>EARS C2</b>	<b>8.5%</b>	<b>92.4%</b>	<b>68.0%</b>	<b>57.1%</b>	<b>38.8%</b>	<b>95.8%</b>	<b>25.1%</b>	<b>0.31</b>
EARS C3	7.4%	92.9%	54.2%	8.5%	35.3%	95.6%	24.7%	0.29
OutbreakP	59.9%	37.4%	70.4%	67.9%	66.1%	94.4%	6.5%	0.12

# Performances des méthodes

Method	FPR	Specificity	POD	POD1week	Sensitivity	NPV	PPV	$F_1$ -measure
Improved Farrington	1.0%	99.0%	43.3%	34.0%	20.5%	95.0%	58.3%	0.30
Original Farrington	2.3%	97.7%	56.9%	45.5%	29.0%	95.4%	45.0%	0.35
Periodic Poisson GLM	3.3%	96.8%	67.8%	56.6%	35.6%	95.8%	42.3%	0.39
Periodic NegBin GLM	0.7%	99.4%	44.8%	36.3%	20.7%	95.0%	68.4%	0.32
CDC	3.6%	95.5%	45.0%	18.7%	34.2%	95.6%	33.2%	0.34
CUSUM	44.0%	52.7%	80.5%	70.5%	75.4%	97.0%	9.5%	0.17
CUSUM Rossi	39.5%	57.6%	77.0%	65.9%	71.8%	96.9%	10.1%	0.18
CUSUM GLM	44.2%	52.0%	84.4%	73.8%	79.5%	97.5%	9.9%	0.18
CUSUM GLM Rossi	39.9%	56.8%	81.1%	69.5%	76.1%	97.3%	10.4%	0.18
Bayes 1 ( $\alpha = 0.05$ )	10.1%	90.5%	76.2%	66.2%	39.1%	95.7%	21.4%	0.28
Bayes 2 ( $\alpha = 0.05$ )	9.4%	91.0%	80.8%	69.4%	45.7%	96.2%	25.0%	0.32
<b>Bayes 3 (<math>\alpha = 0.05</math>)</b>	<b>11.1%</b>	<b>88.9%</b>	<b>83.4%</b>	<b>71.9%</b>	<b>51.8%</b>	<b>96.5%</b>	<b>23.6%</b>	<b>0.32</b>
RKI 1	8.3%	92.3%	67.8%	58.9%	30.4%	95.3%	20.6%	0.25
RKI 2	5.5%	94.7%	67.8%	57.8%	34.5%	95.6%	30.0%	0.32
RKI 3	7.0%	93.0%	71.3%	60.6%	41.8%	96.0%	28.3%	0.34
GLR NegBin	4.3%	95.7%	50.8%	29.8%	21.6%	94.9%	24.9%	0.23
GLR Poisson	15.5%	84.5%	75.5%	60.3%	45.9%	95.9%	16.4%	0.24
EARS C1	6.9%	93.7%	66.3%	57.4%	25.6%	95.0%	21.2%	0.23
EARS C2	8.5%	92.4%	68.0%	57.1%	38.8%	95.8%	25.1%	0.31
EARS C3	7.4%	92.9%	54.2%	8.5%	35.3%	95.6%	24.7%	0.29
OutbreakP	59.9%	37.4%	70.4%	67.9%	66.1%	94.4%	6.5%	0.12

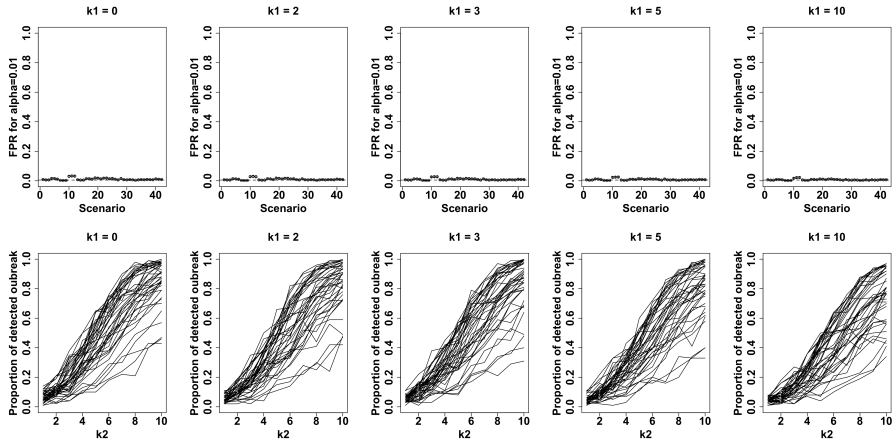
# Current Past Graph (CDC)

CDC (historical limits),  $\alpha = 0.01$

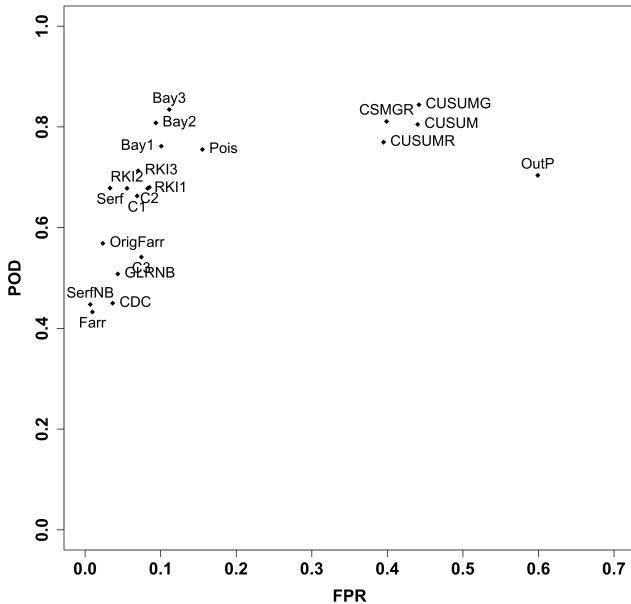


# Farrington amélioré

Improved Farrington,  $\alpha = 0.01$



# POD vs. FPR (alpha = 0.01)





# Discussion

# Utilisation de données simulées

## Force de ce travail

- Excès de cas déterminé *a priori*
- Volume de données important
- Grande diversité de séries et d'épidémies

## Mais repose sur un modèle paramétrique...

- Proche des modèles utilisés pour la détection
- ⇒ Validation croisée sur un autre jeu de données ?



# ○ — Diversité des méthodes testées

## A partir d'un *package* faisant référence

- Utilisation de toutes les méthodes sélectionnées

## Parti-pris opérationnel assumé

- Un seul "réglage" par méthode
- = celui paramétré par défaut dans le package

# ○ — Une évaluation objective

## Des performances limitées...

- ... en termes de sensibilité des méthodes
- Aucune méthode très supérieure aux autres
- Résultats des méthodes les plus "rudimentaires" +++

## Influence des caractéristiques des séries temporelles

- Non présenté aujourd'hui



ISDS Conference 2015 – Denver, CO, USA, 8-10 December 2015



## Performance of early outbreak detection algorithms in public health surveillance from a simulation study

Gabriel BÉDUBOURG<sup>1,2</sup> and Yann LE STRAT<sup>3</sup>



1. French Armed Forces Center for Epidemiology and Public Health, Marseille, France ; 2. UMR 912 SESSTIM, Aix Marseille University, Marseille, France ; 3. French Institute for Public Health Surveillance, Saint Maurice, France



ESCAIDE, Stockholm (SWE), 28-30 November 2016



## Outbreak detection: determinants of statistical algorithms' performances, a simulation based study

Gabriel BÉDUBOURG<sup>1,2</sup> and Yann LE STRAT<sup>3</sup>



- 1 French armed forces center for epidemiology and public health (CESPA), Marseille, France. 2 Aix Marseille Univ, INSERM, IRD, SESSTIM, Sciences Economiques & Sociales de la Santé & Traitement de l'Information Médicale, Marseille, France. 3 Direction des maladies infectieuses, Santé Publique France, Saint Maurice, France.

RESEARCH ARTICLE

## Evaluation and comparison of statistical methods for early temporal detection of outbreaks: A simulation-based study

Gabriel Bédubourg<sup>1,2\*</sup>, Yann Le Strat<sup>3</sup>

1 CESPA, French Armed Forces Center for Epidemiology and Public Health, Marseille, France, 2 Aix Marseille Univ, INSERM, IRD, SESSTIM, Sciences Economiques & Sociales de la Santé & Traitement de l'Information Médicale, Marseille, France, 3 Santé publique France, French national public health agency, F-94415 Saint-Maurice, France

▫ Current address: CESPA, GSBDD Marseille Aubagne, 111 Avenue de la Corse, BP 40026, 13568 Marseille Cedex 02, France

\* [gabrielbedubourg@hotmail.fr](mailto:gabrielbedubourg@hotmail.fr)



## Implementation and Initial Analysis of a Laboratory-Based Weekly Biosurveillance System, Provence-Alpes-Côte d'Azur, France

Michael Huart, Gabriel Bedubourg, Cédric Abat, Philippe Colson, Jean Marc Rolain, Hervé Chaudet, Pierre Edouard Fournier, Didier Raoult, Xavier DeParis

ESCAIDE, Stockholm (SWE), 28-30 November 2016

epiconcept

### French armed forces health surveillance tool (OSEA): an information system for outbreak detection

smart health

G. Bédubourg<sup>1,2</sup>, B. Fétiqueau<sup>3</sup>, M. Rondy<sup>3</sup>, R. Haus-Cheymol<sup>4</sup>, S. Duron<sup>1,2</sup>, F. Delon<sup>1</sup>, D. Khnafo<sup>3</sup>

<sup>1</sup>French armed forces center for epidemiology and public health (CESPA), Marseille, France.

<sup>2</sup>Aix Marseille Univ, INSERM, IRD, SESSTIM, Sciences Economiques & Sociales de la Santé & Traitement de l'Information Médicale, Marseille, France

<sup>3</sup>EpiConcept, Paris, France.

<sup>4</sup>French armed forces Surgeon General Office, Paris, France.



# Plan

1. Introduction générale
2. Objectifs de la thèse
3. Méthode de simulation des épidémies
4. Evaluation des méthodes de détection des épidémies
- 5. Combinaison de méthodes statistiques pour la détection**
6. Analyse étiologique des épidémies
7. Perspectives et conclusion





# Introduction



# ○ — Evaluation des algorithmes

## Des performances limitées

- Hétérogènes selon les scénarios de simulation
- Quelques méthodes très spécifiques...
- ... mais peu sensibles

## Intérêts de la combinaison de méthodes

- Combiner des méthodes spécifiques mais peu sensibles
- Pour améliorer la sensibilité

## Objectif

- Evaluer les performances de la combinaison d'algorithmes de détection



# Méthode

# ○ — Combinaisons de 2, 3 ou 4 méthodes

## Choix des algorithmes

- Parmi 21 algorithmes évalués dans l'étude précédente
- Dont 8 dépendaient de la valeur  $\alpha$  (3 valeurs)
- Evaluation de toutes combinaisons de 2, 3 ou 4 algorithmes

## Règle de décision simple

"Au moins une méthode émet une alarme  $\Rightarrow$  Alarme"

## Evaluation des performances des combinaisons

- A partir du jeu de données simulées présenté
- Mêmes indicateurs de performance



# Résultats

# 615 combinaisons de 2 méthodes

Method 1	Method 2	POD	FPR	POD1week	Se	Sp	PPV	NPV	$F_1$ -measure
Periodic reg. (0.05)	Poisson Bayes 3 (0.05)	85.8	12.0	74.2	53.6	88.1	23.0	96.6	0.32
EARS C1 (0.01)	Periodic Poisson reg. (0.05)	84.5	10.5	73.2	48.4	90.0	24.2	96.3	0.32
Periodic reg. (0.05)	Bayes 2 (0.05)	84.4	10.5	72.8	49.7	89.8	24.4	96.4	0.33
Periodic reg. (0.05)	Poisson EARS C1 (0.05)	87.6	14.3	76.8	50.7	86.5	19.9	96.4	0.29
EARS C1 (0.001)	Bayes 3 (0.05)	85.7	12.6	74.6	52.9	87.6	22.0	96.6	0.31

# 615 combinaisons de 2 méthodes

Method 1		Method 2		POD	FPR	POD1week	Se	Sp	PPV	NPV	$F_1$ -measure
Periodic reg. (0.05)	Poisson	Bayes 3 (0.05)		85.8	12.0	74.2	53.6	88.1	23.0	96.6	0.32
EARS C1 (0.01)		Periodic reg. (0.05)	Poisson	84.5	10.5	73.2	48.4	90.0	24.2	96.3	0.32
Periodic reg. (0.05)	Poisson	Bayes 2 (0.05)		84.4	10.5	72.8	49.7	89.8	24.4	96.4	0.33
Periodic reg. (0.05)	Poisson	EARS C1 (0.05)		87.6	14.3	76.8	50.7	86.5	19.9	96.4	0.29
EARS C1 (0.001)		Bayes 3 (0.05)		85.7	12.6	74.6	52.9	87.6	22.0	96.6	0.31

# 6079 combinaisons de 3 méthodes

Method 1	Method 2	Method 3	POD	FPR	POD1week	Se	Sp	PPV	NPV	$F_1$ -measure
EARS (0.01) C1	Original Farrington (0.05)	Periodic Poisson reg. (0.05)	85.9	11.8	74.6	50.8	88.6	22.7	96.5	0.31
EARS (0.001) C1	Periodic Poisson reg. (0.05)	Bayes (0.05) 2	85.6	11.6	74.4	50.4	88.8	23.0	96.4	0.32
RKI 3	EARS (0.01) C1	Periodic Poisson reg. (0.05)	86.3	12.3	75.1	51.6	88.1	22.2	96.5	0.31
EARS (0.001) C1	Periodic Poisson reg. (0.05)	Bayes (0.05) 3	87.1	13.2	76.1	54.4	87.0	21.6	96.7	0.31
RKI 3	EARS (0.001) C1	Periodic Poisson reg. (0.05)	85.1	10.9	73.8	50.9	89.4	24.1	96.5	0.33
EARS (0.001) C1	Original Farrington (0.05)	Periodic Poisson reg. (0.05)	84.6	10.3	73.2	50.1	90.0	24.9	96.5	0.33

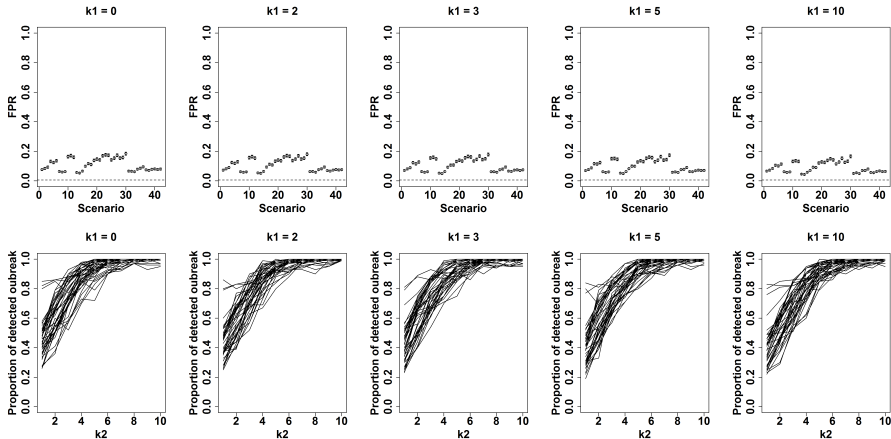
# 6079 combinaisons de 3 méthodes

Method 1	Method 2	Method 3	POD	FPR	POD1week	Se	Sp	PPV	NPV	$F_1$ -measure	
EARS (0.01)	C1 Original Farrington (0.05)	Periodic Poisson reg. (0.05)	85.9	11.8	74.6	50.8	88.6	22.7	96.5	0.31	
EARS (0.001)	C1 Periodic Poisson reg. (0.05)	Bayes (0.05)	2	85.6	11.6	74.4	50.4	88.8	23.0	96.4	0.32
RKI 3	EARS C1 (0.01)	Periodic Poisson reg. (0.05)	86.3	12.3	75.1	51.6	88.1	22.2	96.5	0.31	
EARS (0.001)	C1 Periodic Poisson reg. (0.05)	Bayes 3 (0.05)	87.1	13.2	76.1	54.4	87.0	21.6	96.7	0.31	
RKI 3	EARS C1 (0.001)	Periodic Poisson reg. (0.05)	85.1	10.9	73.8	50.9	89.4	24.1	96.5	0.33	
EARS (0.001)	C1 Original Farrington (0.05)	Periodic Poisson reg. (0.05)	84.6	10.3	73.2	50.1	90.0	24.9	96.5	0.33	



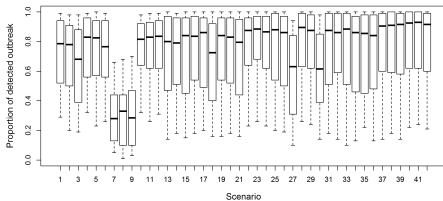
# Combinaison de 3 algorithmes

Combinaison of 0001.earsC1, 005.Farrington.native & 005.Serfling

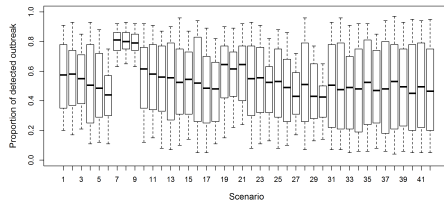


# Effets sur la POD (3 algorithmes)

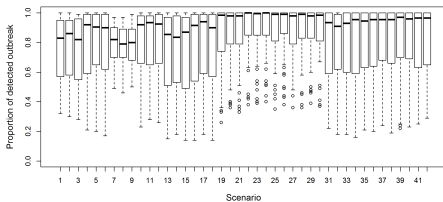
005.Farrington.native



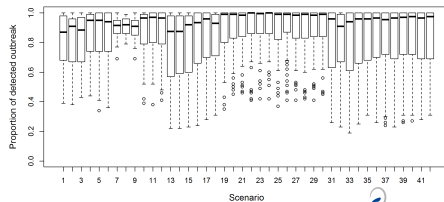
0001.earsC1



005.Serfling



Combination of 005.Farrington.native, 0001.earsC1 & 005.Serfling



# 39850 combinaisons de 4 méthodes

Method 1	Method 2	Method 3	Method 4	POD	FPR	POD1week	Se	Sp	PPV	NPV	$F_1$ -measure
RKI 3	EARS C1 (0.001)	Original Farrington (0.05)	Periodic Poisson reg. (0.05)	86.0	11.8	74.8	52.5	88.4	23.0	96.6	0.32
RKI 2	EARS C1 (0.001)	Original Farrington (0.05)	Periodic Poisson reg. (0.05)	85.4	11.0	74.0	50.6	89.3	23.8	96.5	0.32
EARS C1 (0.001)	Improved Farrington (0.01)	Periodic Poisson reg. (0.05)	Bayes 2 (0.05)	86.0	11.8	74.8	51.1	88.5	22.7	96.5	0.31
EARS C1 (0.001)	Original Farrington (0.05)	Periodic Poisson reg. (0.05)	Bayes 2 (0.05)	86.6	12.5	75.4	52.3	87.8	22.1	96.5	0.31
RKI 2	EARS C1 (0.01)	Original Farrington (0.05)	Periodic Poisson reg. (0.05)	86.4	12.4	75.2	51.3	88.0	22.1	96.5	0.31

# Discussion

# ○ — Des performances améliorées

## Une amélioration de la sensibilité

- Avec une perte limitée en spécificité
- Plusieurs combinaisons présentent une POD  $> 85\%$
- En conservant un FPR  $< 15\%$

## Moins de dépendance aux scénarios de simulation

⇒ Plus de confiance dans les systèmes d'analyses automatisées

# Article soumis

## Soumis à *Epidemiology and Infection* en juillet 2018

- Refusé car "trop technique"

## Re-soumis à PloS One en décembre 2018



**PLOS** | SUBMISSION

Combination of statistical methods for the early time detection of outbreaks: a simulation-based study

Bédubourg Gabriel<sup>1,2,3,\*</sup>, Mayet Aurélie<sup>1,2</sup>, Meynard Jean-Baptiste<sup>1,2</sup>, Le Strat Yann<sup>3</sup>

1 CESPA, French Armed Forces Center for Epidemiology and Public Health, Marseille, France.

2 Aix Marseille Univ, INSERM, IRD, SESSTIM, Sciences Economiques & Sociales de la Santé & Traitement de l'Information Médicale, Marseille, France.

3 Santé publique France, French national public health agency, F-94415 Saint-Maurice, France.

□ Current Address: CESPA, GSBDD Marseille Aubagne, 111 Avenue de la Corse, BP 40026, 13568 Marseille Cedex 02, France.

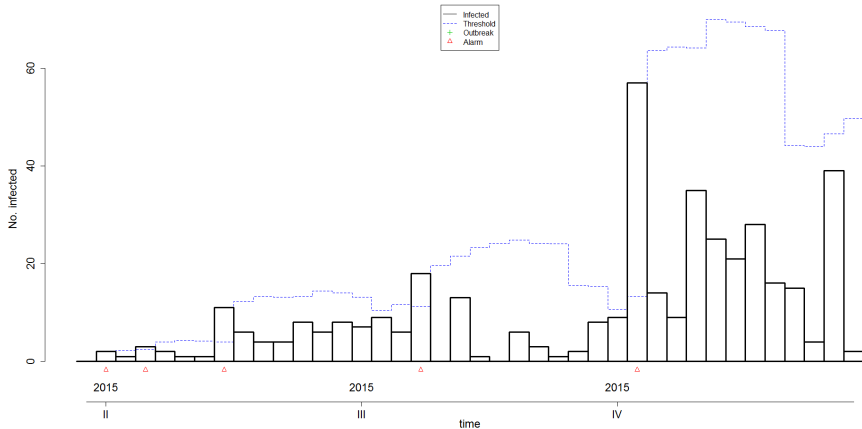
\* gabrielbedubourg@hotmail.fr

# Plan

1. Introduction générale
2. Objectifs de la thèse
3. Méthode de simulation des épidémies
4. Evaluation des méthodes de détection des épidémies
5. Combinaison de méthodes statistiques pour la détection
- 6. Analyse étiologique des épidémies**
7. Perspectives et conclusion

# Série temporelle

Surveillance des diarrhées, Unité Barkhane - semaines 8 à 48/2015





# ○ — Exemple d'application

**Augmentation des foyers de toxi-infection alimentaire collective en France entre 1996 et 2010 :  
réelle ou effet surveillance ?**



# Introduction

# Définition

## Toxi-infections alimentaires collectives (Tiac)

Survenue d'**au moins deux cas** d'une symptomatologie similaire, en général digestive, dont on peut rapporter la cause à une **même origine alimentaire**

# ○ — Grande variété des Tiac

## Agents pathogènes nombreux

- Bactéries (Salmonelles, *C. perfringens*, *S. aureus*...)
- Virus (virus entériques...)
- Parasites (*Cryptosporidium*...)

## Formes cliniques variées

- Digestives +++
- Atteintes neurologiques (toxine du botulisme)
- Atteintes vasomotrices (Histamine)
- *Etc.*

# Autres éléments épidémiologiques

## Gravité

- Hospitalisations (7% des cas entre 2006 et 2008)
- Décès (4,5 pour 10000 cas)

## Réalisation d'enquêtes épidémiologiques

- Identifier le(s) agent(s)/aliment(s) responsable(s)  
⇒ **contrôler l'épisode**
- Identifier les facteurs favorisants  
⇒ **prévenir un nouvel épisode**

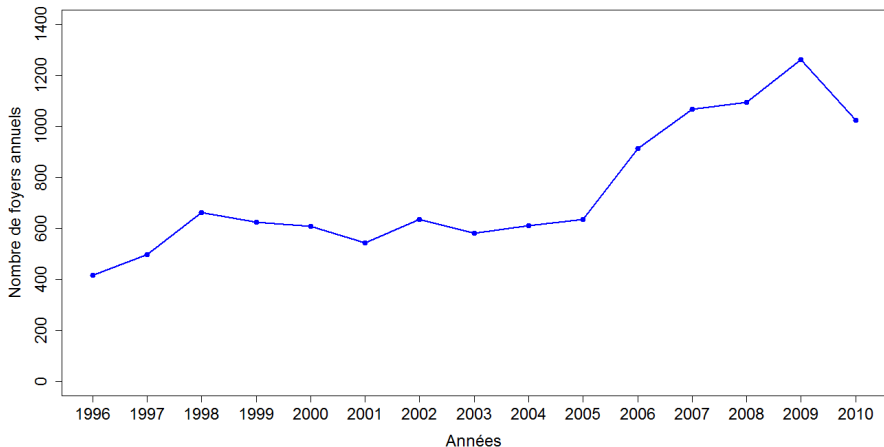
## Déclaration obligatoire (DO)

- Depuis 1987



# Contexte

# Augmentation importante



Source : InVS-2011

# ○ — Modifications épidémiologiques

## Des tendances connues pour quelques agents

- ↘ des infections à Salmonelles
- ↗ des infections à *Campylobacter* depuis 2006



# ○ — Modifications de la surveillance

**2002** : CNR des Virus entériques

**2004** : Début du déploiement du logiciel WinTiac

**2005** : Fin du dépl. WinTiac + formation des acteurs

**2006** : Renfort d'un médecin à l'InVS



# Objectifs

# Objectifs

**Décrire et quantifier l'augmentation observée**

**Identifier la part de cette augmentation**

- Compatible avec des **modifications de l'épidémiologie des Tiac**
- Compatible avec des modifications à l'origine d'un **biais de surveillance**

# ○ — Augmentation réelle : hypothèses

## Identifier la part liée à des modifications de l'épidémiologie des Tiac

Devrait se traduire par

- ↗ **disproportionnée** de certains types de foyers
  - ▶ certains agents
  - ▶ certains modes de restauration
  - ▶ certains aliments...

# ○ — Biais de surveillance : hypothèses

## Identifier la part liée à des modifications du système de surveillance

Devrait se traduire par :

- ↗ de tous les types de foyers
- ↗ +++ des foyers **possiblement moins bien relayés** par les Ddass vers l'InVS avant 2005
  - ▶ Familiaux
  - ▶ Faible nombre de cas
  - ▶ Sans aliment ou agent pathogène identifié



# Méthodes

# Construction de séries temporelles

## Données de la DO

- Variables explicatives (agent pathogène, aliment...)
- Une modalité = une série "explicative"

## Echelle mensuelle

- De janvier 1996 à décembre 2010

⇒ **1 série principale et 37 séries explicatives**

# ○ — Outils statistiques

## Modèles linéaires avec erreurs ARMA

- 6 indicatrices annuelles de 2005 à 2010

$$Z_t = \alpha_0 + \alpha_1 I_{2005t} + \alpha_2 I_{2006t} + \alpha_3 I_{2007t} + \alpha_4 I_{2008t} + \alpha_5 I_{2009t} + \alpha_6 I_{2010t} + \epsilon_t$$

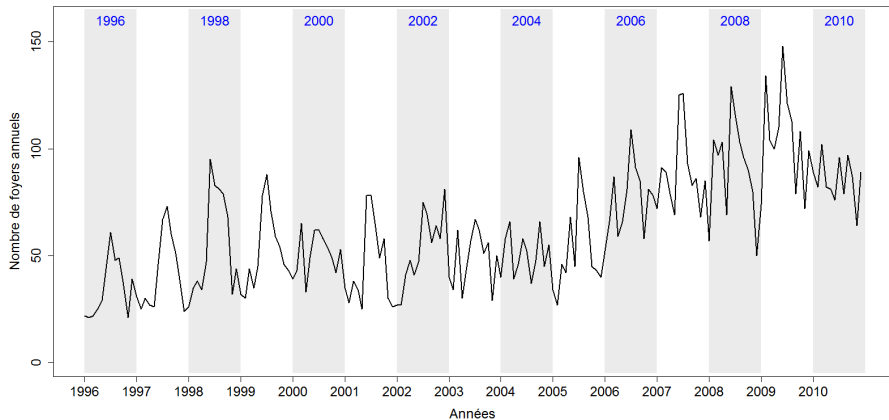
avec  $\epsilon_t = \frac{(1-\omega B)^q}{(1-\rho B)^p} \nu_t$  et  $\nu_t \sim N(0, \sigma^2)$

## Modèle estimé pour

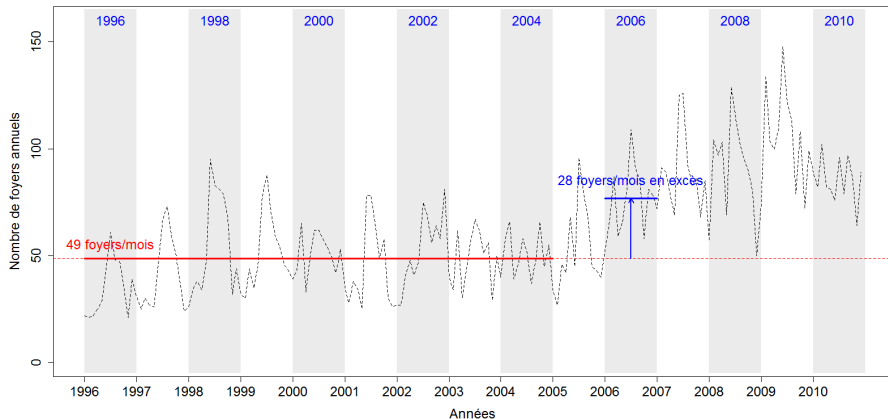
- La série principale
- L'ensemble des séries explicatives



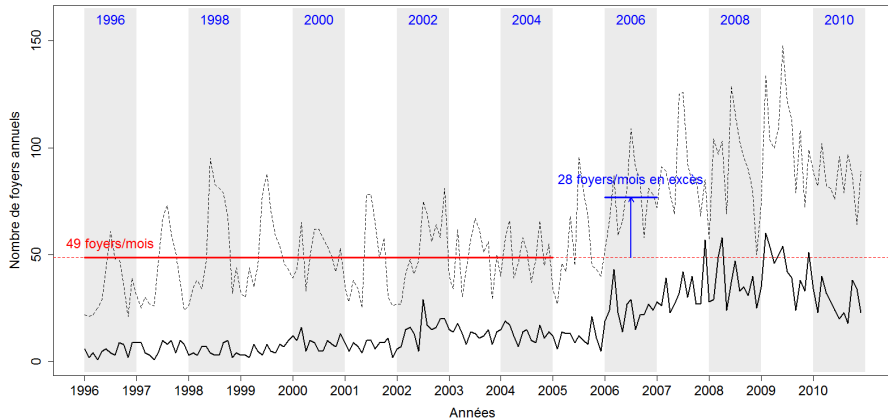
# Part attribuable (PA)



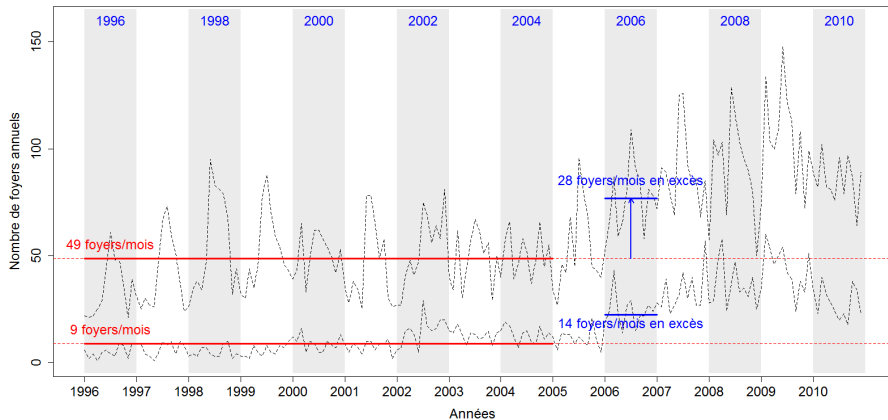
# Part attribuable (PA)



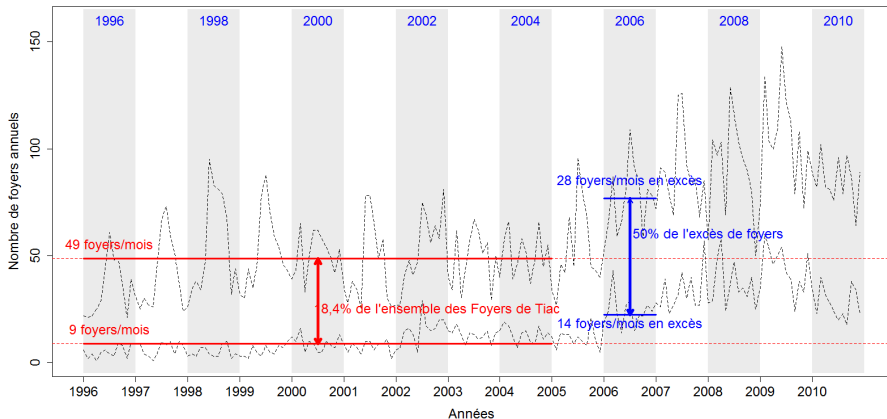
# Part attribuable (PA)



# Part attribuable (PA)



# Part attribuable (PA)





# Résultats

# ○ — Population de l'étude

## Au total

- 11260 foyers de Tiac

## Données exploitables

- 11172 foyers inclus

**Plus de 99% des foyers inclus**

# Série principale

Années	1996–2004	2005	2006	2007	2008	2009	2010
Niv. moyen <sup>1</sup>	<b>48,6</b>	–	–	–	–	–	–
Excès de foy. <sup>2</sup>	–	5,5	<b>28,2</b>	<b>42,5</b>	<b>41,9</b>	<b>56,5</b>	<b>37,2</b>
IC95%	46,7–50,4	-2,1–13,0	20,4–36,0	34,5–50,5	33,8–50,0	48,4–64,7	28,9–45,5
p	< 0,001	0,15	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
% d'augment. <sup>3</sup>	–	11,3%	<b>58,1%</b>	<b>87,4%</b>	<b>86,2%</b>	<b>116,4%</b>	<b>76,6%</b>

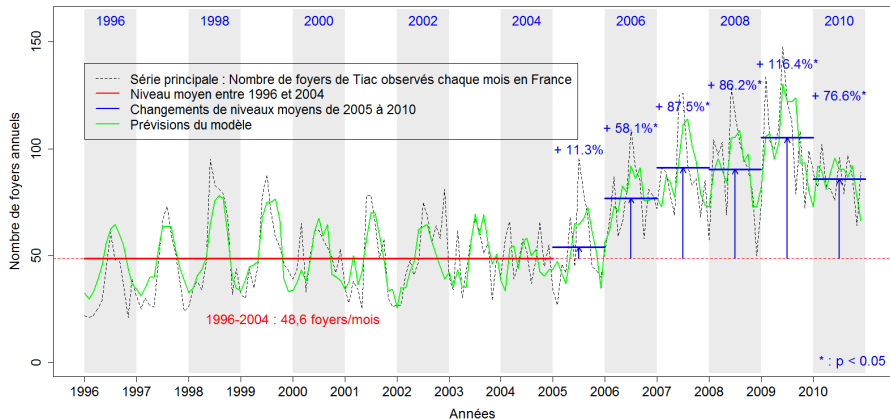
<sup>1</sup> nombre de foyers déclarés chaque mois en moyenne

<sup>2</sup> nombre de foyers déclarés en plus chaque mois par rapport à 1996–2004

<sup>3</sup> par rapport à la période 1996–2004



# Série principale



# Part attribuable de certains agents

Série explicative	% de l'ens. des foy.		% de l'augmentation			
	1996–2004	2006	2007	2008	2009	2010
<b>Agents pathogènes incriminés</b>						
<i>Salmonella sp.</i>	<b>34,2%</b>	– <sup>1</sup>	–	–	–	–
<i>Campylobacter sp.</i>	<b>0,7%</b>	–	1,3%	1,6%	2,0%	3,5%
(...)						
Agent inconnu	<b>18,7%</b>	48,3%	55,8%	63,6%	59,3%	50,8%
Virus entériques	<b>4,0%</b>	13,0%	–	–	7,9%	19,4%
(...)						

<sup>1</sup> PA non estimée : coefficient de régression non significatif

# Part attribuable de certains agents

Série explicative	% de l'ens. des foy.		% de l'augmentation			
	1996–2004	2006	2007	2008	2009	2010
<b>Agents pathogènes incriminés</b>						
<i>Salmonella sp.</i>	<b>34,2%</b>	– <sup>1</sup>	–	–	–	–
<i>Campylobacter sp.</i>	<b>0,7%</b>	–	1,3%	1,6%	2,0%	3,5%
(...)						
Agent inconnu	<b>18,7%</b>	48,3%	55,8%	63,6%	59,3%	50,8%
Virus entériques	<b>4,0%</b>	13,0%	–	–	7,9%	19,4%
(...)						

<sup>1</sup> PA non estimée : coefficient de régression non significatif

# Part attribuable de certains agents

Série explicative	% de l'ens. des foy.		% de l'augmentation			
	1996–2004	2006	2007	2008	2009	2010
<b>Agents pathogènes incriminés</b>						
<i>Salmonella sp.</i>	<b>34,2%</b>	– <sup>1</sup>	–	–	–	–
<i>Campylobacter sp.</i>	<b>0,7%</b>	–	1,3%	1,6%	2,0%	3,5%
(...)						
Agent inconnu	<b>18,7%</b>	48,3%	55,8%	63,6%	59,3%	50,8%
Virus entériques	<b>4,0%</b>	13,0%	–	–	7,9%	19,4%
(...)						

<sup>1</sup> PA non estimée : coefficient de régression non significatif

# Part attribuable de certains agents

Série explicative	% de l'ens. des foy.		% de l'augmentation			
	1996–2004	2006	2007	2008	2009	2010
<b>Agents pathogènes incriminés</b>						
<i>Salmonella sp.</i>	<b>34,2%</b>	– <sup>1</sup>	–	–	–	–
<i>Campylobacter sp.</i>	<b>0,7%</b>	–	1,3%	1,6%	2,0%	3,5%
(...)						
<b>Agent inconnu</b>	<b>18,7%</b>	<b>48,3%</b>	<b>55,8%</b>	<b>63,6%</b>	<b>59,3%</b>	<b>50,8%</b>
Virus entériques	<b>4,0%</b>	13,0%	–	–	7,9%	19,4%
(...)						

<sup>1</sup> PA non estimée : coefficient de régression non significatif

# Part attribuable de certains agents

Série explicative	% de l'ens. des foy.		% de l'augmentation			
	1996–2004	2006	2007	2008	2009	2010
<b>Agents pathogènes incriminés</b>						
<i>Salmonella sp.</i>	<b>34,2%</b>	– <sup>1</sup>	–	–	–	–
<i>Campylobacter sp.</i>	<b>0,7%</b>	–	1,3%	1,6%	2,0%	3,5%
(...)						
Agent inconnu	<b>18,7%</b>	48,3%	55,8%	63,6%	59,3%	50,8%
Virus entériques	<b>4,0%</b>	13,0%	–	–	7,9%	19,4%
(...)						

<sup>1</sup> PA non estimée : coefficient de régression non significatif

# PA de certains types de restauration

Série explicative	% de l'ens. des foy.		% de l'augmentation			
	1996–2004	2006	2007	2008	2009	2010
<b>Lieu de restauration</b>						
Rest. commerciale	<b>23,2%</b>	35,1%	33,3%	31,9%	34,4%	45,2%
Tiac familiales	<b>35,1%</b>	30,9%	23,6%	23,1%	30,1%	36,8%
Lieu inconnu	<b>0,6%</b>	– <sup>1</sup>	–	1,7%	2,5%	1,3%
(...)						

<sup>1</sup> PA non estimée : coefficient de régression non significatif

# PA de certains types de restauration

Série explicative	% de l'ens. des foy.		% de l'augmentation			
	1996–2004	2006	2007	2008	2009	2010
<b>Lieu de restauration</b>						
Rest. commerciale	<b>23,2%</b>	35,1%	33,3%	31,9%	34,4%	45,2%
Tiac familiales	<b>35,1%</b>	30,9%	23,6%	23,1%	30,1%	36,8%
Lieu inconnu	<b>0,6%</b>	– <sup>1</sup>	–	1,7%	2,5%	1,3%
(...)						

<sup>1</sup> PA non estimée : coefficient de régression non significatif



# PA de certains types de restauration

Série explicative	% de l'ens. des foy.		% de l'augmentation			
	1996–2004	2006	2007	2008	2009	2010
<b>Lieu de restauration</b>						
Rest. commerciale	<b>23,2%</b>	35,1%	33,3%	31,9%	34,4%	45,2%
Tiac familiales	<b>35,1%</b>	30,9%	23,6%	23,1%	30,1%	36,8%
Lieu inconnu	<b>0,6%</b>	– <sup>1</sup>	–	1,7%	2,5%	1,3%
(...)						

<sup>1</sup> PA non estimée : coefficient de régression non significatif

# PA de certains types de restauration

Série explicative	% de l'ens. des foy.		% de l'augmentation			
	1996–2004	2006	2007	2008	2009	2010
<b>Lieu de restauration</b>						
Rest. commerciale	<b>23,2%</b>	35,1%	33,3%	31,9%	34,4%	45,2%
Tiac familiales	<b>35,1%</b>	30,9%	23,6%	23,1%	30,1%	36,8%
Lieu inconnu	<b>0,6%</b>	– <sup>1</sup>	–	1,7%	2,5%	1,3%
(...)						

<sup>1</sup> PA non estimée : coefficient de régression non significatif

# Aliments et nombre de malades

Série explicative	% de l'ens. des foy.	% de l'augmentation				
	1996–2004	2006	2007	2008	2009	2010
<b>Principales familles d'aliments</b>						
Inconnu (...)	<b>26,6%</b>	61,8%	58,1%	70,2%	46,8%	29,2%
<b>Nombre de malades</b>						
< 3 cas (...)	<b>15,8%</b>	32,0%	26,5%	31,5%	28,2%	34,3%
≥ 13 cas	<b>31,0%</b>	- <sup>1</sup>	15,1%	23,7%	16,8%	10,5%

<sup>1</sup> PA non estimée : coefficient de régression non significatif

# Aliments et nombre de malades

Série explicative	% de l'ens. des foy.		% de l'augmentation			
	1996–2004	2006	2007	2008	2009	2010
<b>Principales familles d'aliments</b>						
Inconnu	<b>26,6%</b>	61,8%	58,1%	70,2%	46,8%	29,2%
(...)						
<b>Nombre de malades</b>						
< 3 cas	<b>15,8%</b>	32,0%	26,5%	31,5%	28,2%	34,3%
(...)						
≥ 13 cas	<b>31,0%</b>	- <sup>1</sup>	15,1%	23,7%	16,8%	10,5%

<sup>1</sup> PA non estimée : coefficient de régression non significatif

# Aliments et nombre de malades

Série explicative	% de l'ens. des foy.	% de l'augmentation				
	1996–2004	2006	2007	2008	2009	2010
<b>Principales familles d'aliments</b>						
Inconnu (...)	<b>26,6%</b>	61,8%	58,1%	70,2%	46,8%	29,2%
<b>Nombre de malades</b>						
< 3 cas (...)	<b>15,8%</b>	32,0%	26,5%	31,5%	28,2%	34,3%
≥ 13 cas	<b>31,0%</b>	– <sup>1</sup>	15,1%	23,7%	16,8%	10,5%

<sup>1</sup> PA non estimée : coefficient de régression non significatif

# Aliments et nombre de malades

Série explicative	% de l'ens. des foy.	% de l'augmentation				
	1996–2004	2006	2007	2008	2009	2010
<b>Principales familles d'aliments</b>						
Inconnu (...)	<b>26,6%</b>	61,8%	58,1%	70,2%	46,8%	29,2%
<b>Nombre de malades</b>						
< 3 cas (...)	<b>15,8%</b>	32,0%	26,5%	31,5%	28,2%	34,3%
≥ 13 cas	<b>31,0%</b>	– <sup>1</sup>	15,1%	23,7%	16,8%	10,5%

<sup>1</sup> PA non estimée : coefficient de régression non significatif



# Discussion

# ○ — Une augmentation importante

## Entre 2006 et 2010

- Atteignant 116% d'excès en 2009



# ○ — Modifications de l'épidémiologie ?

## Des modifications compatibles avec les données des systèmes de surveillance spécifiques

- Diminution des foyers à Salmonelles
- Augmentation des foyers à *Campylobacter sp.*

⇒ N'explique pas l'augmentation observée

↗ des Foyers en **restauration commerciale**

# ○ — Biais de surveillance ?

## Une augmentation disproportionnée

- Des foyers à Virus entériques
  - ▶ Principale étiologie des épidémies de GEA
  - ▶ Nouvelles capacités diagnostiques (CNR)
- Des foyers peu documentés
  - ▶ Agent et aliment inconnus
  - ▶ Peu de cas...

# ○ — Augmentation réelle ou biais ?

## Nous retrouvons des observations en faveur de

- L'absence de modification importante de l'épidémiologie des Tiac, en dehors des Salmonelles et des *Campylobacter*
- Une meilleure transmission des signalements vers l'InVS compatibles avec une amélioration du système de surveillance par le déploiement de WinTiac, la formation des personnels...

## Nécessité de mieux décrire l'augmentation des Tiac en restauration collective

# ○ — Valorisation

## Article en cours de finition

- revue ciblée : *Epidemiology and Infection*

## Communication orale au congrès ADELFF-EPITER 2014 (Nice)

## Méthodes mises en œuvre dans le cadre d'un Master 2



2<sup>ème</sup> année du Master Sciences, Technologies, Santé :  
mention Santé Publique - Spécialité Epidémiologie

Promotion 2015-2016

**DETERMINANTS DE L'EVOLUTION  
TEMPORELLE DES CAS INCIDENTS DE  
TRAUMATISMES SONORES AIGUS  
DANS LES ARMEES FRANCAISES, 2008-2015**

***Doris Rosario MEDINA-GARIN***

Née le 29 mars 1978

**Maitres de stage : Dr. Franck Berger**

**Chef du Service de promotion, prévention, programmes de santé**

***Dr. Gabriel Bédubourg***

- 97 -

# Plan

1. Introduction générale
2. Objectifs de la thèse
3. Méthode de simulation des épidémies
4. Evaluation des méthodes de détection des épidémies
5. Combinaison de méthodes statistiques pour la détection
6. Analyse étiologique des épidémies
7. Perspectives et conclusion

# ○ — De nouveaux outils en surveillance

## Pour la détection des épidémies

- Une étude d'évaluation et une approche originale
- Des applications concrètes (OSEA, PACASurvE...)

## Pour l'investigation des épidémies

- En exploitant les données de la surveillance
- Applicables pour de nombreux systèmes de surveillance
- Gain de temps, de moyens...

# Perspectives

## Pour la détection des épidémies

- Evaluation des méthodes d'apprentissage profond
- Projet de Master 2 en systèmes d'information en santé

## Projet Bédouin 2

- Place du traitement automatique de la langue et de la science des données en intelligence médicale
- Co-encadrement de thèses





# CESPA

Centre d'épidémiologie et de santé publique des armées  
**Observer Analyser Agir**