

MÉTHODES INNOVANTES EN SURVEILLANCE SYNDROMIQUE : TRAITEMENT DU LANGAGE ET DATA MINING

Rencontres Santé publique France , 4 juin 2019

Yasmine Baghdadi, Pascal Vilain

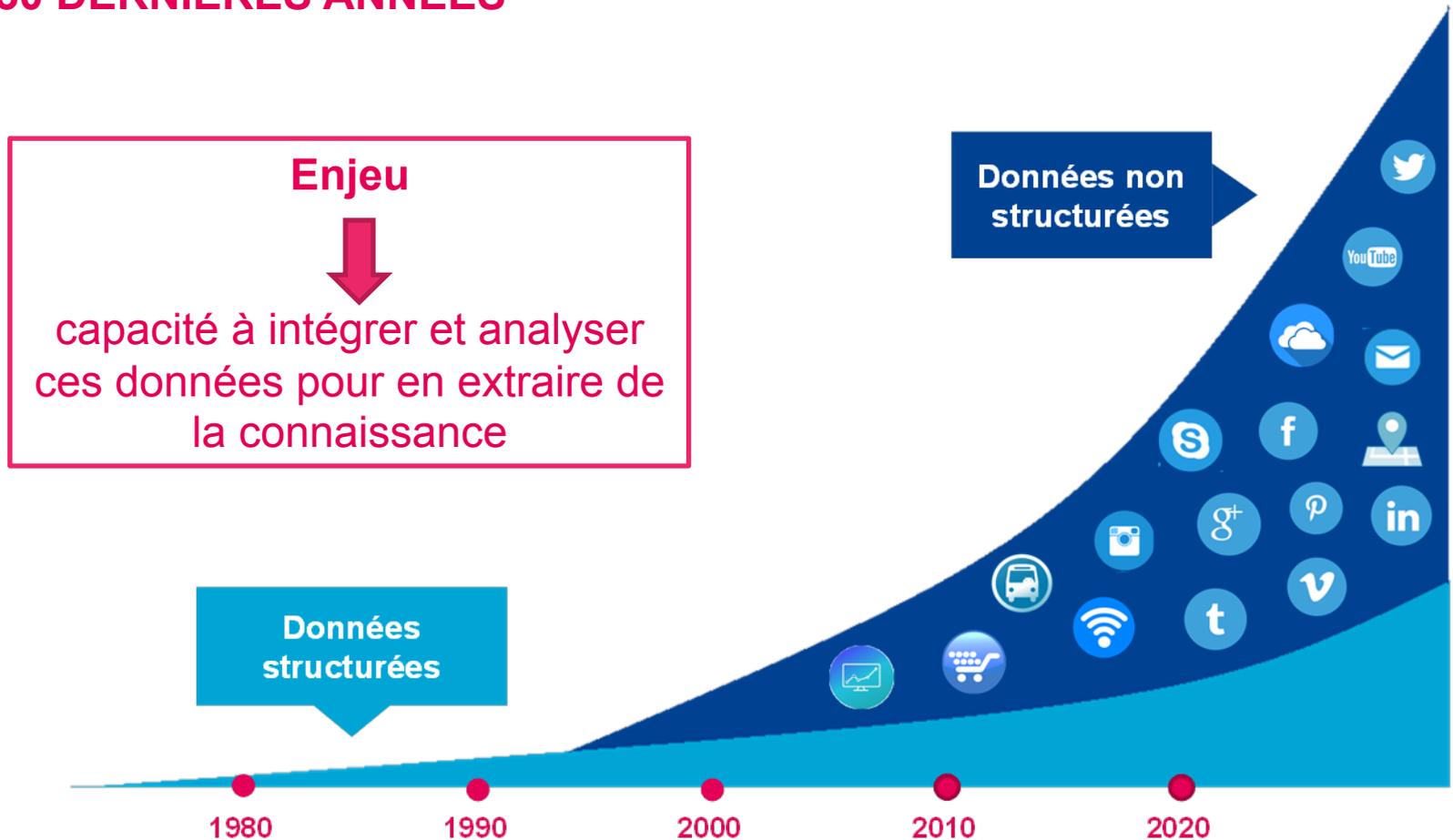
Contribution: Alix Bourrée, Anne Fouillet



Cette intervention est faite en tant que personnel de Santé publique France, organisateur de la manifestation.
Nous n'avons pas de lien d'intérêt avec le sujet traité.

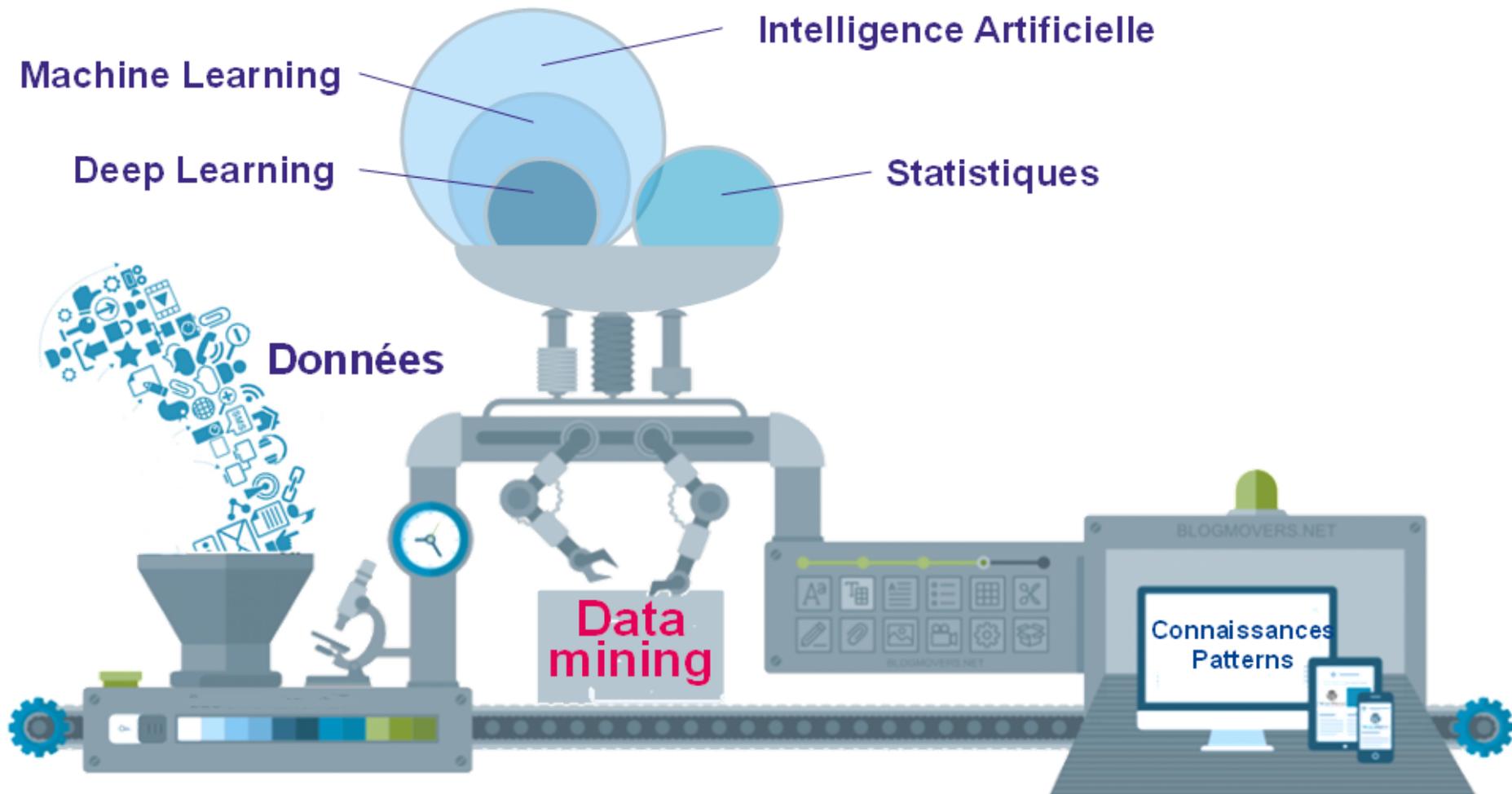
EXPLOSION DES DONNÉES : NOUVEAU CHALLENGE

CROISSANCE EXPONENTIELLE DES DONNÉES NUMÉRIQUES AU COURS DES 30 DERNIÈRES ANNÉES



DATA MINING : KESAKO ?

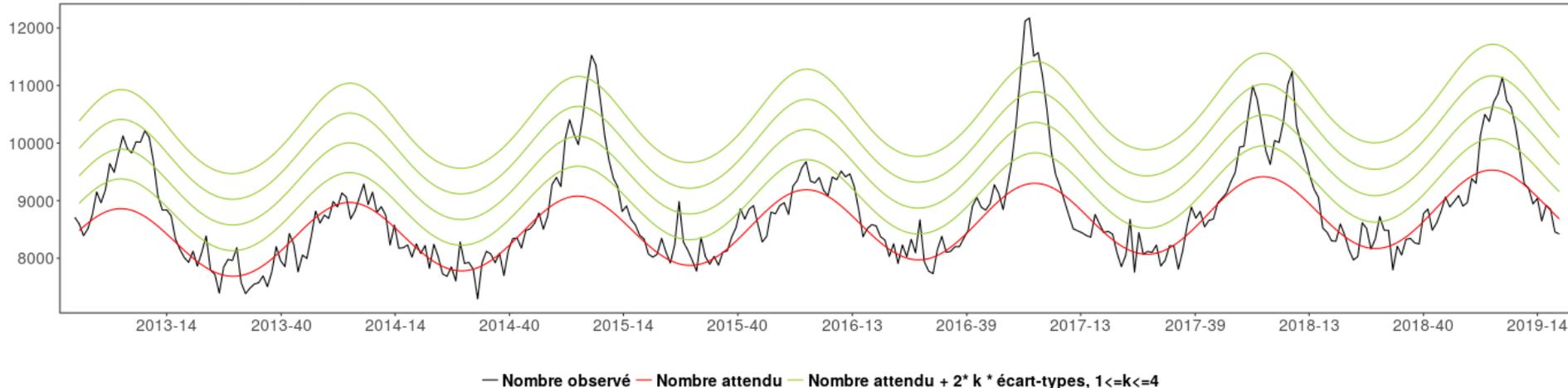
DATA MINING OU FOUILLE DE DONNÉES OU FORAGE DE DONNÉES



LA SURVEILLANCE RÉACTIVE DE LA MORTALITÉ

- Depuis 2005 : fondée sur les données d'état civil fournies par l'INSEE
 - Surveillance quantitative toutes causes : Absence de cause de décès

Effectifs hebdomadaires de mortalité - France - Tous Ages
Sources : Santé publique France - Insee



- Suivi des variations de la mortalité en routine :
 - Détection de variations habituelles ou inhabituelles
 - Estimation au cours ou à la suite d'un événement de l'impact quantitatif sur la santé

L'absence de causes médicales de décès ne permet pas :

- Identifier les principales causes associées à l'excès
- Déterminer la part de l'excès de décès liée à l'événement

Certification électronique des décès

- Depuis 2007: réception des décès issus de la certification électronique
 - Accès aux causes de décès en temps réel sous forme de texte libre

Causes du décès ?

Renseignements confidentiels et anonymes

Partie I : Maladie(s) ou affection(s) morbide(s) ayant directement provoqué le décès ?
La dernière ligne remplie doit correspondre à la cause initiale

Intervalle entre le début du processus morbide et le décès (heure, jours, mois ou ans).

*a) Arret cardio respiratoire	<input type="text"/>	--	▼
due à ou consécutive à : b) rupture anevrysme cerebral	<input type="text"/>	--	▼
due à ou consécutive à : c)	<input type="text"/>	--	▼
due à ou consécutive à : d)	<input type="text"/>	--	▼

Partie II : Autres états morbides, facteurs ou états physiologiques (grossesse...) ayant contribué au décès, mais non mentionnés en Partie I ?

hepatite a et b, insuffisance resp, maladie parkinson	<input type="text"/>	--	▼
	<input type="text"/>	--	▼

* Champs obligatoires

Informations complémentaires

Causes de décès

Facteurs associés

- Texte libre, non structuré, fautes d'orthographe, abréviation

Surveillance réactive de la mortalité à partir des causes de décès (1/2)

Principe : Suivi en routine des variations des regroupements syndromiques de mortalité

ETAPE 1: DÉFINITION DES RS

- **Regroupement syndromique (RS)**

Un regroupement syndromique est la **réunion** de **termes, expressions, de causes médicales de décès** ayant un sens pour nos objectifs de surveillance réactive

Evènements	Causes infectieuses	Causes environnementales	Autres causes
Inhabituels- innattendus	- MDO - Gastroenterite/Diarhée - Arboviruse/ Fièvre virale moustiques - Méningite virale - Sepsis	- Autre évènement environnemental (Avalanche, tempête)	- Insuffisance respiratoire aiguë - Asphyxie et anomalies de la respiration - Fièvre d'origine inconnue - Effet toxique de substances médicinales - Nausées et vomissements - Déshydratation
Attendus	- Grippe - Infections respiratoires aigües basses - Bronchiolite	- Chaleur - Froid	- Suicide - Accident de la route - AVC/ hémorragie cérébrale - Embolisme pulmonaire

ETAPE 2 : CLASSIFICATION DES CAUSES DE DÉCÈS DANS LES RS À L'AIDE DE 2 MÉTHODES DE TRAITEMENT AUTOMATIQUE DES LANGUES ET ÉVALUATION DES PERFORMANCES

Certificat de décès

ID	Champs	Causes de décès
259874	1	rupture anevrysme cerebral
259874	2	hepatite a et b,insuffisance resp, malaidie parkinson

Associé à

RS
AVC/ Hémorragie cérébrale
Hépatite Virale
Maladies chroniques du système nerveux
Insuffisance respiratoire aigue

- **Méthode à base de règles linguistiques**

Recherche d'une correspondance entre la cause de décès contenue dans le certificat et une ressource répertoriant une grande variété d'expressions de causes de décès et le RS associé

- **Méthode par apprentissage automatique : SVM (Support Vector Machine)**

- On cherche à classer des objets (champs du certificat) dans des classes (RS) grâce à un ensemble d'apprentissage
- Le principe est de séparer le plus distinctement des objets (définis par un vecteur) dans différentes classes

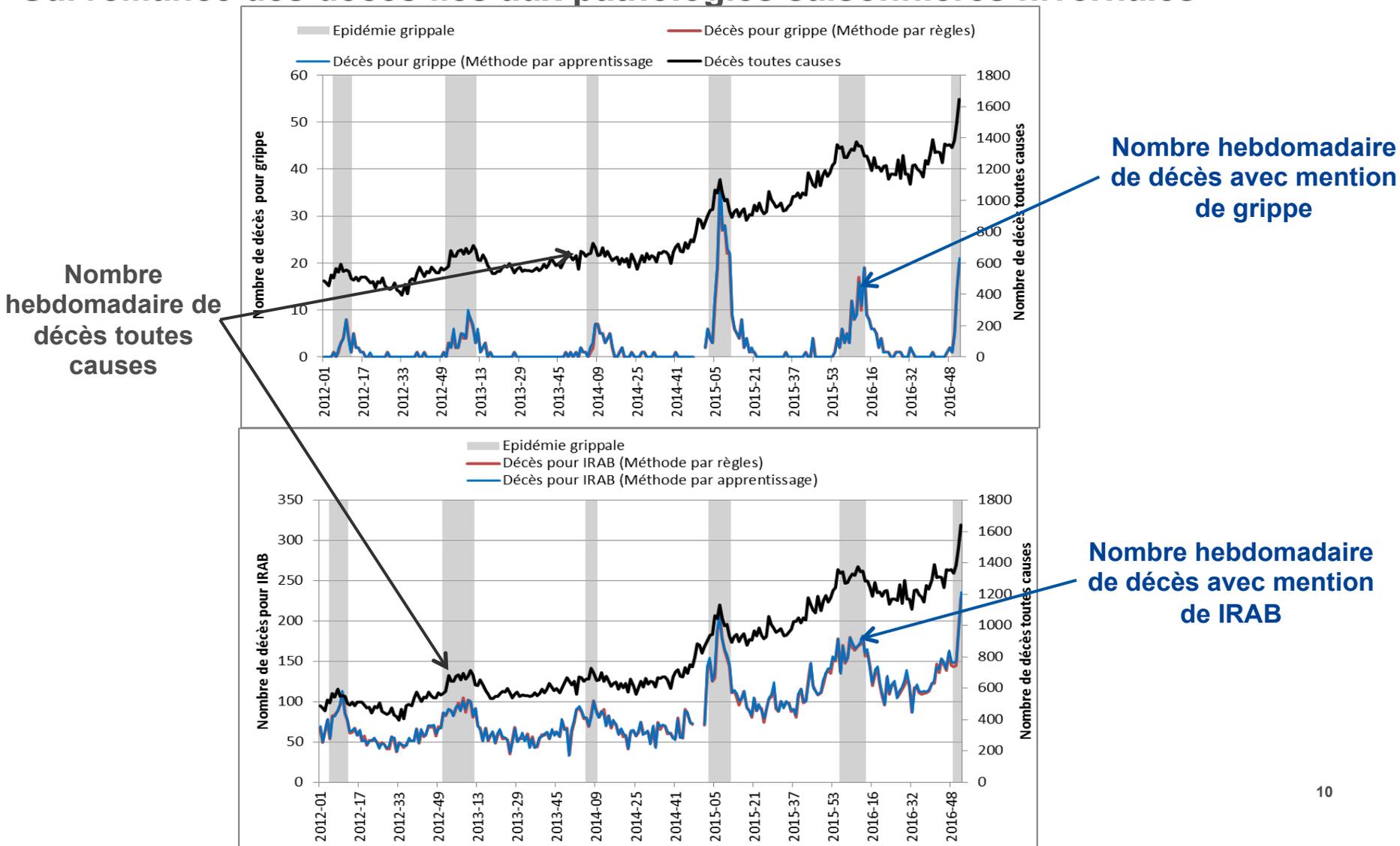
→ **PERFORMANCES ÉLEVÉES (F-MESURE >0,95) POUR 30 RS**

ETAPE 3 : ANALYSE EXPLORATOIRE RÉTROSPECTIVE DES VARIATIONS DE LA MORTALITÉ PAR CAUSES

- Classification des certificats de décès électroniques entre 2012 et 2016 (203 000 certificats)
- Méthodes par règles (rouge)
- Méthode par apprentissage (bleu)
- Analyse des variations hebdomadaires du nombre de décès contenant les RS d'intérêts pour l'alerte

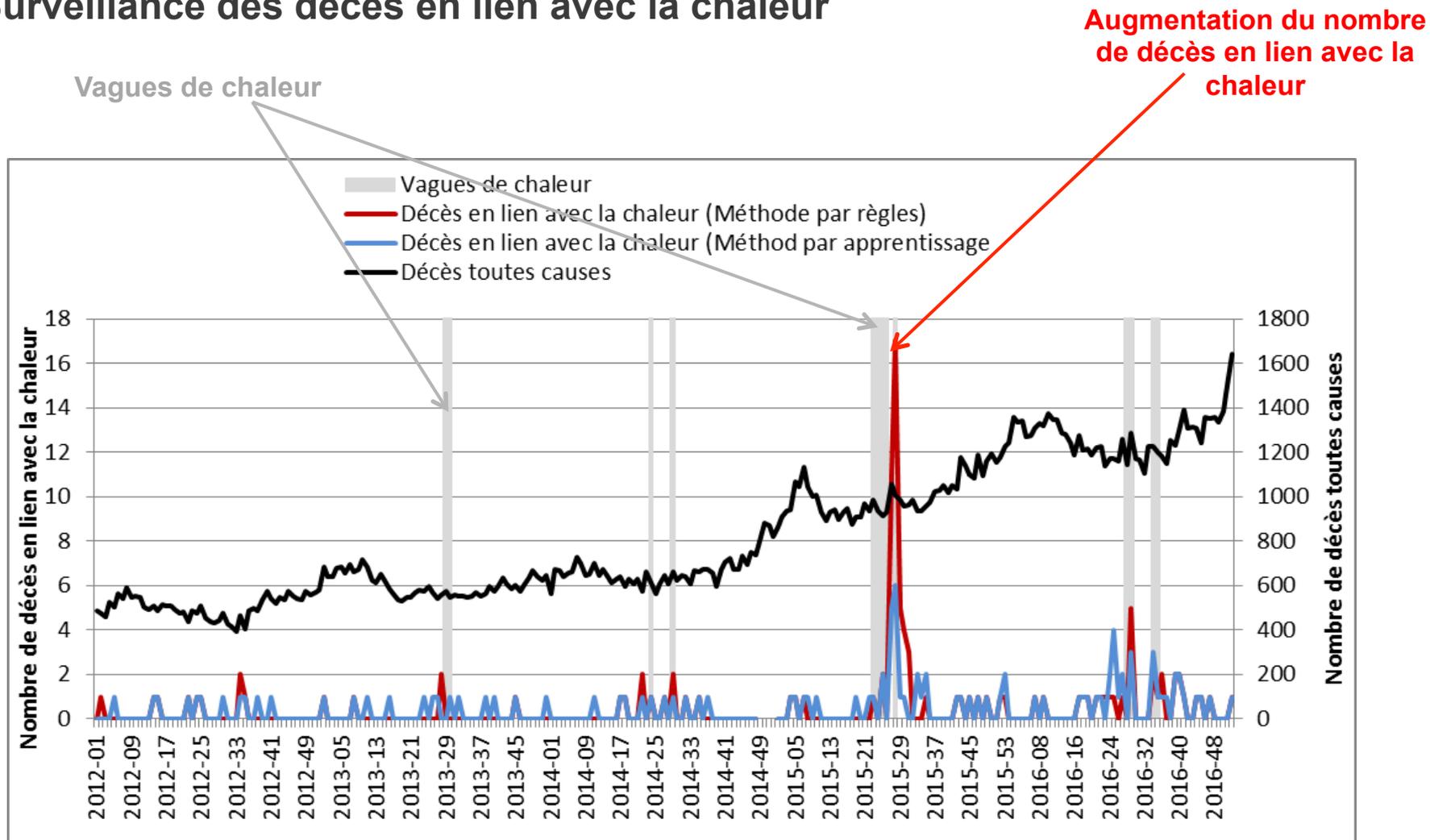
Surveillance réactive de la mortalité à partir des causes de décès : Exemples (1/3)

Surveillance des décès liés aux pathologies saisonnières hivernales



Surveillance réactive de la mortalité à partir des causes de décès : Exemples (2/3)

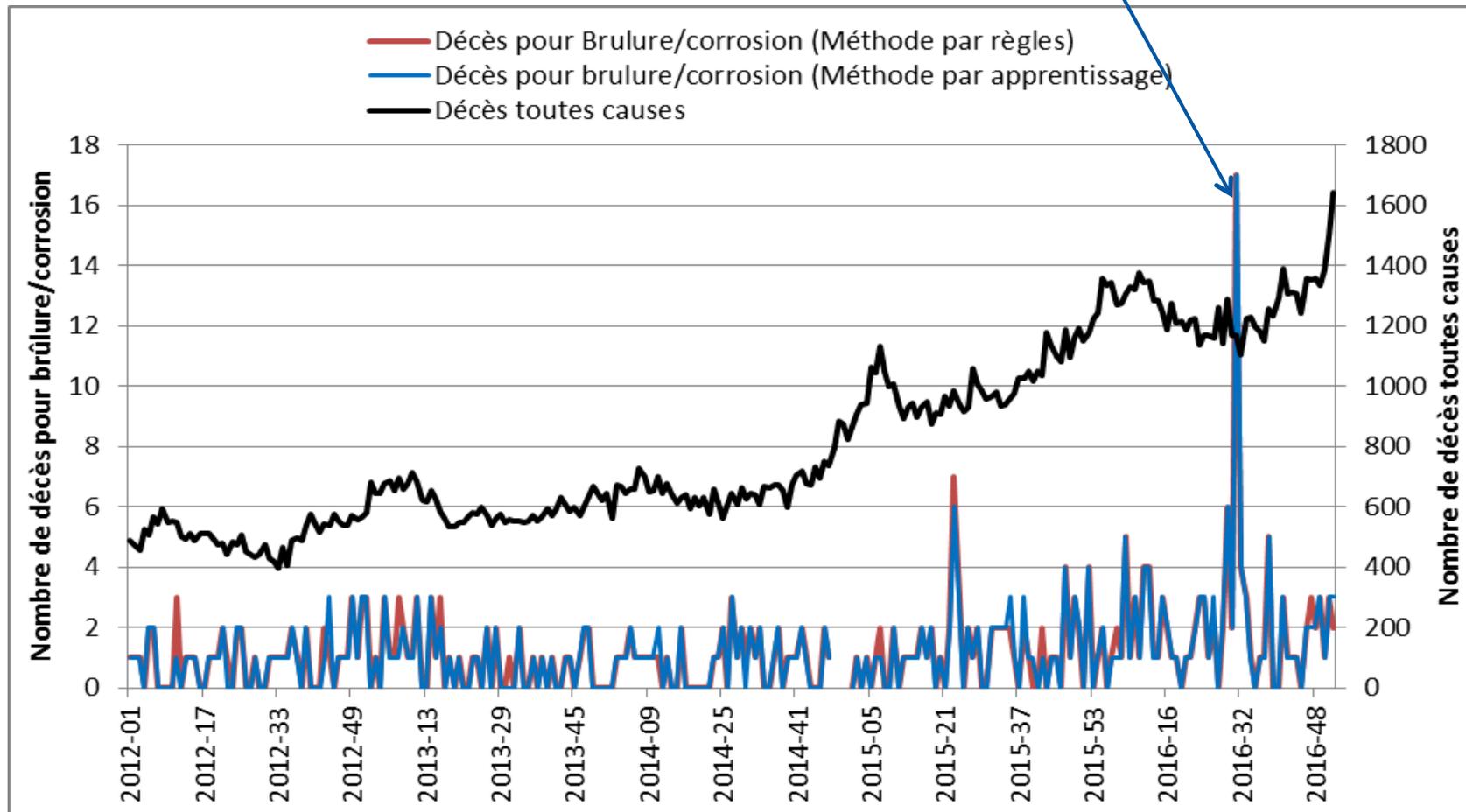
Surveillance des décès en lien avec la chaleur



Surveillance réactive de la mortalité à partir des causes de décès : Exemple (3/3)

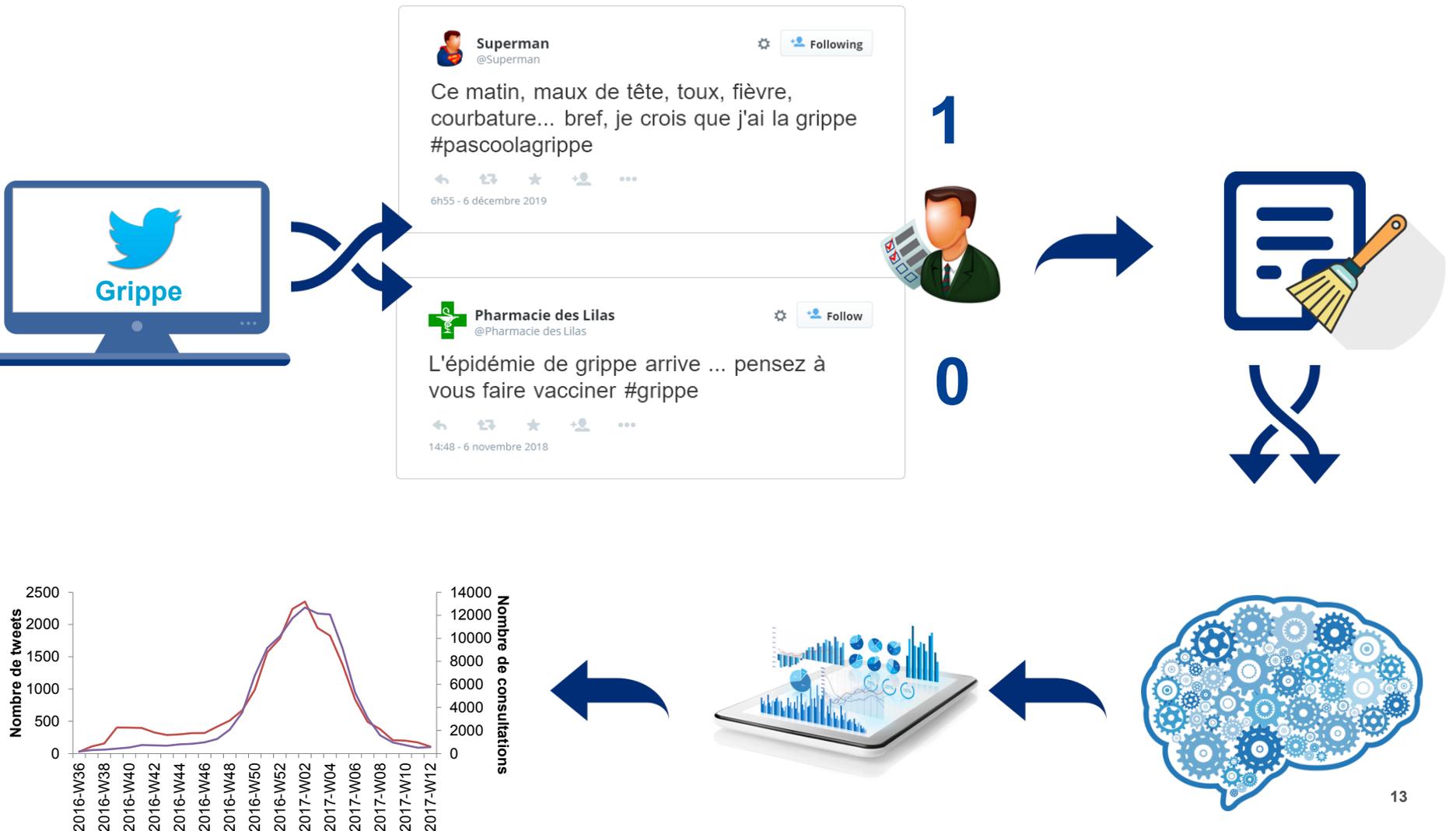
Détection d'évènements inhabituels : Décès en lien avec des brûlures ou corrosions

Détection d'un pic de décès en lien avec des brûlures ou corrosions

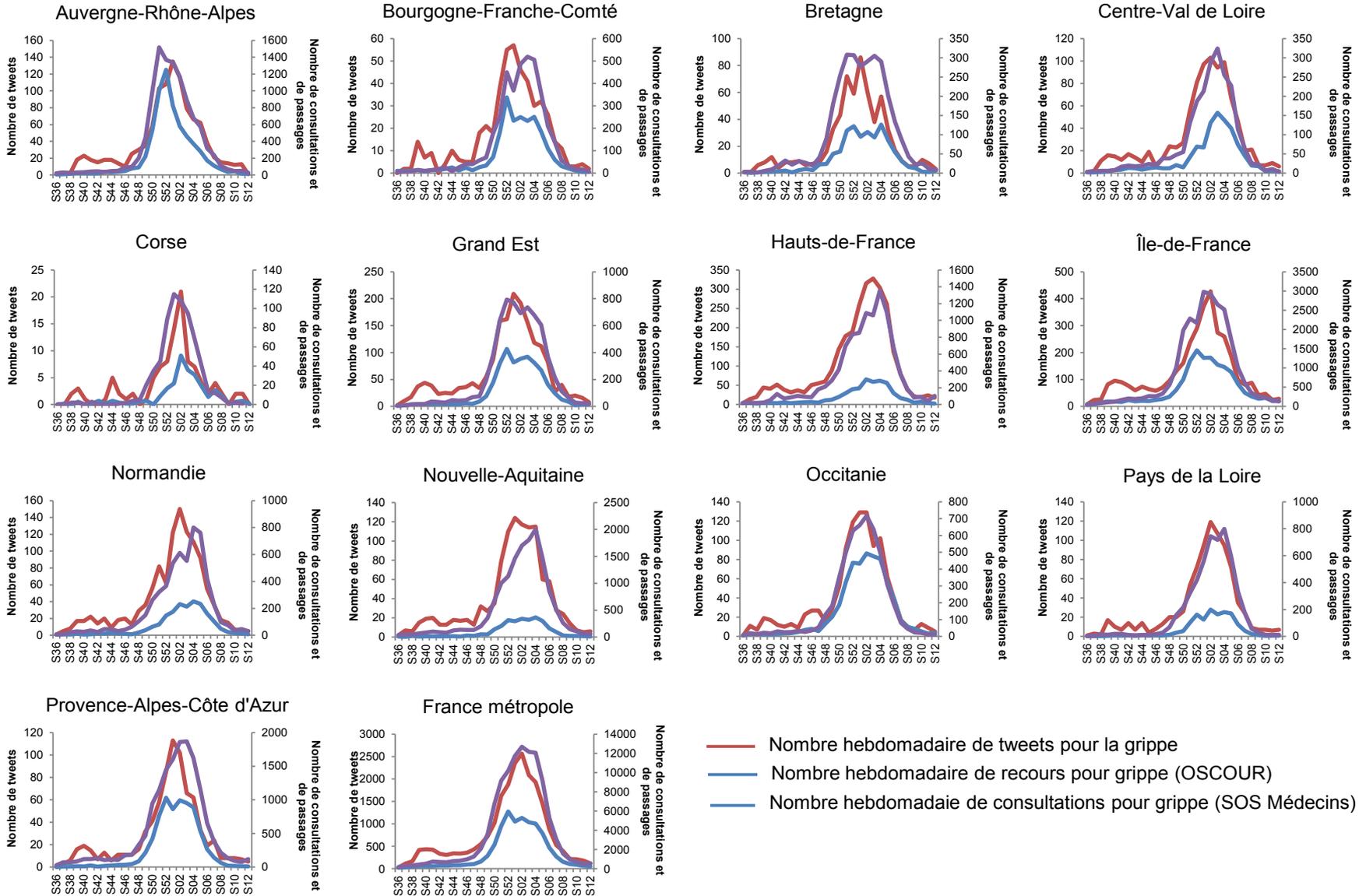


GRIPPE ET TWITTER (1)

ETUDE EXPLORATOIRE



GRIPPE ET TWITTER (2)



- Développement du Data mining en surveillance syndromique : challenge passionnant pour l'avenir
- Sources de données hétérogènes MAIS complémentaires
- Nécessité d'une réflexion sur les systèmes d'information (écosystème capable de stocker des grandes quantités de données, vitesse dans le traitement des données, outil de visualisation, etc.)
- Outil d'aide à la décision pour l'épidémiologiste

**STRATÉGIE PROMETTEUSE DANS LE CADRE D'UNE SURVEILLANCE
RÉACTIVE DE LA POPULATION
(MORTALITÉ, PATHOLOGIES SAISONNIÈRES, ETC.)**