

## Fiche Indicateur n°4 : Analyse de Tweets<sup>1</sup>

### Contexte :

L'Observatoire Régional sur les Risques Majeurs en PACA a été créé en 2014 dans sa nouvelle configuration. Un portail des risques naturels est néanmoins en ligne sur Internet depuis 2012. L'ORRM a fait l'objet d'une présentation en 2015 aux acteurs du risque en Région et a été présenté aux Assises Nationales des Risques à Marseille en mars 2016. Aujourd'hui, cet Observatoire dispose de différents outils (Moi Citoyen, RiskPACA...), et d'une base de données documentaire importante, le tout ayant comme objectif d'améliorer la réponse au citoyen en termes d'information sur les risques.

Dans le cadre de cet Observatoire, différents événements considérés comme significatifs à l'échelle régionale ont été examinés au regard de l'activité qu'ils ont généré sur le réseau-social Twitter par le BRGM (BRGM/RP-66593-FR) et ainsi identifiés comme un indicateur intéressant au niveau régional. Pour ce faire le BRGM a utilisé la plate-forme d'analyse Visibrain qui permet d'assurer une veille continue de Twitter à l'aide d'un système de filtres complet, et d'une interface permettant l'exploration instantanée des données. Ainsi, pour les besoins de cette analyse, des collections de tweets relatives aux 5 événements listés ci-dessous ont été constituées :

- 2014 janvier – Inondation Var (83)
- 2014 novembre – Glissement de terrain à Saint-Blaise (06)
- 2015 – Inondation Cannes (06)
- 2015 – Glissement de terrain Chambon (38-05)
- 2016 – Feux de forêts à Vitrolles (13)

La recherche a été réalisée sur des fenêtres temporelles englobant chaque événement, en utilisant des mots-clés issus d'une déclinaison « grand public » du champs lexical relatif aux différents phénomènes considérés. Par ailleurs, seuls les messages « natifs » ont été retenus en excluant les « retweets », de telle sorte à se concentrer sur les messages signifiants en s'affranchissant du phénomène de caisse de résonance (ou de « buzz ») propre à la viralité des réseaux-sociaux. En plus du contenu textuel des messages/tweets et des attributs associés (date de publication du message, nom de l'émetteur, géolocalisation native lorsque GPS activé, etc.), étaient également disponible une indication de géolocalisation inférée par Visibrain : cette géolocalisation n'est qu'indicative et ne peut seulement être utilisée pour tirer des tendances quant à la provenance des Tweets.

Dans le cadre de cette étude, il n'a pas été fait appel à des techniques d'analyse automatique du langage, et le contenu textuel des tweets n'a donc pas été analysé en détails. Il a au contraire été privilégié une approche visant à identifier des tendances globale et une caractérisation d'événement par l'activité qu'il a généré sur Twitter.

Néanmoins, des nuages de mots (à partir du contenu des Tweets) sont proposés pour chaque collection de Tweets sur chacun des 5 événements. On appelle nuage de mots une représentation visuelle des mots-clefs (tags) les plus utilisés dans un texte ou comme ici une

---

<sup>1</sup> D'après BRGM/RP-66593-FR – Rapport final

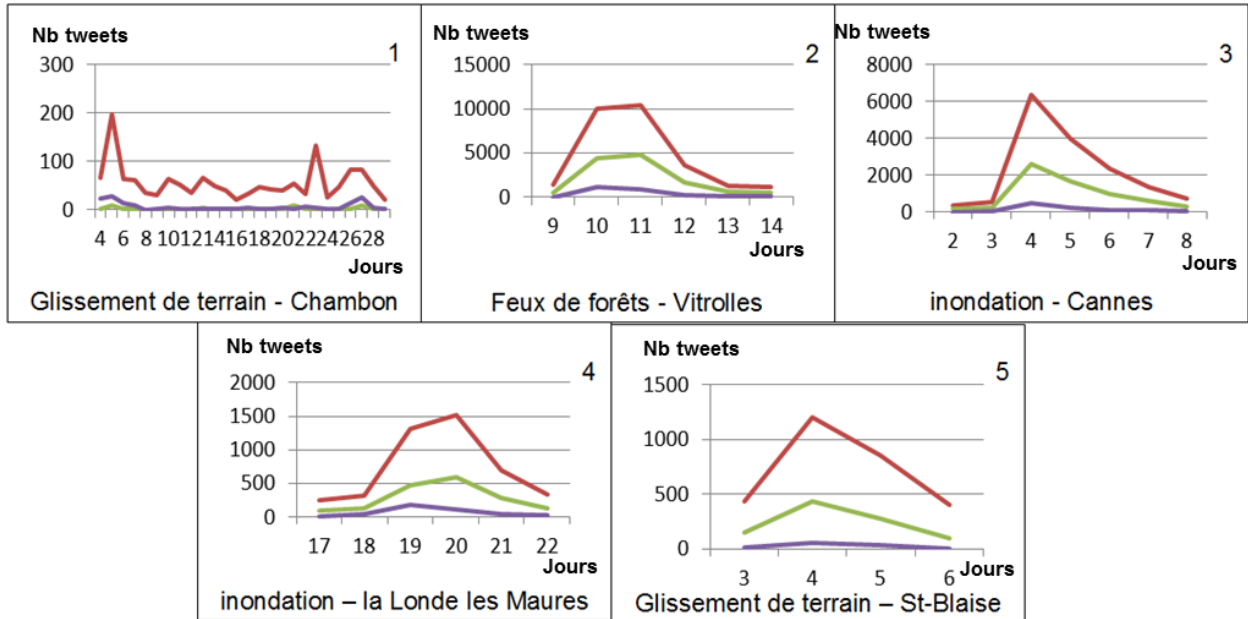
série de Tweets. Généralement, les mots s'affichent dans des polices de caractères d'autant plus grandes qu'ils sont utilisés. Cela permet de se rendre compte rapidement du contenu des messages et de l'importance accordée à un certain nombre de mots.

### Indicateurs pertinents :

1. **Nombre de Tweets sur un événement donné (par mots clefs et sur une période donnée) au regard de la valeur attribuée pour les impacts sociaux sur l'échelle de dommages** produite par le BRGM : cet indicateur montre pour chaque événement étudié le nombre de Tweets émis au total, depuis le territoire Français et régionalement (région PACA principalement et Rhône-Alpes concernant le glissement du Chambon). Cet indicateur montre une certaine cohérence entre l'indicateur « nombre de Tweets échangés » et le niveau de dommages lié aux impacts sociaux pour cet événement. Les Tweets peuvent donc être un bon indicateur pour apprécier cet impact social
2. **Courbe du nombre de Tweets sur un événement donné (par mots clefs et sur une période donnée)** : cet indicateur montre l'évolution du nombre de Tweets émis au total, depuis le territoire Français et régionalement (région PACA principalement et Rhône-Alpes concernant le glissement du Chambon) au cours du temps (nombre de jours), selon la durée de l'événement. On remarque que l'on peut mettre en évidence trois types d'événements :
  - a. Les événements se déroulant sur un temps relativement long (plus d'une semaine) avec différents pics d'activité : Ex. : Le glissement du Chambon étudié sur près d'un mois avec plusieurs « crises » liées au fait qu'il avait été annoncé un déplacement du glissement mettant les riverains, gestionnaires et scientifiques en alerte, ce qui s'est traduit sur Twitter par un pic de Tweets échangés avant la réelle deuxième mise en mouvement fin juillet 2015.
  - b. Les événements se déroulant sur un temps court (quelques jours pour l'événement en lui-même, puis quelques jours supplémentaires pour l'évaluation de ses impacts) avec un palier sans crise nette mais plutôt des échanges de Tweets se maintenant à un niveau élevé pendant 1 à 2 jours : Ex. : C'est le cas par exemple du feu de forêts de Vitrolles en 2016 et des inondations de la Londe en janvier 2014 qui se sont poursuivis sur plusieurs jours.
  - c. Les événements se déroulant également sur un temps court de quelques jours, mais présentant un pic bien de tweets bien visible représentant la phase aiguë de la crise ou étant signe de victimes : Ex. : C'est le cas par exemple pour les inondations à Cannes en 2015 et pour le glissement de terrain au cours des inondations dans les Alpes-Maritimes (St Blaise) en novembre 2014.

On peut remarquer que les événements courts montrent des échanges de Tweets nombreux dans un temps court et que finalement le nombre de Tweets n'est pas proportionnel au temps d'évaluation des impacts qui n'excède pas les 2 à 3 jours sur les exemples présentés ici. Le cas du Chambon est particulier dans le sens où l'événement a beaucoup fait communiquer alors que le glissement de terrain final a été attendu pendant quasiment un mois, d'où des échanges de Tweets sur une longue période.

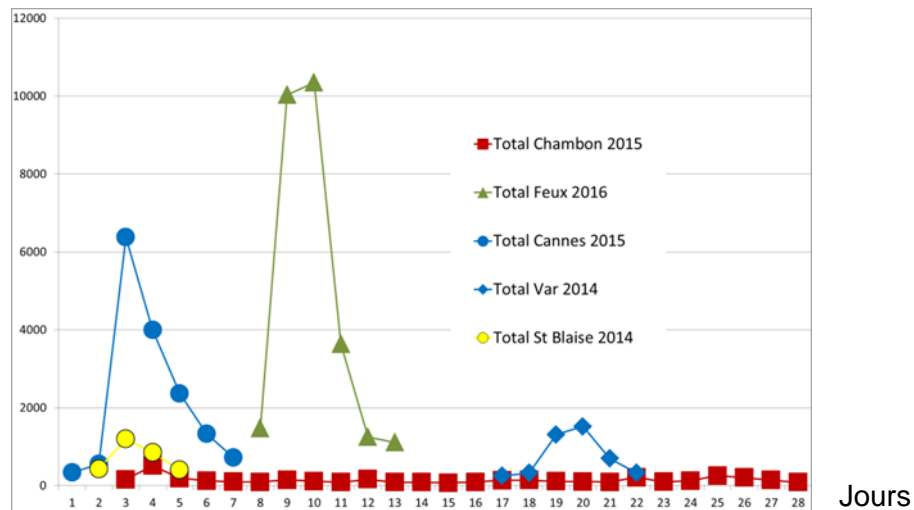
**Graphiques :**



- Total des Tweets
- Total des Tweets émis depuis la France
- Total des Tweets émis depuis la région PACA

*Illustration 1 : Différents profils de graphiques d'échanges de Tweets sur 5 événements significatifs en région PACA*

Nb tweets



*Illustration 2 : Présentation des 5 événements sur un même graphique en nombre de Tweets totaux et calés par rapport au temps origine de survenue de chaque évènement*

Nb tweets

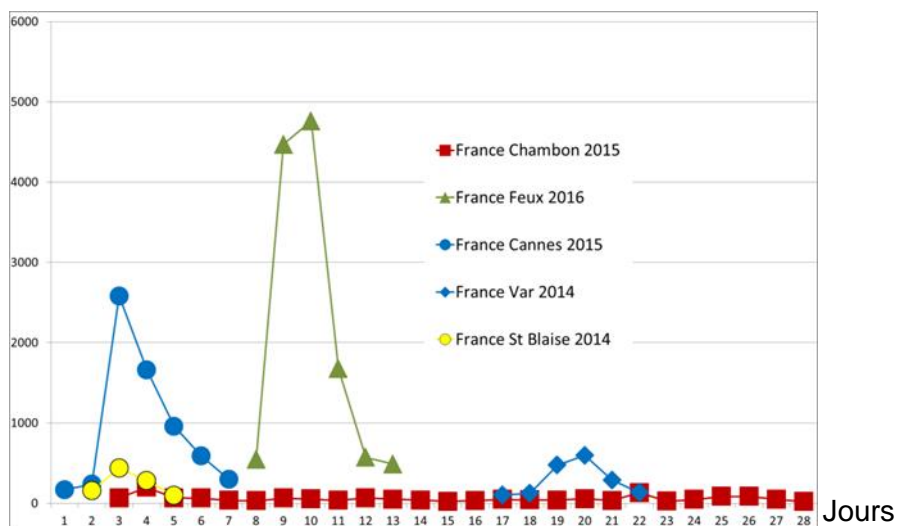


Illustration 3 : Présentation des 5 événements sur un même graphique en nombre de Tweets émis depuis la France, et calés par rapport au temps origine de survenue de chaque événement

Nb tweets

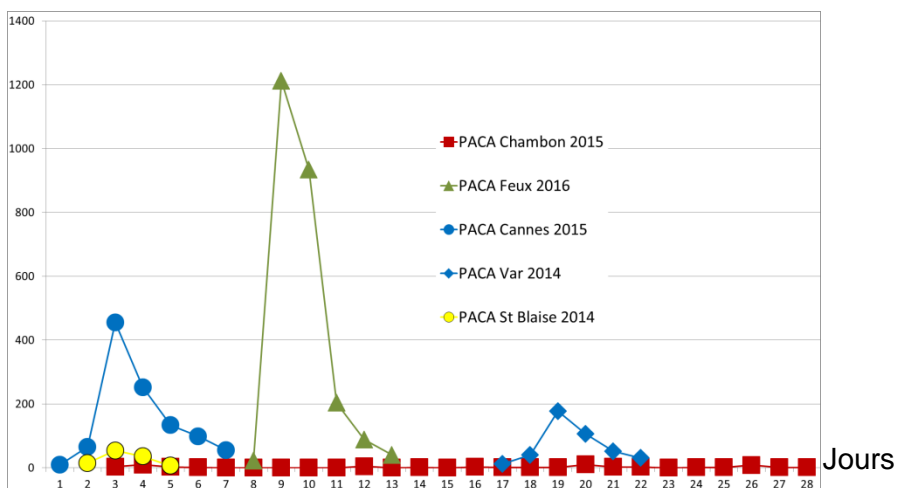


Illustration 4 : Présentation des 5 événements sur un même graphique en nombre de Tweets identifiés comme émis depuis la région PACA, et calés par rapport au temps origine de survenue de chaque événement

## Tableaux :

	Nombre de Tweets			Valeur attribuée pour les impacts sociaux
	Total	France	Paca/Rha	
Cannes 2015	15 685	6 495	1 067	4
Chambon 2015	3 778	1 477	50/170	3
Inondation Var 2014	4 443	1 720	416	3
St Blaise 2014	2 884	974	109	2
Feux 2016	27 853	12 510	2 497	5

Illustration 5 : Tableau représentant le nombre de Tweets échangés pour chaque événement et la valeur attribuée du niveau des impacts sociaux pour chaque événement d'après l'échelle de dommages établie par le BRGM dans l'étude BRGM/RP-66593-FR (juillet 2017), voir grille d'échelle de dommage ci-dessous extraite du rapport final :

Grilles de caractérisation des dommages		Niveaux de dommages					
Types de dommages		1	2	3	4	5	6
<b>A</b>	Dommmages aux personnes	0 blessés	0 mort	1 mort	entre 2 et 10 morts	entre 10 et 50 morts	plus de 50 morts
<b>B</b>	Dommmages fonctionnels	Privé/ Individuel	Communal / Intercommunal	Départemental	Régional	National	International
<b>C</b>	Impacts économiques	Privé/ Individuel	Communal / Intercommunal	Départemental	Régional	National	International
<b>D</b>	Coût des dommages	Moins de 300 k€ (maison)	Entre 300 k€ et 1 Million	Entre 1 Million et 50 Millions	Entre 50 Millions et 500 Millions	Entre 500 Millions et 1 Milliard	Plus de 1 Milliard
<b>E</b>	Impacts sociaux	Privé/ Individuel	Communal / Intercommunal	Départemental	Régional	National	International
<b>F</b>	Impacts environnementaux	Privé/ Individuel	Communal / Intercommunal	Départemental	Régional	National	International (UNESCO)

Illustration 6 : Echelle de dommages proposée par le BRGM dans l'étude BRGM/RP-66593-FR (juillet 2017)

## Nuages de mots :

### 1- Inondation – Cannes (Alpes-Maritimes) – 3 octobre 2015

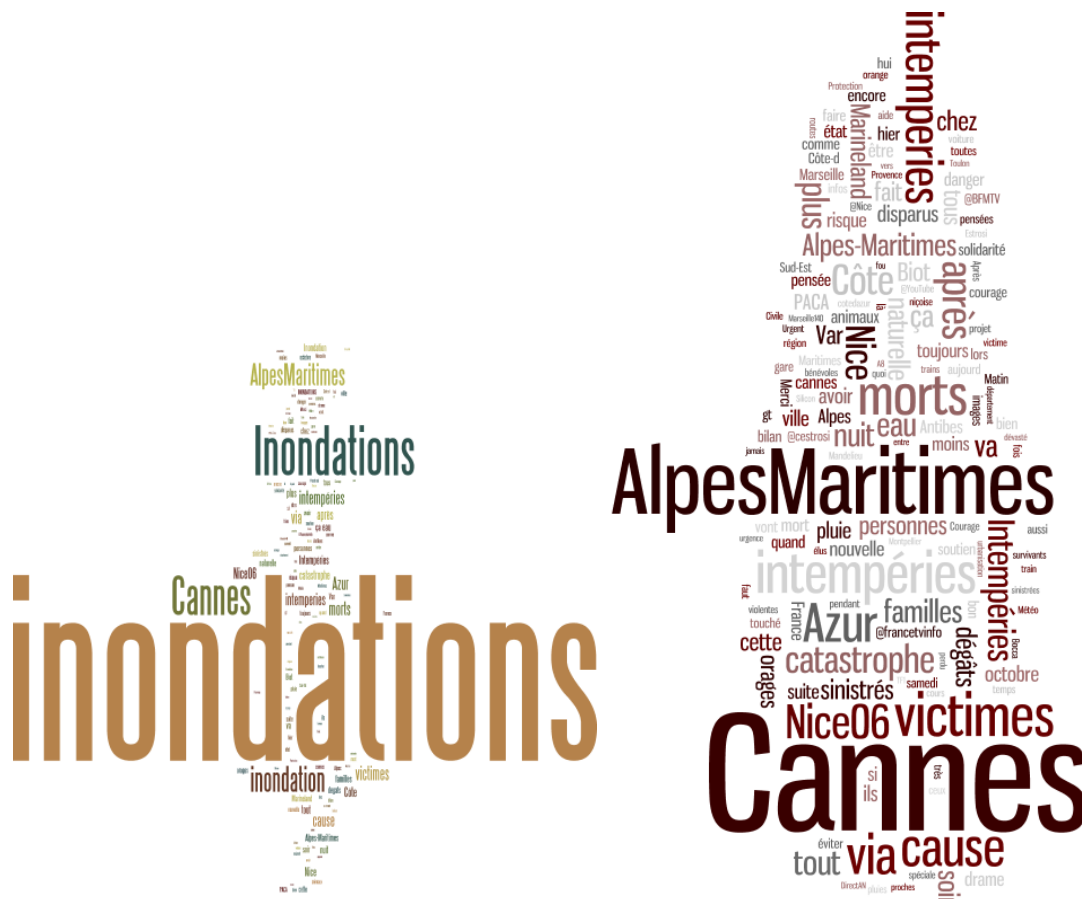


Illustration 7 : Nuage de mots produit à partir des 1067 Tweets émis sur la région PACA (à gauche) après avoir supprimé les mots clés utilisés pour la recherche de Tweets (à droite)

### 2- Glissement de terrain – Chambon (Hautes-Alpes – Isère) – juillet 2015



Illustration 8 : Nuage de mots produit à partir des 170 Tweets émis sur la région Rhône-Alpes (à gauche) et sur la région PACA (à droite)









5- Feu de forêt – Vitrolles (Bouches-du-Rhône) – août 2016



Illustration 12 : Nuage de mots produit à partir des 2497 Tweets émis sur la région PACA (à gauche) après avoir supprimé les mots clefs utilisés (à droite)

## Bibliographie :

**Marçot N. et Arnal C. (2017)** – Analyse multirisques. Production d'une échelle de dommages. Rapport final. BRGM/RP-66593-FR, 86 p., 49 Ill., 2 An.

Sites internet :

Observatoire Régional des Risques Majeurs (ORRM) de la région PACA : <http://observatoire-regional-risques-paca.fr/>

Création de nuages de mots : <http://www.wordle.net/create>

Sites Internet de médias :

<http://www.lemonde.fr>

<http://www.lesechos.fr>

<http://www.francetvinfo.fr>

<http://www.ledauphine.com>

<http://www.lepoint.fr>

<https://www.francebleu.fr>

<http://www.nicematin.com>

<http://www.letelegramme.fr>

<http://tempsreel.nouvelobs.com>