Titre : Utilisation de la donnée issue des sciences participatives et météorologique pour décrire la distribution des morsures et zoonoses liées aux tiques en France.

Hilami Kh, Godard V, Cosson J-F et al.

Abstract

Connaître la distribution et la densité des tiques dans notre environnement, au niveau national, n'est pas une tâche facile, surtout quant on se limite à l’utilisation des seuls atouts des chercheurs universitaires. Les sciences participatives citoyennes ont pris le relais en alimentant des bases de données contenant des rapports de piqûres géolocalisés à l'aide d'une application disponible en format fixe et mobile. En France, depuis juillet 2017, 22.000 rapports concernant l'homme ont été acquis. Pour comprendre les mécanismes de distribution de ces piqûres de tiques dans le temps et dans l'espace, nous avons jugé important de les coupler aux conditions météorologiques au lieu et à la date de la piqûre. Ces déterminants météorologiques résultent de la fouille de vastes entrepôts de données. C'est l'automatisation de cette classification des rapports de piqûres, qui passe par le développement informatique, qui est discutée ici.

Mots clé:

Environnement, Épidémiologie spatiale, Fouille de données, Analyse de donnée, Météorologie, Sciences participatives

# Arrière-plan

## Maladies transmises par les tiques

Les maladies transmises par les tiques sont en constante augmentation en Europe et en Amérique du Nord, sans que l’on sache vraiment si cela correspond à une augmentation de leur incidence, ou à une augmentation de leur détection par les services de santé (Steere et al. 2016; Kilpatrick et al. 2017). Largement répandue dans les régions tempérées de l’hémisphère nord, la borréliose de Lyme (BL), la plus emblématique de ces maladies à tiques, est la maladie à transmission vectorielle la plus fréquente sur l’ensemble de son aire de distribution depuis de nombreuses années (Ward and Brown 2004). Depuis sa mise en évidence chez l’Homme en 1977 (Steere et al. 1977), elle est considérée comme maladie émergente, car son incidence a rapidement augmenté au cours des dernières décennies (Stone, Tourand, and Brissette 2017). Longtemps associée à la ruralité et à ses forêts, la BL semble pourtant poser une nouvelle problématique de santé publique dans les espaces fortement urbanisés (Rizzoli et al. 2014), et ce, en raison d’une forte concentration humaine et de la capacité polarisante des grands massifs forestiers périurbains. Pour autant, la présence des tiques ne se cantonne pas à la forêt, elles sont également rencontrées dans les jardins, les pâtures, les friches… (Mulder et al. 2013; Zeman, Benes, and Markvart 2015). Alors qu’elle est maintenant assez bien connue à l’échelle planétaire, la répartition des tiques à l’échelle régionale et à une échelle plus fine encore, celle des massifs forestiers, par exemple, n’est qu’imparfaitement connue, voire méconnue.

Du rural à l’urbain dense en passant par un périurbain plus ou moins lâche, la présence des tiques est attestée, mais la connaissance des lieux d’interaction avec la population - là où les morsures se produisent – reste à construire. De même, cette morsure se produit-elle au printemps, en été voire en hiver, par temps sec, par temps humide ? Là encore, les connaissances sont lacunaires. Jusqu’à présent, il paraissait difficilement envisageable d’aller plus loin que la simple écologie des espèces de tiques après des campagnes de collectes au drapeau sur la végétation (Cohnstaedt et al. 2012; Bord et al. 2014) ou par collectage sur des animaux abattus à la chasse (Gilot et al. 1994; Kriz et al. 2014) ou bien encore sur des troupeaux domestiques (L’Hostis 1994; Grech-Angelini et al. 2016; Bord and Agoulon 2017), en extrapolant cette infestation aux paysages fréquentés. De plus, dans le cas de la BL qui n’est pas une maladie à déclaration obligatoire en France, malgré la présence de réseaux de médecins volontaires, comme ceux du réseau ‘‘ Sentinelles ’’[[1]](#footnote-2), il est extrêmement difficile de connaître la répartition géographique des cas. Au mieux, c’est à l’échelle régionale, grâce aux cartographies « Sentinelles », qu’il est possible de percevoir les zones et les périodes où la contamination semble se concentrer croître ou décroître. De ce fait, pour faire de la prévention ou du « porté à connaissance » il est quasi impossible à l’heure actuelle d’analyser les déterminants paysagers et météorologiques qui pourraient être les marqueurs d’une situation favorable pour contracter des morsures de tiques, première étape vers une BL ou une autre maladie transmise par les tiques, si celles-ci sont infectées.

Pour progresser dans la connaissance des lieux et conditions de morsures, donc de contamination potentielle, il est indispensable d’accroître le volume des données à traiter en dépassant la seule collecte effectuée par des équipes scientifiques (Laaksonen et al. 2017). Impliquer les citoyens dans la constitution de bases de données sur des questions de santé est maintenant bien documenté (Ranard et al. 2014). Recourir à la contribution citoyenne pour collecter des données portant sur les morsures de tiques est relativement récent. En Europe, les première expérimentations sont attribuées aux Pays-Bas (Garcia-Marti et al. 2018) et consultables sur les sites de Natuurkalender[[2]](#footnote-3) et Tekenradar[[3]](#footnote-4). C’est la solution qui a été retenue pour faire rentrer des gros volumes d’informations sur des situations et des lieux variés qui seuls pourront permettre de faire ressortir les déterminants d’exposition aux morsures et donc progresser vers la connaissance du risque. Actuellement la partialité des données ne permet pas de savoir si, pendant la saison acarologique, les jardins ruraux sont plus ‘‘à risque’’ que les forêts périurbaine, en nombre de morsures constatées par exemple. Elle ne permet pas non plus de connaître s’ils le restent toute la saison ou seulement au printemps, avant la dessiccation liée aux chaleurs estivales, qui se feront moins sentir dans les litières forestières ? Cette progression des connaissances est un préalable à la constitution d’observatoires dédiés aux maladies transmises par les tiques dont l’objectif serait de faire de la prévention ou du ‘‘porté à connaissance’’ à partir de signaux météorologiques avant-coureurs concernant les espaces les plus favorables à l’interaction tiques-humains, ceux où se produisent les morsures.

Les maladies à tiques sont le résultat d’une combinaison complexe de facteurs écologiques et sociologiques. En quelques années elles sont devenues un problème important en santé publique et vétérinaire dans de nombreux pays développés (Vayssier-Taussat, Cosson, et al. 2015; Vayssier-Taussat, Kazimirova, et al. 2015). Elles sont complexes et relèvent typiquement d’une approche dite ‘‘One Health’’, tant la santé humaine, animale et environnementale y sont intriquées et impliquées dans leur épidémiologie (Vayssier-Taussat, Cosson, et al. 2015). Leur impact économique (direct et indirect) est considérable (Lohr et al. 2015; van den Wijngaard et al. 2017). Ces maladies convoient également une très forte charge émotionnelle dans les sociétés occidentales, en Amérique du Nord comme en Europe, où les citoyens s’inquiètent de cette recrudescence et interpellent les gouvernements et les services de santé (Salomon and Vayssier-Taussat 2016). Dans ce contexte anxiogène, plusieurs pays européens comme les Pays-Bas, la Belgique, la Suisse et la France, ont initié des plans nationaux ou des projets de recherche participative (Cosson 2018) dans le but d’impliquer les citoyens pour améliorer les connaissances et la prévention de ces maladies. En France, le projet CiTIQUE[[4]](#footnote-5), initié par l’INRAe, la DGS et l’Anses, lancé le 17 juillet 2017[[5]](#footnote-6) a tout de suite emporté une très forte participation. Les résultats présentés ici s’appuient sur l’implication des citoyens dans la collecte d’informations sur les conditions de la survenue de piqûres de tiques, grâce à l’application « Signalement Tique »[[6]](#footnote-7). Rappelons que l’utilisation d’internet puis des smartphones pour la collecte des données de santé remonte aux années 2000 (Freifeld et al. 2010; Dickinson et al. 2012) et au début des années 2010 pour les signalements de piqûres de tiques (Haddad et al. 2015). La littérature scientifique croisant sciences participatives, collecte de tiques et applications mobiles était encore peu abondante ces dernières années. Quelques exemples commencent d’être disponibles (Bord et al. 2019). Une des toutes premières publications, concernant une application pour animaux de compagnie, essentiellement les chiens, a été proposée en 2018, située aux Pays-Bas (Jongejan et al. 2019) et à destination des vétérinaires.

Pour notre part, sur les trente premiers mois de fonctionnement, entre juillet 2017 et janvier 2020, après nettoyage de la base de données, plus de 22.000 signalements ont été enregistrés, dont plus de 17.000 sur des humains et 5.000 sur les animaux de compagnie. Parmi les plus de 20.000 tiques qui ont été archivées dans les ‘‘tiquothèques’’ de l’INRAe, à la suite d’envois liés au programme de sciences participatives CiTIQUE, plus de 1.400 tiques piqueuses ont analysées à ce jour pour étudier la distribution géographique en France des agents pathogènes qu’elles contenaient[[7]](#footnote-8). Ce n’est pas cet aspect qui sera présenté ici, mais celui de la participation citoyenne aux signalements et surtout l’enrichissement de ces données par de la donnée météorologique pour essayer de comprendre les conditions d’exposition aux piqûres de tiques au travers des conditions météorologiques concomitantes ou préalables aux piqûres.

Si les premières analyses de cette masse de données révèlent la nécessité de mieux comprendre les ressorts de la participation citoyenne pour interpréter les résultats, il est intéressant de compléter et qualifier ces signalements par de la donnée météorologique, l’objet de cette publication, pour faire le lien entre aléa (présence des tiques dans le paysage), et enjeu (présence concomitante des humains dans ce même paysage) pour évoquer un risque d’exposition. En effet, la littérature scientifique nous renseigne sur l’écologie des tiques et en particulier sur celle qui a le plus fréquemment été fournies par les participants lors des signalements, à savoir Ixodes Ricinus dans 90 p.100 des cas pour la France métropolitaine[[8]](#footnote-9). Si les conditions de températures, de précipitations, hygrométriques, etc. des signalements s’écartent significativement des données disponibles dans la littérature, qui, comme le rappellent Ostfeld et Brunner(Ostfeld and Brunner 2015), diffèrent également entre le laboratoire et le terrain, alors nous aurons des information, non seulement sur les conditions d’exposition réelles au risque acarologique sur ces trois dernières années en France métropolitaine, mais également la possibilité de nous projeter vers les conditions futures qui conditionneront nos promenades dans les espaces de nature relativement au changement climatique à venir.

# Méthode

La connaissance de l’aire de distribution d’Ixodes Ricinus s.l. en France est bien décrite et depuis longtemps. Une synthèse reprenant les travaux de Gilot (Gilot, B. 1985), de Gilot et Pérez-Eid (Gilot and Perez-Eid 1998), ou de Pérez-Eid (Pérez-Eid 2007), puis complétée par des travaux du réseau VBORNET[[9]](#footnote-10) et de l’EFSA[[10]](#footnote-11) permet de considérer que peu de parties du territoire métropolitain français échappent à cette tique dure (Fontenille Didier et al. 2013, p.3). Cette synthèse indique qu’à l’exception de zones au-delà de 1 500 m d’altitude et des zones les plus sèches du pourtour méditerranéen, Ixodes Ricinus s.l. est potentiellement partout.

Depuis les années 1980, différents auteurs se sont penchés sur l’évolution de l’aire de répartition des tiques hématophages en général et d’Ixodes Ricinus s.l. en particulier. Une synthèse d’une large part des travaux disponibles permet de constater que cette aire de répartition a tendance à gagner en altitude et en latitude vers le nord (Léger et al. 2013). Il est souvent admis que les principaux facteurs qui influencent cette extension se regroupent en une composante paysagère et une composante climatique (Marsot et al. 2017). Toutes les deux étant influencées par les activités humaines (Medlock et al. 2013). C’est de la composante climatique sur le court terme dont il s’agit dans cet article et en particulier de la collecte des données météorologiques au moment, et aux jours précédant le moment, où le participant a ‘‘collecté’’ la tique responsable de la piqûre détectée. Cette donnée météorologique ponctuelle est à l’échelle des paysages car elle provient du distributeur et agrégateur de données météorologique darksky.net ainsi que de Météo-France, et non pas de capteurs disséminés à hauteur des tiques dans les sous-bois, les parcs ou les jardins. Cette donnée météo doit venir qualifier les données issues des signalements envoyés par les ‘‘collecteurs’’ et fournies par l’application ‘‘Signalement Tique’’[[11]](#footnote-12), que ce soit via un smartphone ou un ordinateur, fixe ou portable.

Les données de signalements consistent principalement en : un identifiant unique généré à la saisie ; des dates (de piqûre, de saisie) ; un couple de coordonnées géographiques (obtenu par un pointeur sur une carte digitale géo-référencée) ; le sexe du répondant et son âge ; le nombre de tiques collectées ; une estimation de la précision du lieu et de la date de la collecte de la tique qui a occasionné la piqûre ; un descriptif du paysage supposé de la collecte (forêt, jardin privé, parc urbain, prairie, zone agricole cultivée ou autre à préciser) ; l’activité qui était pratiquée (activité professionnelle, lieu de résidence, randonnée-promenade-pique-nique, sortie scolaire, scoutisme ou autre à préciser). Un formulaire papier comprenant toutes ces rubriques est également téléchargeable sur le site de CiTIQUE[[12]](#footnote-13).

## La donnée de signalements brute

Le jeu de donnée brut issue des collectes du projet CiTIQUE comportait un nombre de signalements concernant humains et animaux. Ces signalements s’étalent dans une fourchette temporelle allant du 17 janvier 2017 au 04 avril 2020. Les ‘‘saisons’’ de collecte, quant à elles, sont simplifiées selon les trois années principales d’activité, soit : 2017, 2018, 2019. L’analyse directe en partant de la donnée brut s’est avéré impossible. Les formulaires de collecte ayant choisi des séparateurs de donnée ambigus, ainsi que des champs libres où ces séparateurs mêmes, ayant une valeur sémantique[[13]](#footnote-14) interprétée, sont permis, rendait ce traitement directe impossible et compromettaient la structure de la donnée. Une phase de pré-traitement et de nettoyage fut indispensable avant de qualifier les signalements géographiquement et météorologiquement. Un géo-codage inverse[[14]](#footnote-15) a permis de déterminer les localités administratives exactes de chaque lieu de morsure. Les serveurs de géocodage publiques étant sujet à des quotas d’utilisation, il fut nécessaire de compiler un serveur local de géocodage sur une machine sous le système d’exploitation GNU/Linux. Une informatisation de ces processus grâce au langages Python[[15]](#footnote-16) et SQL[[16]](#footnote-17) a permis d’effectuer ces tâche avec précision en tenant compte des irrégularités textuelles de la donnée brute. Dans les deux cas de langages utilisés, un recours intensif aux expressions régulières (Aho 1990) fut nécessaire. Dans le cadre de la politique dite ‘‘Open Data’’ les codes et scripts ayant permis ce traitement sont disponible dans un dépôt à contrôle versionné ou Git[[17]](#footnote-18). Le tableau suivant présente le nombre de signalements brut, pré-traités et nettoyé puis finales et qualifié prêtes à l’analyse :

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Saison de collecte | Humains | | | Animaux | | |
| **PT**[[18]](#footnote-19) | **T** | **NET** | **PT** | **T** | **NET** |
| 2017 | 3586 | 3271 | 2979 | 1920 | 991 | 952 |
| 2018 | 8636 | 7546 | 6412 | 2376 | 2052 | 1914 |
| 2019 | 5516 | 5277 | 5266 | 2096 | 2070 | 2060 |
| Total | 17134 | 16094 | **14657** | 5762 | 5113 | **4926** |

Une fois la donnée géolocalisée, qualifiée et nettoyée, elle est stockée dans une base de donnée relationnelle de type Postgresql augmentée de l’extension spatiale Postgis[[19]](#footnote-20). La qualification par la météo se fait sur la base de la coordonnée géographique exacte de la morsure ainsi que le temps de signalement. Le silo de donnée météorologique darksky.net offre un ensemble d’outils programmatiques (API)[[20]](#footnote-21) permettant d’interroger leurs installations sur les conditions météo survenue à une instant ‘‘T’’ dans une localité ‘‘L’’. Le prog

## Constitution de la base de signalements et qualification par la donnée météo

### Script python

Faire un topo sur la collecte des données DS.net ou Météo France où description du point d’enregistrement (station standard ou données sat ?)

### Base postgredsql

## Choix des paramètres

Les principaux paramètres météorologiques quotidiens sont les suivants : températures enregistrées et ressenties (moyenne, min, max, jour et nuit avec l’heure d’enregistrement) ; couvert nuageux estimé (en pourcentage) ; point de rosée (en degrés C°) ; humidité (en pourcentage) ; précipitations [quantité, intensité moyenne par heure (mm/h), maximum d’intensité (mm/h)…] ; pression atmosphérique en hectopascal (hPa) au niveau de la mer ; indice de rayonnement ultra-violet (plus le temps d'enregistrement du maximum de rayonnement ultra-violet) ; visibilité atmosphérique (en km) ; vent (direction, force moyenne, vitesse des rafales et heure d’enregistrement) et sources des données météorologiques. La liste complète et extensive est disponible sur le site de Dark Sky API[[21]](#footnote-22).

Si les données qui vont être présentées ici sont bien des données qui caractérisent et/ou influencent le comportement humain, elles caractérisent aussi l’interaction entre les tiques à l’affût et les hôtes potentiels qui fréquentent leur biotope. Elles ne permettent probablement pas de reprendre in extenso l’intégralité des données climatiques accumulées dans la littérature scientifique sur les tiques, mais nous pouvons faire l’hypothèse de la transposabilité de certains des paramètres connus comme significatifs, issus de ces articles, pour apporter un éclairage sur les conditions de la rencontre hôte-tique, en fonction de la météorologie de cette interaction ou de celle qui la précède.

Ce sont ici les :

* Variables abiotiques Cat thèse p.175-177 t°, précipitation, humidité, éclairement (Cat 2017; Cat et al. 2017)
* Humidité atmosphérique (Paltridge, Arking, and Pook 2009)
* Climate humidity dans le rapport chercher Willet surface humidity ? (Arndt et al. n.d.)
* Effet des températures et hausse des températures (Greenfield 2011)
* Climatic data et le fameux saturation deficit (SD) (Hauser et al. 2018)
* Using volunteered observations to map human exposure to ticks (collecte de piqûres a posteriori)(Garcia-Marti et al. 2018)

# Analyse

# Resultats

## Cartographie des signalements de piqures

Cf. « Apport de CiTIQUE à la prévention contre les maladies à tique\_v7\_vg.docx » p. 12

Carte de risques ≠ cartes des signalements

Collecte protocolisée

# Discussion

Réchauffement climatique

Températures

Si l’on se base sur l’évolution du climat passé en Ile-de-France, c’est bien une hausse nette des températures qui est décrite dans la littérature. « Sur les 50 dernières années, les températures annuelles ont augmenté d’environ 0,3 °C par décennie. Les cinq années les plus chaudes (2011, 2014, 2015, 2017 et 2003) ont été observées au XXIe siècle. À l’échelle saisonnière, c’est l’été qui se réchauffe le plus, avec des hausses de l’ordre de 0,4 °C par décennie (températures minimales et maximales), suivi par le printemps avec une hausse de 0,3 °C par décennie. En automne et en hiver, les tendances sont également à la hausse mais les valeurs moins fortes : 0,2 à 0,3 °C par décennies. » (Bardis and Hissem 2018). De même, selon la même source, sur la période 1959-2009, le nombre de jours de gel a été réduit de 3 à 4 jours par décennies. Le réchauffement moyen annuel en France est du même ordre de grandeur et s'explique là encore principalement par l'augmentation des températures au printemps et en été (Gibelin et al. 2014). Quels que soient les scénarios dit RCP (Representative Concentration Pathways) du forçage radiatif (Guivarch and Rozenberg n.d.; Moss and Intergovernmental Panel on Climate Change 2008), les projections montrent une poursuite de l’élévation des températures jusqu’en 2050. « En hiver, l’augmentation des températures serait comprise entre 0.6 et 1.3 °C pour le milieu du siècle par rapport à la période de référence 1976-2005 selon les deux modèles climatiques diffués par le portail Drias[[22]](#footnote-23). A la fin du siècle, elle serait de +0.9 à +3.6°C suivant les scénarios. », elle serait plus marquée en été, de l’ordre de +1.3 à +5.3 °C (Aubé, D. 2016) .

Précipitations

En Île-de-France, le signal n’est pas très évident. Il y a bien une légère augmentation des cumuls annuels, mais elle n’est pas aussi nette que l’augmentation des températures. D’un point de vue saisonnier, « les automnes sont légèrement plus secs, tandis que les autres saisons sont légèrement plus humides, mais les variations restent faibles. » avec une forte variabilité interannuelle (Bardis and Hissem 2018). Ceci se traduit par un assèchement des sols au printemps et en été. A l’échelle de la France, une faible augmentation des précipitations est caractérisée sur le quart nord-est de la France assortie d’une diminution dans le sud-est. « A l’échelle saisonnière, les diminutions des précipitations dans le sud de la France sont principalement marquées en hiver et en été. Les tendances à l’augmentation sur le centre et le nord du bassin Rhône-Méditerranée sont principalement en automne (météo France/climatHD). » (Aubé, D. 2016). Selon les projections de Terray et Boé sur la France (Terray and Boé 2013), la diminution des précipitations estivales sera surtout marquée dans le sud. Globalement, elle pourrait être de l’ordre de -0.6mm/jour avec le scénario RCP 8.5, le plus défavorable. Elle serrait en partie compensée par une augmentation modérée en hiver, +0.3mm/jour sous le même scénario.

Additionnal files

Acknoledgements

Availiability of Data

References

1. <https://websenti.u707.jussieu.fr/sentiweb/?page=carte> [↑](#footnote-ref-2)
2. [http://www.natuurkalender.nl](http://www.natuurkalender.nl/) [↑](#footnote-ref-3)
3. <https://www.tekenradar.nl/> [↑](#footnote-ref-4)
4. <https://www.citique.fr/> (consulté le 25/03/2020) [↑](#footnote-ref-5)
5. <http://presse.inra.fr/Communiques-de-presse/Signalement-Tique> (consulté le 25/03/2020) [↑](#footnote-ref-6)
6. <https://www.citique.fr/signalement-tique/> (consulté le 25/03/2020) [↑](#footnote-ref-7)
7. https://www.citique.fr/actualites-2/ (consulté le 25/03/2020) [↑](#footnote-ref-8)
8. Résultats préliminaires présentés aux rencontre naturalistes 2018 <http://www.arb-idf.fr/article/retour-rencontres-naturalistes-2018> (consulté le 25/03/2020) [↑](#footnote-ref-9)
9. http://www.vbornet.eu/ [↑](#footnote-ref-10)
10. https://efsa.onlinelibrary.wiley.com/doi/10.2903/j.efsa.2010.1723 [↑](#footnote-ref-11)
11. https://www.citique.fr/signalement-tique/ (consulté le 25/03/2020) [↑](#footnote-ref-12)
12. https://www.citique.fr/telechargements/download-info/formulaire-papier-signaler-une-et-envoyer-une-piqure-de-tique/ (consulté le 14/04/2020) [↑](#footnote-ref-13)
13. Toute expression interprétée comme ayant une valeur algorithmique et non pas simplement textuelle. [↑](#footnote-ref-14)
14. Attribution d’un contexte spatiale exacte à partir d’une paire de coordonnées géographiques. [↑](#footnote-ref-15)
15. python.org [↑](#footnote-ref-16)
16. postgresql.org [↑](#footnote-ref-17)
17. https://github.com/spicysardine/pycitique [↑](#footnote-ref-18)
18. PT=pré-traitement. T=nettoyage. NET=nettoyé [↑](#footnote-ref-19)
19. postgis.net [↑](#footnote-ref-20)
20. Application Programming Interface [↑](#footnote-ref-21)
21. https://darksky.net/dev/docs (consulté le 15/04/2020) [↑](#footnote-ref-22)
22. <https://darksky.net/dev/docs> (consulté le 15/04/2020) [↑](#footnote-ref-23)