

Questionner la qualité de service mobile avec les mesures issues d'applications tierces (crowdsourcing)

Comment mobiliser les mesures issues d'applications tierces au bénéfice de l'information de l'utilisateur final ?

Afin d'accompagner l'utilisation des données de crowdsourcing mises à disposition sur « Mon réseau mobile » par les consommateurs, les élus ou encore la presse spécialisée, ce document met en lumière les atouts et les limites du crowdsourcing et propose des « cas d'usages » qui ne sont que des exemples. Chacun est invité à s'inspirer de ces propositions pour construire sa propre vision de la connectivité mobile. Il s'agit d'une version mise à jour du document publié en 2022. La méthode retenue par l'Arcep pour représenter ces mesures sur son site cartographique « Mon réseau mobile » est présentée en partie 3.4.

1 Introduction

1.1 Qu'est-ce que le crowdsourcing ?

La production collaborative ou « *crowdsourcing* », dans le cadre de l'évaluation de la qualité de service des réseaux mobiles, correspond au recueil de données par des applications de test. Ces données peuvent résulter soit de tests déclenchés par les utilisateurs sur l'application de test qu'ils ont téléchargée, soit de tests effectués de manière indépendante de l'utilisateur par l'application. Ces tests portent généralement sur des indicateurs caractéristiques de la qualité de service mobile : débit montant ou descendant, latence, etc.

1.2 La démarche de l'Arcep : régulation par la donnée, code de conduite de la mesure de la qualité de service

Réalisées en conditions réelles, les mesures de qualité de service n'offrent pas une vision exhaustive du territoire, mais permettent de connaître de façon précise le niveau de service proposé par chaque opérateur dans tous les lieux mesurés. Depuis 1997, l'Arcep mène, chaque année, une campagne d'évaluation de la qualité des services mobiles des opérateurs métropolitains. Les mesures réalisées visent à évaluer la performance des réseaux des opérateurs de manière comparable, et ce dans différentes situations d'usage (en ville, en zone rurale, dans les transports, etc.) et pour les principaux services utilisés (appels, SMS, chargement de page web, visionnage de vidéo, téléchargement de fichiers, etc.). Cette enquête s'inscrit dans la stratégie de régulation par la donnée de l'Arcep et permet de mieux informer les utilisateurs. Chaque année, plus d'un million de mesures en 2G, 3G, 4G et 5G sont réalisées sur l'ensemble du territoire, dans tous les départements (à l'intérieur et à l'extérieur des bâtiments) et dans les transports (métros, TGV, routes). En 2017, l'Arcep a lancé son outil cartographique et interactif « Mon réseau mobile », qui permet de visualiser les cartes de couverture mobile des opérateurs ainsi que l'ensemble des données de cette enquête de qualité de service. Depuis juillet 2018, les territoires d'outre-mer y figurent également.

Pour autant, ces mesures sont réalisées par échantillonnage, et ne peuvent pas, par nature, donner une vision détaillée de la qualité de service mobile sur l'ensemble du territoire. L'Arcep a ainsi souhaité

compléter cette démarche, en mobilisant la « puissance de frappe » des applications de mesure crowdsourcées.

L'Arcep avait identifié en 2017 le besoin d'une plus grande transparence des méthodologies de mesure. Elle a publié en décembre 2018 une première version du Code de conduite de la qualité de service internet à destination des acteurs de la mesure. Ce Code de conduite porte sur deux aspects : d'une part, inviter les outils à accompagner la publication des résultats par une explication claire des choix méthodologiques réalisés afin que toute personne tierce soit en mesure d'analyser les résultats présentés ; d'autre part, indiquer les bonnes pratiques permettant l'obtention de mesures plus robustes. Cette approche permet d'inciter les acteurs à un niveau adéquat de transparence et de robustesse, à la fois pour les protocoles de test, mais aussi pour la présentation des résultats.

La démarche de co-construction, retenue pour l'élaboration de l'édition 2018 du Code de conduite puis sa mise à jour en 2020, a mobilisé plus d'une vingtaine d'acteurs dont des éditeurs d'outils de mesure en crowdsourcing, des organismes de protection des consommateurs, des opérateurs et des acteurs académiques. Ce Code de conduite comporte deux volets :

- le premier concerne le protocole de test de l'outil de mesure, c'est-à-dire à la fois les méthodologies de mesure des différents indicateurs (débit, latence, temps de chargement des pages web et qualité du streaming vidéo), les mires de test ainsi que les autres tests que l'outil propose ou les informations qu'il communique à l'utilisateur final ;
- le second concerne les publications agrégées, dont un engagement général sur la mise en place d'algorithmes visant à exclure les mesures erronées, manipulées ou non pertinentes. Par ailleurs, pour garantir la représentativité statistique, les outils respectant le Code de conduite s'engagent à publier la période couverte, le nombre de mesures et les facteurs susceptibles d'introduire un biais significatif dans l'analyse des catégories comparées.

Ce Code de conduite met aussi l'accent sur un certain nombre de biais de mesure à expliciter dans les publications agrégées des outils de mesure.

1.3 Les acteurs déclarés conformes au code de conduite, et ceux ayant partagé des données avec l'Arcep

L'Arcep a publié le 14 septembre 2020, l'édition 2020 du Code de conduite de la qualité de service internet. Dès début 2021, plusieurs outils s'y déclaraient conformes. Les outils qui étaient déjà conformes à la version 2018 ont renouvelé leur déclaration de conformité et de nouveaux outils ont manifesté leur intérêt pour rejoindre la démarche de co-construction de l'Arcep. En ce qui concerne la qualité de service mobile, les outils de test qui se sont déclarés conformes à la version 2020 du Code de conduite de la qualité de service internet sont :

- 5GMark, développé par QoSi (groupe Mozark) ;
- Ça capte ? : le testeur de connexion du département du Finistère, développé par QoSi (groupe Mozark) ;
- DébiTest 60 : le testeur de connexion de 60 Millions de consommateurs, développé par QoSi (groupe Mozark) ;
- Gigalis : le testeur de connexion de la région des Pays de la Loire, développé par QoSi (groupe Mozark) ;
- KiCapte : le testeur de connexion du département d'Ille-et-Vilaine, développé par QoSi (groupe Mozark) ;
- LatenceTech : la solution d'analyse du réseau avec un accent particulier sur la latence ;
- nPerf, développé par nPerf ;

- QuelDébit : le testeur de connexion de l'association UFC-Que Choisir, développé par QoSi (groupe Mozark) ;
- Speedtest, développé par Ookla ;
- Tadurezo : le testeur de connexion de la région Bourgogne-Franche-Comté, développé par QoSi (groupe Mozark) ;
- Tu Captes ? : le testeur de connexion de la région Hauts-de-France et les Départements de l'Aisne, du Nord, de l'Oise, du Pas-de-Calais et de la Somme, développé par QoSi (groupe Mozark) ;

La solution de crowdsourcing Tutela, développé par Tutela. D'autres outils de test de débit existent, mais ils ne se sont pas encore déclarés conformes au Code de conduite 2020.

L'Arcep les remercie de leur collaboration, et invite tous les acteurs qui souhaiteraient s'inscrire dans une démarche similaire à rejoindre cette liste.

A titre illustratif, des analyses contenues dans les parties 2 et 3 de ce rapport ont été réalisées en 2021 en s'appuyant sur les données de deux acteurs en particulier : 5GMark (groupe Mozark) et Speedchecker, qui ont accepté de partager avec l'Arcep des données issues de leurs mesures crowdsourcées. Ces résultats sont parfois, dans la suite de ce document, comparés aux résultats issus de la campagne annuelle de mesure de la qualité de service de l'Arcep.

Les résultats présentés sont anonymisés : les deux acteurs de la mesure sont renommés « 1 » et « 2 », et les opérateurs sont renommés « A », « B », « C » et « D ». L'Arcep a en effet souhaité procéder ainsi afin de mettre l'accent sur les possibilités ouvertes par l'analyse des mesures crowdsourcées plutôt que sur les résultats des comparaisons¹.

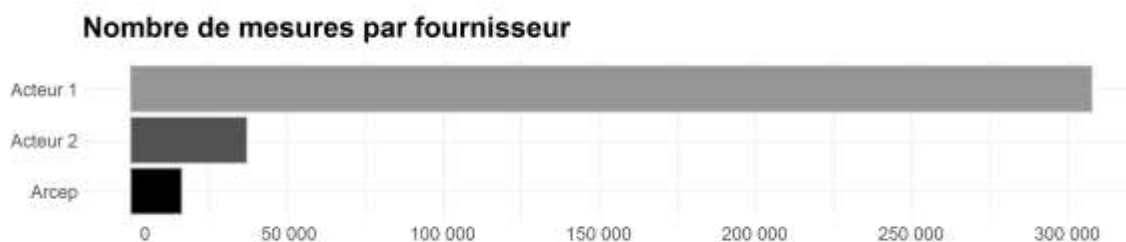
2 Atouts et limites du crowdsourcing

2.1 Le crowdsourcing présente des atouts complémentaires aux mesures dites de « drive-test » habituellement mobilisées par l'Arcep

L'Arcep réalise tous les ans des campagnes de mesure dites « drive-test ». Ce sont des mesures réalisées par des techniciens en voiture, à pied, ou en transport en commun qui vont tester les quatre opérateurs dans des conditions identiques (terminaux et protocoles identiques, au même moment et au même endroit). Il est ainsi possible de comparer objectivement les opérateurs entre eux en un point et un instant donné. Les choix des lieux et des heures de mesure sont en outre effectués pour avoir une répartition la plus représentative possible des usages. Cependant, ces « drive-test », également considérés comme « en environnement maîtrisé », ont un coût important qui limite le nombre de points de mesure et leur répartition géographique. C'est sur ce critère que le crowdsourcing apporte un avantage loin d'être négligeable.

En effet, l'atout principal du crowdsourcing est son nombre important de mesures permis par la sollicitation des utilisateurs ayant téléchargé une application.

¹ Dans la suite de ce document les mesures fournies par 5GMark et Speedchecker qui ont été utilisées pour la réalisation des visuels diffèrent des mesures affichées dans la version actuelle de « Mon réseau mobile », qui sont plus récentes.



Le crowdsourcing permet également plus de diversité en ce qui concerne :

- les terminaux utilisés, avec des modèles à la fois récents et anciens ;
- les heures de test, avec des mesures également tard le soir ou tôt le matin ;
- la localisation des tests, avec des lieux variés et en grand nombre
- les conditions de tests, notamment les mesures dans les espaces de travail ou dans les maisons et appartements.

2.2 Il présente *a contrario* des spécificités qu'il faut prendre en compte lors de l'analyse des résultats obtenus

Bien que le crowdsourcing présentent un certain nombre d'avantages, il présente également des spécificités qui doivent être prises en compte lors des analyses des résultats obtenus.

2.2.1 Les facteurs influençant les résultats

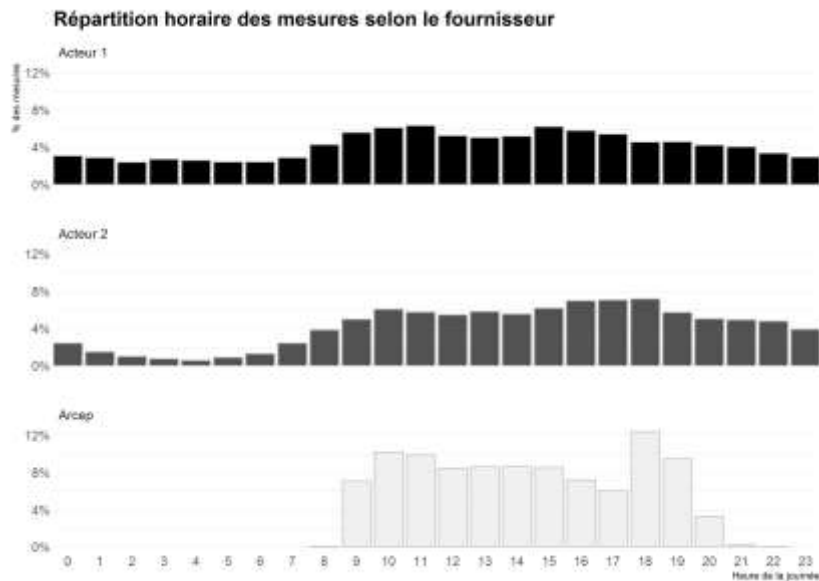
Contrairement aux mesures terrain réalisées dans le cadre des enquêtes de qualité de service, les mesures réalisées par les utilisateurs d'une application ne sont pas en environnement dit « maîtrisé », ce qui complique fortement l'interprétation des résultats ainsi que la comparaison des performances entre les opérateurs. Les facteurs de diversité dans les tests évoqués plus haut engendrent donc mécaniquement un certain nombre de points d'attention pour l'interprétation des résultats, notamment :

- Le mobile utilisé

Les mobiles plus anciens sont généralement moins performants et ne fonctionnent pas sur toutes les bandes de fréquences déployées de nos jours par les opérateurs. L'environnement logiciel (applications s'exécutant sur le mobile, OS et navigateur utilisé sur ce dernier etc.) pourront en outre avoir un impact sur la mesure.

- Les heures de mesure

A un même point géographique, réaliser le test en pleine nuit lorsque le réseau n'est sollicité par personne d'autre pourrait donner de très bons résultats, tandis qu'un test en fin d'après-midi aux heures les plus chargées donnera des performances dégradées.



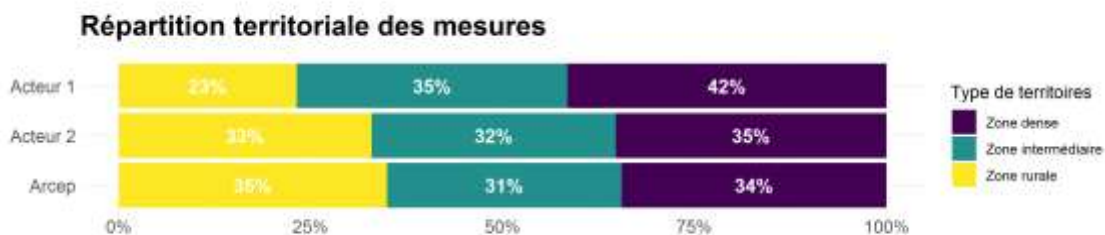
Commentaire : les mesures de l'enquête Arcep (en environnement maîtrisé) sont concentrées sur une plage horaire plus étroite que les mesures issues du crowdsourcing. On voit cependant que les « pics de charge » sont représentés dans chacune des sources de données.

- Les conditions du test

Les performances peuvent varier grandement entre un test en extérieur et un test à l'intérieur d'un bâtiment qui atténue la puissance du signal reçu. Le fait d'être statique ou bien en déplacement dans une voiture ou un train modifie également la qualité du service testée. Ce type d'information sur les conditions de la mesure n'est d'ailleurs pas toujours remonté par l'application de crowdsourcing, ce qui en fait l'un des paramètres les plus compliqués à prendre en compte dans l'interprétation des résultats.

- La localisation géographique

Mécaniquement, le nombre de tests réalisés via les applications de crowdsourcing sera le plus important sur les zones les plus denses. Or, les performances des réseaux en zones denses et en zones rurales sont différentes. Le débit moyen mesuré, si aucune pondération entre les typologies de zones n'est appliquée, sera alors impacté en fonction de la proportion de tests dans une zone par rapport à une autre.



Définitions : la zone dense regroupe l'ensemble des unités urbaines (selon la définition donnée par l'INSEE en 2010) peuplées d'au moins 400 000 habitants (population Insee 2017), la zone intermédiaire regroupe les unités urbaines peuplées de 10 000 à 400 000 habitants, et la zone rurale est définie comme le reste du territoire.

- Le protocole utilisé

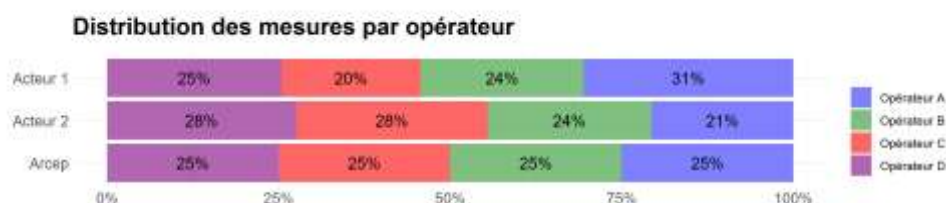
Deux acteurs de crowdsourcing peuvent utiliser deux protocoles différents pour mesurer un indicateur qui, au premier abord, paraît identique. L'exemple le plus marquant concerne le débit montant ou descendant, qui peut être mesuré en « monothread » ou en « multithread ». L'Arcep a fait le choix, depuis plusieurs années, de réaliser des tests de débits en « mono-connexions », c'est-à-dire consistant à ouvrir et à utiliser un seul flux de connexion entre le mobile et le serveur où est hébergé le fichier à télécharger. Il s'agit aujourd'hui de l'usage majoritaire sur internet. Même lorsque plusieurs applications sont ouvertes sur un mobile, en pratique, la plupart du temps, un seul des flux est utilisé à un instant donné. Ce protocole semble donc à ce stade plus proche de l'usage client et permet de révéler les efforts d'optimisation des réseaux des opérateurs en ce sens.

Il est également possible de réaliser des tests en « multi-connexions », qui mettent plus en avant la « capacité maximale » du réseau en termes de débits. C'est ce qui est fait sur de nombreuses applications de crowdsourcing, et cela peut expliquer les différences entre les débits mesurés par l'Arcep et ceux issus de telles applications.

Le paramétrage et le dimensionnement des serveurs de tests peut également avoir une influence déterminante sur les résultats.

- Les biais de sur-représentation

Il peut exister un biais de sur-représentation des cas très défavorables (l'utilisateur fait son test de débit pour « diagnostiquer » sa connexion) ou très favorables (l'utilisateur fait son test pour mettre en valeur sa bonne connexion). Un utilisateur peut avoir le réflexe de réaliser des mesures dans un cas plus qu'un autre. Il est également possible qu'un opérateur soit plus représenté sur une source de données et moins sur une autre :



2.2.2 Les variations de résultats selon les sources de données

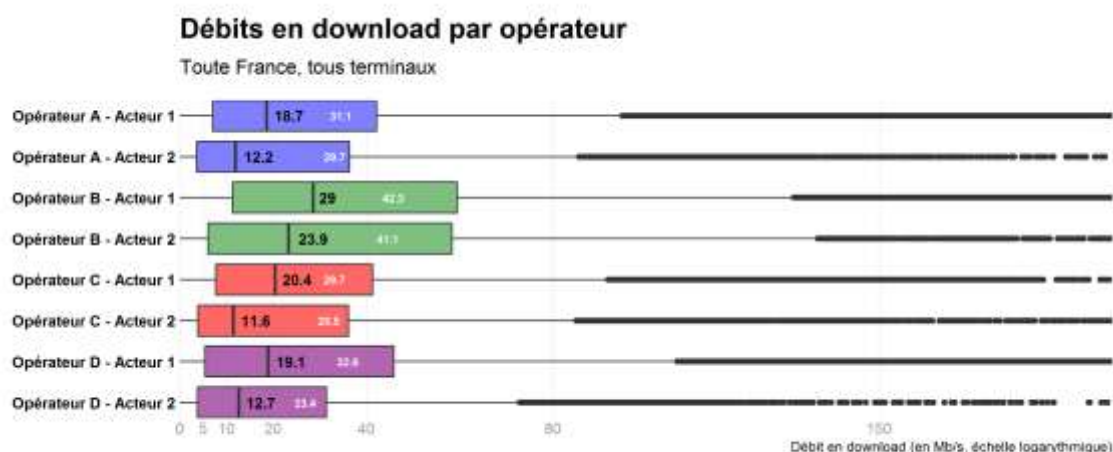
Les mesures de drive-test en environnement maîtrisé, comme le fait l'Arcep tous les ans, permettent de comparer les opérateurs dans des conditions rigoureusement identiques en évitant ces biais. Avec

les données issues du crowdsourcing, la comparaison entre opérateurs est donc plus difficile mais ce dernier permet d'avoir une information sur une plus grande partie du territoire.

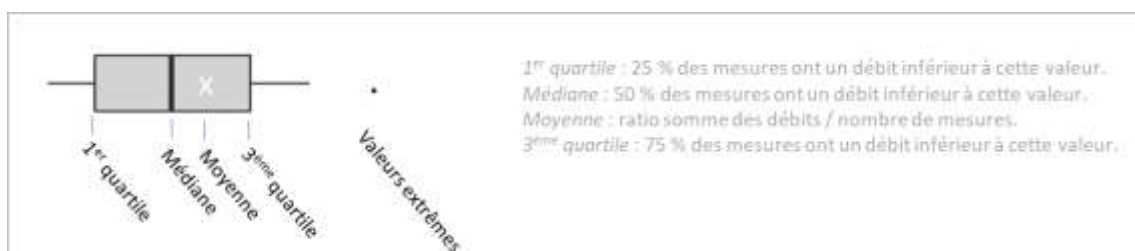
Bien que la répartition des mesures ne soit pas contrôlée, la plupart des acteurs appliquent des corrections et pondérations pour établir des métriques agrégées (par exemple pour obtenir la même répartition entre zones denses et zones rurales d'un opérateur à l'autre).

Concernant les différences de protocole précisés précédemment, l'Arcep ne considère pas qu'un protocole donné est meilleur qu'un autre. Chaque protocole met en avant un résultat différent mais non moins pertinent, présentant ainsi différentes « facettes » de la connectivité mobile. Dans le cas évoqué précédemment du débit en mono-connexion par rapport au multi-connexion, le premier permet de connaître le débit généralement effectif chez l'utilisateur, tandis que le multithread se rapprochera du débit correspondant plus à la « capacité maximale » proposée par l'opérateur ou à un cas d'usage spécifique. Il n'y a donc pas une meilleure façon de mesurer la QoS mobile ou « une bonne source de données » mais « des sources de données qui montrent des informations différentes ». L'enjeu est de bien comprendre l'information qui est présentée.

Avec des protocoles différents et les différents biais listés précédemment il n'est ainsi pas étonnant d'observer des résultats variables entre les sources de données.



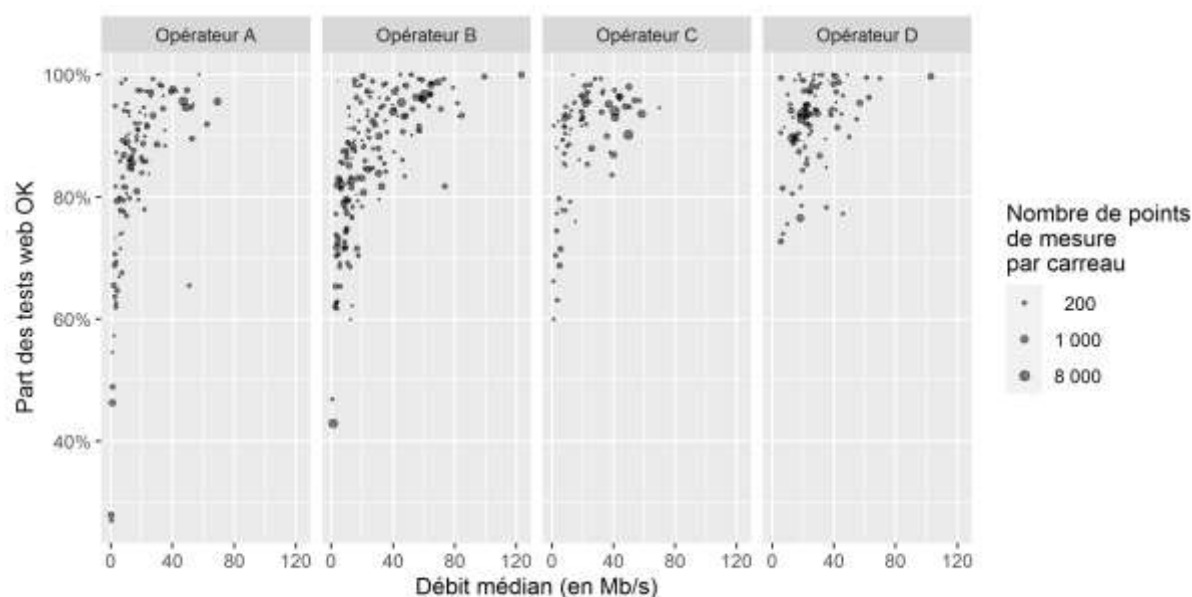
Distribution des débits de 2 opérateurs en fonction de la source des données



Ces résultats agrégés nécessitent une attention particulière pour être interprétés correctement, et en tout état de cause, toute publication en ce sens se doit d'être la plus transparente possible concernant les protocoles de mesures et les méthodologies utilisées.

Il est également possible d'utiliser ce type de données plus localement, sans agrégation, pour savoir si le service d'un opérateur est disponible à un endroit donné. C'est l'objet de la partie 3.

2.3 La corrélation entre le débit descendant et la « qualité d'expérience » (navigation web, streaming)



Distribution du taux de pages web chargées en moins de 10s par rapport au débit médian par carreau d'agrégation

Bien que les mesures crowdsourcing de débit et de web (téléchargement d'une page web en moins de 10s) ne soient pas réalisées au même moment, à des coordonnées géographiques identiques et dans des conditions identiques, on retrouve une corrélation entre débit et web lorsque l'on étudie les résultats en agrégeant les résultats de mesure sur chaque carreau d'agrégation. Plus le débit médian de la zone est élevé, plus le taux de succès des pages web augmente. On note cependant également qu'un débit de quelques Mbit/s peut suffire à naviguer sur le web sans difficulté.

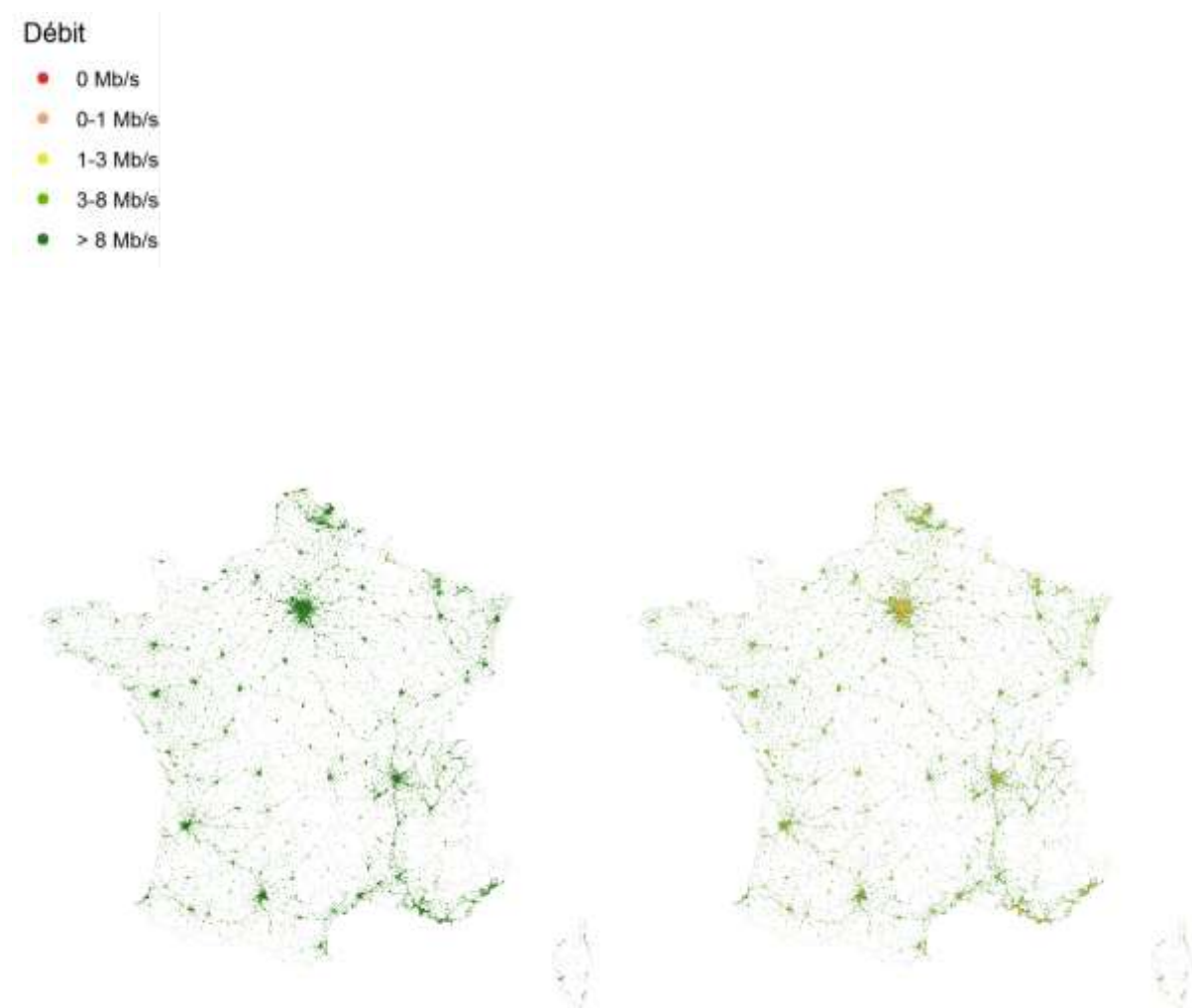
Le nombre important de points de mesure permet de mener des analyses de ce type avec une robustesse statistique qui augmente avec le nombre de données disponibles, et les éventuels retraitements ou corrections qui peuvent être effectués sur les mesures.

3 Quelles utilisations possibles pour les données du crowdsourcing ?

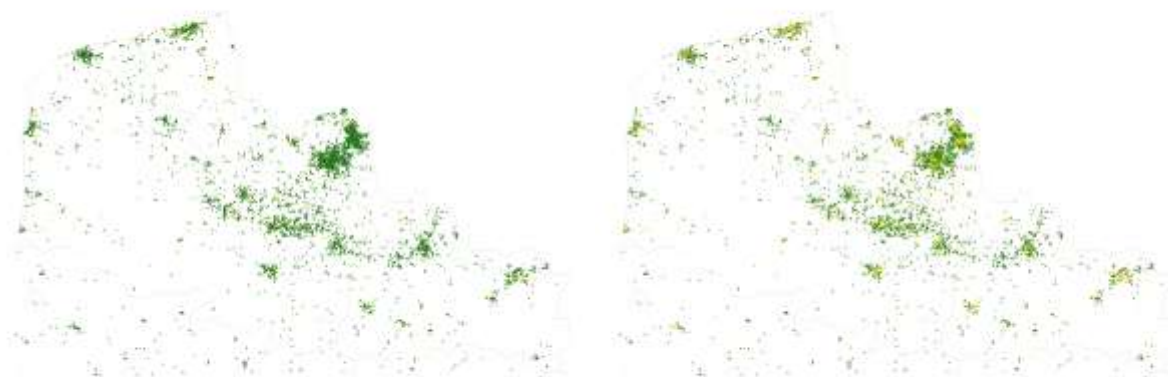
L'Arcep propose dans cette dernière partie des pistes de réutilisation de ces données, afin d'apporter un nouvel éclairage sur la qualité de service des réseaux mobiles.

Chacun de ces exemples est accompagné de commentaires, en particulier des précautions méthodologiques à appliquer lors de l'établissement de ces cartes.

3.1 Une carte représentant le débit descendant mesuré, sans agrégation



Représentation cartographique des résultats de l'ensemble des tests de débit pour un opérateur donné, sur l'ensemble de la France métropolitaine. Les points présentant les résultats les plus élevés sont affichés au-dessus.



Représentation cartographique des résultats de l'ensemble des tests de débit pour un opérateur donné, sur les départements du Nord et du Pas-de-Calais. Les points présentant les résultats les plus élevés sont affichés au-dessus.

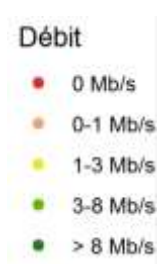
Représentation « point par point »	
Avantages :	Limites :
- Ne transforme pas l'information - Informe sur la densité de mesures	- Ne permet pas une comparaison simple des résultats

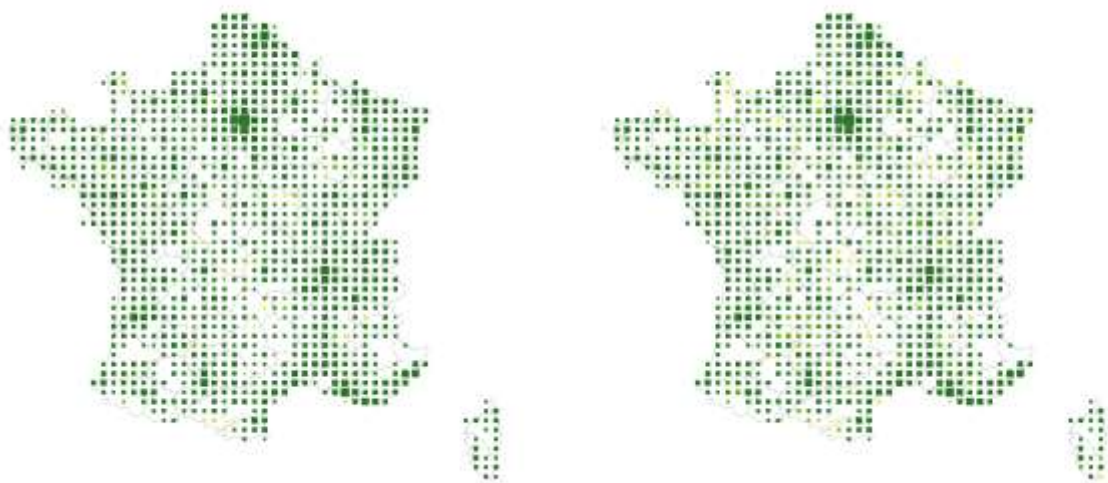
Avantages et limites d'une représentation cartographique « point à point »

Une telle carte montre, pour un opérateur donné, les résultats « bruts » : elle ne transforme pas l'information mais n'est pas particulièrement lisible à des niveaux de zoom faibles. Le principale avantage de cette carte est qu'elle permet de connaître la qualité de service qui a été constatée par un utilisateur à un endroit donné précis.

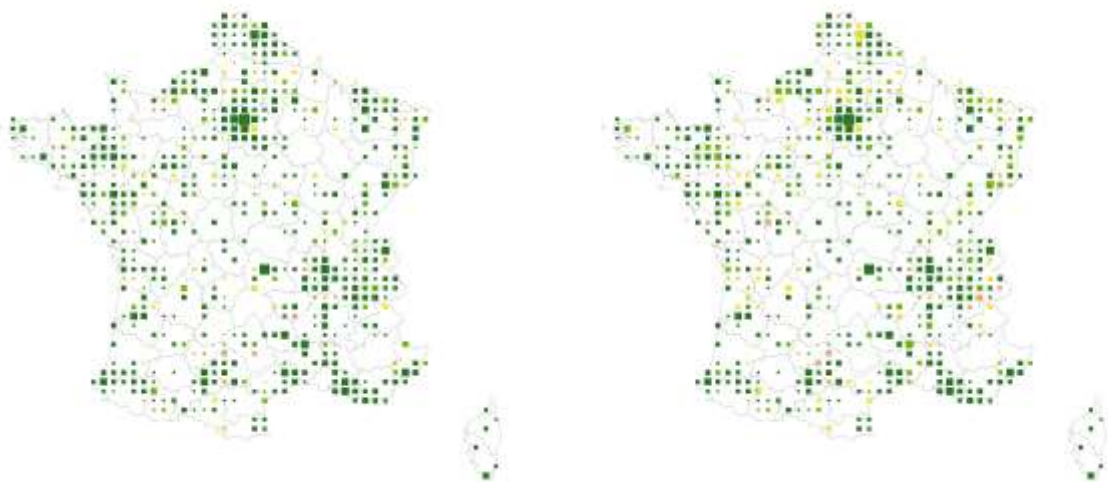
C'est l'approche que l'Arcep avait adoptée, dans une approche expérimentale, pour les premières publications de données du crowdsourcing sur « Mon réseau mobile » dans sa version initiale. La méthode de représentation retenue depuis 2025 est présentée en partie 3.4.

3.2 Une carte représentant le débit descendant mesuré moyen ou médian sur une zone de quelques kilomètres de côté, avec une représentation du nombre de données





Carte des débits moyens (à gauche) et médians (à droite) mesurés par l'Acteur 1 pour un opérateur donné sur des carrés de 22 km de côté. La taille des carrés est proportionnelle au nombre de mesures disponibles dans le carré.



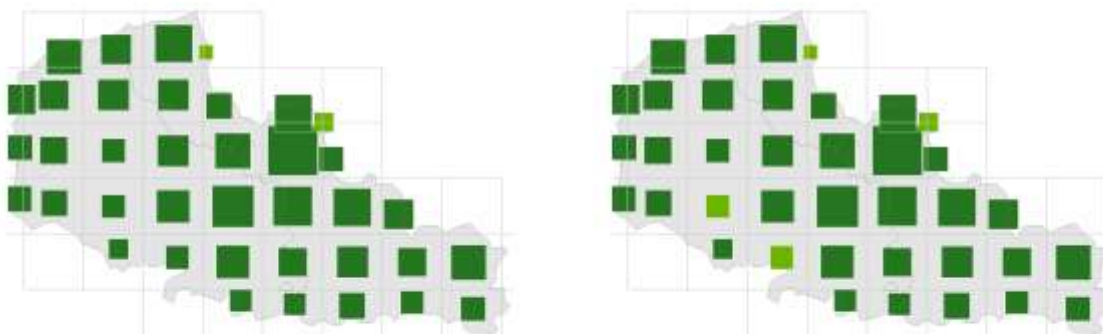
Carte des débits moyens (à gauche) et médians (à droite) mesurés par l'Acteur 2 pour un opérateur donné sur des carrés de 22 km de côté. La taille des carrés est proportionnelle au nombre de mesures disponibles dans le carré.

Représentation « par grille », avec une taille de carré proportionnelle au nombre de mesures	
Avantages : - Permet une comparaison plus simple des résultats - Informe sur la densité de mesures et donc sur la robustesse de l'information présentée	Limites : - Transforme l'information en fonction de la taille de la zone et distorsion de la surface représentée

Avantages et limites d'une représentation cartographique « par grille » avec taille de carré proportionnelle au nombre de mesures dans chaque case de la grille.

L'agrégation des données permet une meilleure visibilité à un niveau de zoom faible et facilite la comparaison entre opérateurs lorsque l'on met les cartes les unes à côté des autres. Cependant, cette agrégation transforme l'information en coloriant toute une zone d'une couleur unique. Il est possible, comme dans l'exemple ci-dessus, de modifier la taille des carrés de couleur en fonction du nombre de résultats dans la zone et ainsi de juger de la robustesse statistique de l'information et de donner à l'utilisateur une idée du « niveau de confiance » qu'il peut avoir concernant l'information.

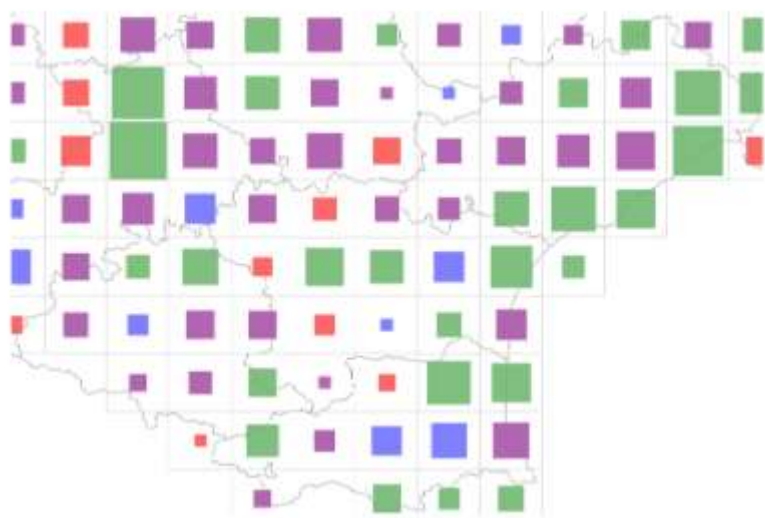
Cette approche soulève cependant une question méthodologique : « Quand considérer que l'on a suffisamment de mesures pour "colorier" un carré ? ». Il est intuitif que colorier un carré alors que l'on a qu'une ou deux mesures sur une zone de plusieurs kilomètres carrés ne donnera pas une information suffisamment robuste. La difficulté est de déterminer un seuil, en fonction de la taille de la zone d'agrégation, qui garantit une pertinence minimale de l'information.



Carte des débits moyens (à gauche) et médians (à droite) mesurés par l'Acteur 1 pour un opérateur donné sur des carrés de 22 km de côté, pour les départements du Nord et du Pas-de-Calais. La taille des carrés est proportionnelle au nombre de mesures disponibles dans le carré.

Il est à noter que le choix de l'indicateur peut changer les observations, même pour deux indicateurs relativement similaires. Par exemple, le débit médian mettra en avant le débit disponible dans 50% des cas, tandis que le débit moyen peut être influencé par des valeurs extrêmes et rares qui vont tirer la moyenne dans un sens ou dans un autre.

3.3 Une carte de l'opérateur proposant le meilleur débit médian sur une zone donnée



Carte représentant l'opérateur proposant le meilleur débit médian mesurés par l'Acteur 1 sur chaque carré de 22 km de côté, zoomé sur le sud de la France métropolitaine. La taille des carrés est proportionnelle au nombre de mesures

Si l'on souhaite connaître rapidement quel est le « meilleur » opérateur dans une région en fonction de l'indicateur étudié, il est possible d'afficher par zone (par exemple, pour chaque carreau de quelques kilomètres de côté) l'opérateur qui aura le meilleur débit médian, débit moyen, débit maximum atteint, ou encore avec le meilleur taux de pages web chargées en moins de 10s par exemple.

3.4 Comment les données de *crowdsourcing* sont-elles représentées sur « Mon réseau mobile » ?

En 2025, l'application « Mon réseau mobile » a été complètement repensée, dans son interface et son fonctionnement. La méthode de représentation des tests de *crowdsourcing* a également évolué.

Les données de *crowdsourcing*, tout comme les données de l'[enquête annuelle de qualité de service mobile](#), sont représentées sous forme d'hexagones sur le site « [Mon réseau mobile](#) ». Chaque test possède différentes classes de résultats (exemple : succès, succès partiel, échec). L'ensemble des tests sont comparés sur une grille d'hexagones de 200 mètres de côté. La classe la plus représentative de chaque hexagone est représentée en son centre comme dans l'exemple ci-dessous.



Exemple de sélection de la classe la plus représentative au sein d'une grille hexagonale. Cas d'un test à deux classes (succès, échec)

Le détail de la méthode d'agrégation des résultats par hexagone est décrit dans la [fiche pédagogique sur la représentation des tests de qualité réseau](#).

4 Conclusion

L'Arcep poursuit sa démarche de « régulation par la donnée », en enrichissant régulièrement les données disponibles pour renforcer l'information mise à disposition des utilisateurs.

L'Arcep remercie les acteurs qui ont accepté de partager leurs données avec elle. Elle invite les autres acteurs intéressés à s'inscrire dans cette même démarche. Elle appelle également les utilisateurs à se saisir des applications de crowdsourcing, aussi bien pour réaliser leur « diagnostic » de la connectivité mobile que pour enrichir « Mon réseau mobile » avec les mesures qu'ils réaliseront.