

Observer le Réseau Électrique: Mesurer la fiabilité de l'électricité à Accra (Ghana), indépendamment des fournisseurs d'électricité

Noah Klugman Joshua Adkins Emily Paszkiewicz Molly G. Hickman
Matthew Podolsky Jay Taneja Prabal Dutta

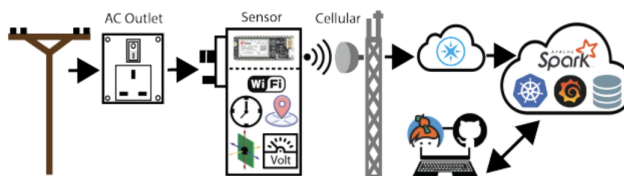
2021

Note du Traducteur Ce document est une traduction française de deux sections d'un article publié en anglais. La version originale de l'article, dont l'intitulé est «Watching the Grid : Utility-Independent Measurements of Electricity Reliability in Accra, Ghana », est disponible ici.

4. Système PowerWatch

PowerWatch consiste en des capteurs que l'on branche afin d'obtenir un ensemble de mesures relativement simples - la mesure première étant d'identifier si les capteurs sont alimentés en courant - et d'un algorithme de détection des pannes qui combine et analyse les mesures des capteurs afin d'obtenir des informations importantes sur les performances du réseau électrique. Dans cette section, nous discutons les mesures de performance utilisées afin d'évaluer un réseau, les conditions auxquelles un système doit répondre pour collecter ces mesures, ainsi que le système qui répond à ces conditions.

FIGURE 1 – Architecture du système PowerWatch. PowerWatch mesure le réseau en se branchant sur les prises de courant des particuliers ou des entreprises, puis en transmettant des données sur la qualité de l'électricité via le réseau cellulaire; ces données sont alors partitionnées en fonction des caractéristiques temporelles et spatiales des coupures de courant.



4.1 Estimations clés fournies par PowerWatch

Les fournisseurs d'électricité quantifient généralement la fiabilité du réseau en calculant l'indice de durée moyenne d'interruption du système (SAIDI) et l'indice de fréquence d'interruption moyenne du système (SAIFI); ces indices

FIGURE 2 – Zone de déploiement de PowerWatch. Les capteurs sont déployés dans trois des 26 districts d’Accra. Le déploiement couvre une zone d’environ 130 kilomètres carrés.

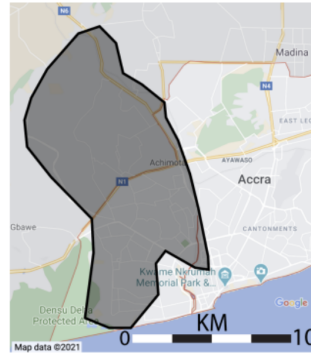


FIGURE 3 – PowerWatch déployé. (a) PowerWatch PCB (circuit imprimé) avec la radio cellulaire, la carte SD, et les circuits de détection. (b) Capteurs PowerWatch assemblés, avec un QR code scanné au moment de l’installation pour associer le capteur à un participant (c) Un agent de terrain installe un capteur PowerWatch sur une prise de courant domestique.



sont définis dans l’équation 1 et l’équation 2 [11]. Ces indicateurs sont calculés pour une unité de temps (le plus souvent mensuelle) ; le nombre de compteurs électriques desservant une zone est souvent utilisé pour définir le nombre de consommateurs [2].

$$\text{SAIDI} = \frac{\text{Durée totale des interruptions prolongées}}{\text{Nombre total de consommateurs impactés}} \quad (1)$$

$$\text{SAIFI} = \frac{\text{Nombre total d’interruptions prolongées}}{\text{Nombre total de consommateurs impactés}} \quad (2)$$

Le calcul précis du SAIDI et du SAIFI nécessite des informations sur la performance du réseau (le numérateur), la configuration électrique sous-jacente et la composition de la clientèle (le dénominateur). PowerWatch améliore l’estimation du numérateur ; améliorer la précision du dénominateur est un projet futur. Au Ghana, le dénominateur est difficilement défini en raison du manque d’informations précises sur l’infrastructure du réseau et sur les clients. Ces lacunes en matière d’information ne sont pas rares ; de nombreux projets sont en cours dans le monde entier pour cartographier les infrastructures et les clients ; ceci permet d’améliorer les opérations des fournisseurs d’électricité

et également les mesures d'indicateurs de fiabilité. [7, 8, 19].

Pour quantifier la performance du réseau sans estimer le nombre de clients impactés par une panne détectée, nous définissons deux nouveaux indices : Sous-échantillon SAIDI (S-SAIDI) et Sous-échantillon SAIFI (S-SAIFI), comme indiqué dans les équations 3 et 4.

$$\text{S-SAIDI} = \frac{\text{Durée totale des interruptions prolongées dans le sous-échantillon}}{\text{Taille totale du sous-échantillon}} \quad (3)$$

$$\text{S-SAIFI} = \frac{\text{Nombre total d'interruptions prolongées dans le sous-échantillon}}{\text{Taille totale du sous-échantillon}} \quad (4)$$

Nous constatons que lorsque la taille du sous-échantillon augmente et devient plus proportionnelle à la densité de la population, S-SAIDI et S-SAIFI se rapprochent de SAIDI et SAIFI.

4.2 Configuration requise

Les objectifs principaux de PowerWatch sont de mesurer S-SAIDI, S-SAIFI, ainsi que d'autres indicateurs de la fiabilité du réseau, tout en restant indépendant des fournisseurs d'électricité. Ces conditions guident les décisions de conception suivantes.

4.2.1 Déploiement

Afin de rester indépendant des fournisseurs d'électricité, nous ne pouvons pas compter sur eux pour installer les capteurs sur les infrastructures électriques ; cela pourrait biaiser l'échantillonnage si le fournisseur ne met qu'une partie de l'infrastructure à disposition [5]. Au lieu de cela, nous concevons nos capteurs de manière à ce que des non-experts puissent déployer et déboguer ces derniers, ce qui nous permet d'installer des capteurs chez des clients que nous sélectionnons.

4.2.2 Quantités détectées

Les capteurs PowerWatch doivent détecter la panne de courant pour calculer S-SAIFI, et doivent également détecter le rétablissement du courant pour calculer S-SAIDI. Pour saisir avec précision le rétablissement de l'alimentation, le capteur doit être capable de traquer le temps sans être alimenté par le réseau. Pour tous les horodatages, le capteur doit conserver une résolution temporelle en secondes. Nous supposons que cette résolution est suffisamment rapide pour observer le comportement du réseau (les pannes ont un impact sur le réseau de l'ordre de quelques minutes [22]). Les capteurs doivent également indiquer leur emplacement à quelques dizaines de mètres près pour permettre à PowerWatch d'estimer l'étendue d'une panne sans s'appuyer sur des cartes de l'infrastructure du réseau sous-jacente. Enfin, les capteurs PowerWatch doivent détecter la tension et la fréquence du réseau, caractéristiques souhaitées par les parties intéressées.

4.2.3 Communication

Étant donné que les capteurs PowerWatch ne seront pas nécessairement collectés à la fin du déploiement, et que les données qu'ils collectent pourront être utilisées en temps réel à l'avenir, un capteur doit disposer d'une connexion fiable au réseau étendu, avec une capacité mesurée en mégaoctets par mois. Cette connexion est utilisée pour collecter des données, pour suivre la fiabilité du système, et pour effectuer des mises à jour du micrologiciel. De courtes coupures de réseau sont tolérables car les données peuvent être stockées localement et envoyées lorsque le réseau est rétabli.

4.2.4 Méthodologie de détection

Pour rester indépendant des fournisseurs d'électricité, nous déployons des capteurs PowerWatch dans les foyers et les entreprises. Nous déployons les capteurs avec l'objectif que tous les capteurs d'un site soient desservis par le même transformateur. Ceci est difficile à garantir en raison de la forte densité du réseau à Accra.

Alors que le déploiement dans les maisons et les entreprises des clients introduit beaucoup de bruits dans les données - les participants débranchent les capteurs, les compteurs individuels prépayés n'ont plus de crédit, et des générateurs rétablissent artificiellement le courant - nous supposons qu'avec un filtrage minutieux, nous pouvons, depuis nos données, extraire des schémas qui nous assurent qu'un capteur fait partie d'une véritable panne, et que le capteur mesure approximativement à la fois l'étendue spatiale de la panne et le niveau de tension du réseau auquel la panne s'est produite.

Plus précisément, nous recherchons des changements spatiaux et temporels dans l'état de l'alimentation sur deux capteurs ou plus pour classer un événement comme une panne, et nous déployons trois capteurs sur chaque site afin que les pannes puissent toujours être détectées dans l'éventualité où un seul capteur tomberait en panne. En exigeant un partitionnement spatio-temporel avant de classer une panne, nous filtrons le bruit créé par le fait que les capteurs sont auprès des utilisateurs finaux. Cependant, en ne considérant pas les rapports d'un seul capteur comme de véritables pannes, nous rejetons les petites pannes qui n'affectent qu'un seul capteur.

Pour déterminer à quel endroit de la hiérarchie du réseau une panne s'est produite, nous mesurons le nombre de capteurs qui observent une panne. Dans le cas d'une grappe de seulement deux ou trois capteurs, le point de défaillance se situe très probablement dans l'infrastructure à basse tension. Dans le cas d'une panne couvrant plusieurs sites de déploiement, le point de défaillance se situe probablement à un niveau supérieur du réseau.

4.4 Architecture

L'architecture PowerWatch, illustrée figure 1, consiste en : (1) un capteur de détection des pannes déployé dans les habitations et les entreprises qui sont clientes des fournisseurs d'électricité, (2) un réseau cellulaire relié au cloud, et (3) d'une analyse de données basée sur le cloud qui partitionne les rapports de plusieurs capteurs en pannes. Le personnel de terrain aide au déploiement et à la gestion.

4.4.1 Capteur

Le capteur PowerWatch, illustré figure 3, se branche sur une prise de courant et signale l'état du réseau par l'intermédiaire d'une liaison cellulaire via un modem Particle Electron [14]. Nous avons choisi une liaison cellulaire avant d'aller sur le terrain en raison du pourcentage relativement élevé d'utilisateurs de téléphones portables dans la région. Les habitants ont en moyenne 1.37 abonnement mobile et 90.0 % possèdent un téléphone portable [15, 21].

Toutes les deux minutes, le capteur prélève un échantillon haute fréquence de la forme d'onde de la tension à la prise afin de calculer la tension efficace et la fréquence du réseau. Il enregistre également d'autres signaux, notamment la position GPS et le temps, la qualité du réseau cellulaire, et le nombre de signaux WiFi à proximité (à des fins d'évaluation en tant que potentiel canal latéral). Les versions plus récentes du capteur s'interrompent en cas de changement d'état d'alimentation et enregistrent l'heure (à partir d'une horloge temps réel, HTR) et l'accélération (pour aider à filtrer les événements liés au débranchement par l'utilisateur). Ces mesures sont stockées localement sur une carte SD et transmises au cloud lorsqu'une connexion cellulaire est disponible. Le capteur contient une batterie de 2000 mAh, qui peut fonctionner pendant plusieurs jours, plus longtemps que la plupart des pannes à Accra.

L'intervalle d'échantillonnage de deux minutes a été choisi comme compromis entre la résolution des données et les coûts de communication. Nous notons qu'il est plus élevé que le taux d'échantillonnage de 15 minutes utilisé par

la plupart des compteurs intelligents, qui sont considérés comme étant le dernier cri dans le calcul du SAIDI et du SAIFI. De plus, pour les capteurs contenant une fonctionnalité d’horodatage des pannes, le capteur peut signaler des pannes avec une précision de deuxième niveau, et d’une durée inférieure à une minute, ce qui est suffisant pour détecter une interruption prolongée telle que définie par IEEE [11].

Le capteur PowerWatch a été optimisé à l’origine pour sa fiabilité et sa facilité de fabrication en petites quantités, plutôt que pour son coût, et il a été mis à la disposition des organismes de financement pour 187 dollars américains par unité. Les éléments qui ont le plus contribué à ce coût sont le circuit imprimé peuplé avec l’alimentation électrique, le GPS et les circuits de détection (87\$), le Particle Electron (38\$), le boîtier (20\$) et l’assemblage (10\$). La communication non optimisée et l’infrastructure cloud coûtent actuellement 8 dollars par capteur et par mois. Les nouveaux capteurs, qui conservent des fonctionnalités identiques ou supérieures, devraient coûter entre 30 et 40 dollars, et les coûts optimisés de communication et de cloud devraient être inférieurs à 1 dollar par capteur et par mois.

4.4.2 Déploiement

Notre méthodologie de déploiement est décrite de manière détaillée dans d’autres articles [13] et n’est présentée ici que brièvement pour éclairer notre discussion sur les performances du système.

Accra est divisée en 26 districts; nous nous sommes déployés dans trois d’entre eux, illustrés figure 2. Nous avons choisi nos sites de déploiement en fonction de l’emplacement prévu des transformateurs, fournis par le fournisseur d’électricité [3], afin que nos données puissent être utilisées par nos collaborateurs dans le cadre d’une évaluation formelle de l’impact de certains projets.

Les capteurs PowerWatch sont déployés par une équipe de 15 employés locaux temporaires (appelés agents de terrain) et entretenus par une équipe de quatre employés locaux à temps plein. Les agents de terrain filtrent les occupants des sites potentiels des capteurs (appelés participants) pour s’assurer que leur maison ou leur entreprise soit connectée au réseau, et qu’elle l’utilise. Après avoir été informés de la collecte, du stockage et de l’utilisation des données, les participants consentent à ce que nous collections leurs données personnelles et les données PowerWatch. Toutes les interactions avec les participants sont approuvées par notre protocole IRB. Une fois que le participant a donné son consentement, l’agent de terrain utilise SurveyCTO [16] pour collecter des informations sur le participant, le lieu de déploiement, et le capteur déployé.

Les agents de terrain branchent le capteur au domicile ou à l’entreprise du participant ou, plus récemment en raison du COVID-19, indiquent aux participants de brancher le capteur sur une prise de courant disponible. Les participants reçoivent une multiprise pour s’assurer qu’ils n’ont pas besoin de sacrifier une prise de courant, et ils se voient automatiquement transférer 5 GHC (environ 1 dollar américain) de temps de communication par mois pour les inciter à garder le capteur installé et pour compenser les coûts d’électricité encourus par leur participation.

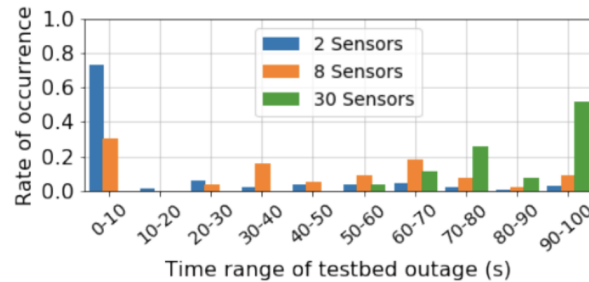
Les coûts de déploiement se sont élevés à 100 000 dollars pour un peu plus d’un an de fonctionnement, y compris le travail sur le terrain pour le déploiement et la maintenance, un chef de projet à temps plein, et les compensations des participants. Pour réduire les coûts liés à la compensation des participants et à la maintenance, nous prévoyons de piloter une application qui montre aux participants des informations sur la fiabilité (par exemple, en les alertant lorsqu’il y a une panne à leur domicile) pour les inciter à (continuer à) participer et à garder leur capteur branché sur le réseau.

4.4.3 Cloud

Le cloud du PowerWatch reçoit les données des capteurs PowerWatch et les stocke dans une base de données PostgreSQL/TimescaleDB [20]. Les données sont ensuite associées aux métadonnées de déploiement pour une analyse plus approfondie. Les données ne sont pas supprimées du capteur tant que celui-ci n’a pas reçu la confirmation que les données ont été stockées dans la base de données.

D'autres services cloud intégrant PowerWatch comprennent des tableaux de bord pour surveiller le déploiement et informer les agents de terrain lorsque des capteurs ne fonctionnent pas, des systèmes pour transférer des compensations aux participants, et des visualisations des données sur les pannes.

FIGURE 4 – Durée des pannes du banc d'essai. Un banc d'essai composé de capteurs et de prises programmables a généré deux cents pannes de différentes tailles dans un environnement contrôlé. Nous avons observé la précision de l'horodatage des pannes, en notant que pour une panne donnée, les capteurs peuvent signaler que la même panne s'est produite avec jusqu'à 100 secondes d'intervalle. Cela nous permet de paramétrer les algorithmes de partitionnement utilisés pour détecter les pannes sur le terrain. Un micrologiciel plus récent réduit la variance temporelle à moins de 10 secondes.



4.4.4 Partitionnement des pannes

Comme indiqué à la section 4.3, les rapports de panne provenant de plusieurs capteurs sont combinés pour garantir la validité d'une panne. Nous considérons qu'il suffit que deux capteurs fassent un rapport commun pour qu'une panne soit signalée. Pour effectuer ce partitionnement, nous utilisons STDBSCAN [4], qui partitionne les rapports de panne en grappes basées sur la densité à la fois dans le temps et dans l'espace.

STDBSCAN nécessite des paramètres pour spécifier le nombre minimum de points dans une grappe basée sur la densité, et la distance maximum entre les points à la fois dans le temps et dans l'espace.

Pour sélectionner le paramètre temporel de STDBSCAN, nous avons créé un banc d'essai afin de générer des pannes artificielles de différentes tailles et d'observer la distribution temporelle des rapports de pannes dans ce cadre contrôlé. Le banc d'essai se compose de trois prises programmables, avec 2, 8 et 30 capteurs connectés à chaque prise respectivement. Comme les capteurs sont connectés à la même prise, nous pouvons nous assurer qu'ils subissent une panne artificielle au même moment. Les capteurs du banc d'essai sont programmés avec la version du micrologiciel qui contient l'horodatage le moins précis.

Les temps résultant de cette expérience sont présentés dans la figure 4. Lors des 200 pannes artificielles, tous les capteurs configurés pour subir une panne ont signalé avec succès qu'une panne s'était produite. Pour une panne donnée, l'intervalle de temps des rapports de panne a varié jusqu'à 100 s (c'est-à-dire qu'un capteur a signalé que la panne s'est produite au maximum 100 s après qu'elle se soit réellement produite). Nous utilisons donc de manière prudente 100 s comme paramètre temporel pour STDBSCAN, ce qui permet à l'algorithme de partitionner deux capteurs avec un écart de temps de signalement allant jusqu'à 100 s. Nous constatons que les données provenant de versions plus récentes du capteur signalent toutes les pannes à moins de 4 s l'une de l'autre, ce qui nous permettra de réduire davantage le paramètre temporel de partitionnement.

Pour dériver le paramètre spatial de STDBSCAN, nous avons exploré la distribution spatiale des capteurs dans notre déploiement. Nous avons observé que des sites adjacents sont susceptibles (mais pas nécessairement) de subir

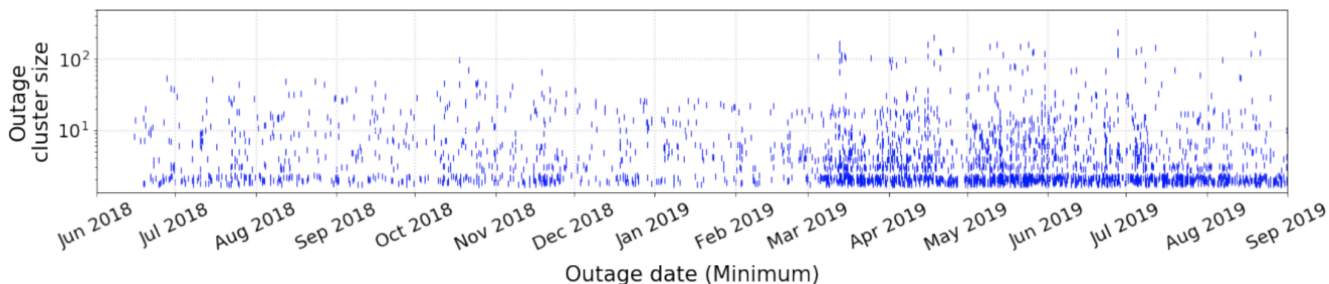
une panne au même moment [1, 22]. Par conséquent, pour tous les sites, nous calculons la distance maximale entre un site et le deuxième site le plus proche. Ce faisant, et en excluant les valeurs aberrantes dont le deuxième site le plus proche se situe au-delà de 3x de l'écart interquartile de la distribution, nous obtenons un paramètre de partitionnement spatial de 2,4 km.

6. Performance des applications

Entre juin 2018 et septembre 2019, PowerWatch a détecté 3,123 pannes, couvrant des pannes importantes provenant de défauts de haute et moyenne tension en amont du consommateur, mais aussi des pannes plus petites provenant de défaillances du réseau basse tension à proximité du consommateur. L'ensemble des pannes est présenté dans la figure 5.

Nous examinons plus en détail les performances de PowerWatch en considérant à la fois les méthodes utilisées pour extraire les pannes à partir d'un ensemble de données de rapports de panne contenant du bruit, ainsi que la pertinence de notre échantillon pour estimer S-SAIFI et S-SAIDI.

FIGURE 5 – Toutes les pannes détectées par PowerWatch de juin 2018 à septembre 2019. Les pannes sont visualisées sur une ligne de temps où l'axe vertical montre la taille de la panne (mesurée en nombre de capteurs impactés) sur une échelle logarithmique. De petites perturbations sont ajoutées à l'emplacement des lignes pour faciliter la distinction des pannes de même taille. PowerWatch a détecté 3,123 pannes, d'une durée moyenne de 1.7 heure. La panne la plus longue a duré plus de 48 heures. La panne la plus importante a touché une zone de près de 80 km², représentant deux tiers de nos capteurs déployés.



6.1 Extraction des pannes

PowerWatch extrait les pannes à partir de rapports de capteurs individuels qui sont agrégés en identifiant des grappes spatio-temporelles. Évaluer cette technique serait une tâche relativement simple si l'on disposait de mesures de référence ; cependant, le fournisseur d'électricité n'a que très peu de données de référence aux niveaux haute et moyenne tension, et pratiquement aucune donnée de référence utilisable aux niveaux basse tension.

Les données limitées qui existent - les données du système SCADA (système de contrôle et d'acquisition de données) pour les pannes à haute et moyenne tension, et les appels des clients et les registres des camions pour les pannes à moyenne et basse tension - n'ont pas été mises à la disposition de notre équipe, ou ne contiennent pas d'informations spatiales et temporelles suffisamment précises pour valider les événements de pannes individuelles.

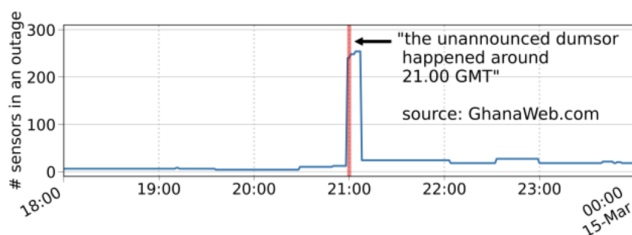
Par conséquent, notre évaluation complète les données de base limitées dont nous disposons par des évaluations des relations spatio-temporelles entre les rapports des capteurs (qui ne correspondent qu'à de véritables pannes).

6.1.1 Vérifications initiales

Nous cherchons tout d'abord à confirmer que des pannes similaires à celles détectées se produisent à Accra. Les pannes conséquentes sont parfois rapportées dans les journaux et peuvent être détectées par PowerWatch (voir figure 6)[9].

En outre, bien que les registres de réparation rapportés par les fournisseurs d'électricité ne soient pas suffisamment précis pour valider les pannes individuelles, nous pouvons comparer leur nombre relatif aux pannes détectées par PowerWatch. Les registres de réparation que nous avons obtenus indiquent 1,449 réparations en un peu plus de cinq mois dans un district, alors que PowerWatch a détecté 575 pannes au cours de la même période et dans le même district. Il est normal que PowerWatch détecte moins de pannes ; PowerWatch ne couvre qu'une partie du réseau. En extrapolant linéairement notre couverture à l'ensemble du district, nous anticipons que PowerWatch aurait détecté 1,801 pannes au cours de cette période, soit un nombre similaire à celui publié dans les journaux.

FIGURE 6 – Une panne d'électricité («dumsor») reportée par GhanaWeb (une source d'information populaire) comme s'étant produite «vers 21 heures»le 14 mars ; cette panne est parfaitement enregistrée par les capteurs PowerWatch et les algorithmes de partitionnement [9].



6.1.2 Modèles temporels dans les rapports des capteurs

Lors d'une panne de courant, l'ensemble du réseau en aval est rapidement mis hors tension. Cela déclenche des rapports individuels de panne de capteurs PowerWatch proches les uns des autres dans le temps. Inversement, nous nous attendons à ce que les rapports de panne faussement positifs causés par des participants qui débranchent un capteur ou par des compteurs prépayés qui expirent soient distribués de manière relativement aléatoire dans le temps. La transition entre ces deux groupes temporels devrait se produire autour de la grappe de temps maximale des capteurs signalant une véritable panne, soit environ 100 s, comme l'illustre la figure 4.

Pour tester cette hypothèse, nous examinons la fonction de distribution cumulative du temps écoulé entre un rapport de panne et le rapport de panne le plus proche, illustré figure 7. Conformément à nos attentes, nous observons une distribution bimodale des temps de signalement des pannes entre les capteurs, avec une transition entre ces deux modes autour de 100 s. Cela confirme notre hypothèse selon laquelle les vraies pannes peuvent être identifiées dans le domaine temporel.

6.1.3 Modèles spatiaux dans les rapports des capteurs

Nous nous attendons à ce que la plupart des vraies pannes soient spatialement denses, car la distribution spatiale des pannes (en particulier des petites pannes) est contiguë. De plus, nous anticipons que très peu de capteurs soient alimentés dans le périmètre d'une panne détectée, bien que cela reste possible en raison des générateurs et des structures concaves du réseau.

FIGURE 7 – Répartition des délais entre les différents rapports de débranchement des capteurs. Plus de 40 % des débranchements de capteurs ont lieu dans les 100 secondes qui suivent un autre rapport de débranchement. En outre, la section plate au milieu du graphique indique que les rapports de débranchement de capteurs se produisent principalement selon deux modes : ceux qui sont fortement corrélés dans le temps avec d’autres événements de débranchement, et ceux qui se produisent de manière beaucoup plus aléatoire dans le temps. Nous pensons que la corrélation temporelle est due aux pannes et que la présence de cette corrélation peut être utilisée pour distinguer les véritables événements de débranchement de ceux qui ne sont pas dus à une défaillance du réseau.

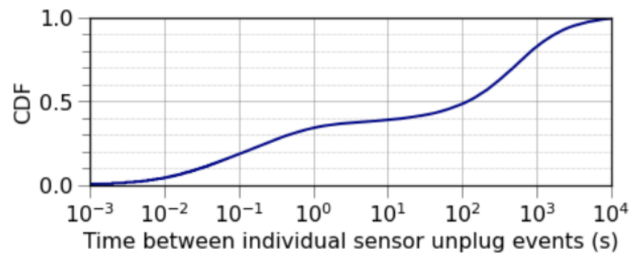


TABLE 1 – Nombre de capteurs alimentés dans l’enveloppe convexe d’une panne. Nous constatons que pour toutes les tailles de pannes, très peu de capteurs alimentés - au maximum 2 - se trouvent dans l’enveloppe convexe d’une panne détectée. Cela nous conforte dans l’idée que les pannes détectées par PowerWatch sont de vraies pannes, car nous ne nous attendons pas à ce que les capteurs situés dans une zone de panne soient alimentés (en dehors d’anomalies telles que la présence d’un générateur ou de formes concaves du réseau où des infrastructures alimentées séparément empiètent sur l’enveloppe convexe d’une panne).

Taille de la panne	Nombre de capteurs alimentés dans l’enveloppe convexe d’une panne			
	Moyenne	Moyenne %	Max	Max %
Pannes de 3-10 capteurs	0,03	0,33%	2	20%
Pannes de 10-30 capteurs	0,09	0,51%	2	11,76 %
Pannes de 30+ capteurs	0,31	0,60%	2	4,65 %

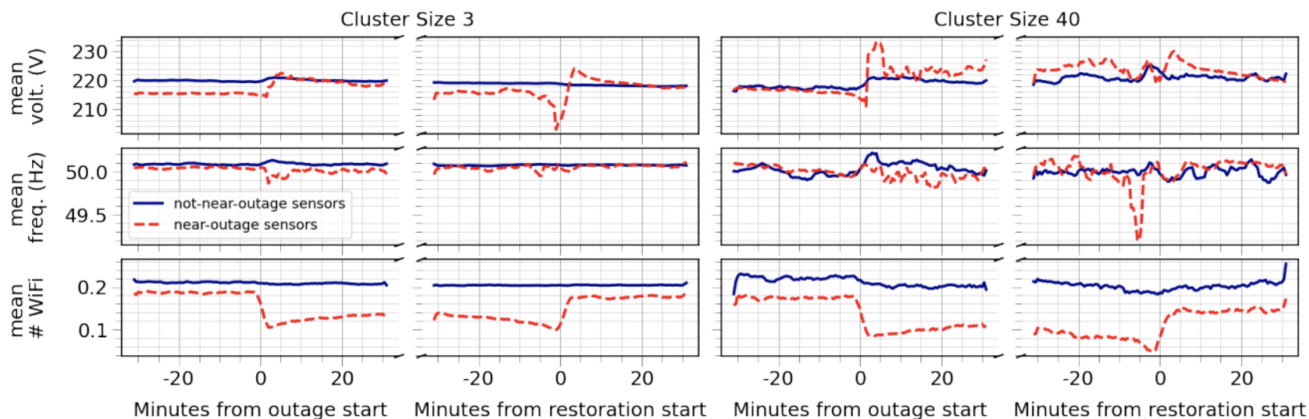
Pour vérifier si ces propriétés sont vérifiées dans l’ensemble des données PowerWatch, nous évaluons le nombre de capteurs alimentés à l’intérieur de l’enveloppe convexe des pannes détectées dans le tableau 1. Nous constatons que pour toutes les tailles de pannes, le nombre de capteurs alimentés à l’intérieur de l’enveloppe convexe est faible, avec pas plus de deux capteurs alimentés à l’intérieur de l’enveloppe convexe de n’importe quelle panne. L’absence de capteurs alimentés dans l’enveloppe convexe indique que PowerWatch détecte les vraies pannes.

6.1.4 Autres signaux corroborants

Pour confirmer que les pannes extraites par notre algorithme de partitionnement sont de vraies pannes, nous examinons d’autres signaux collectés par PowerWatch, à la recherche de signes indiquant qu’une panne s’est produite. Dans la figure 8, nous analysons la tension, la fréquence et le nombre de réseaux WiFi détectés par les capteurs proches d’une panne et par les capteurs qui ne sont pas proches de la panne et qui n’ont donc pas été touchés par celle-ci.

Pour les capteurs situés à proximité des petites pannes, nous observons une nette augmentation de la tension après la panne et une nette diminution de la tension au moment du rétablissement. Pour les pannes plus importantes,

FIGURE 8 – Tension, fréquence et nombre de réseaux WiFi avant et après une panne. Nous prenons un temps de référence, et nous prenons la moyenne de la tension, de la fréquence et du nombre de réseaux WiFi observés par les capteurs PowerWatch pendant les petites (grappes de 3 capteurs) et les grandes (grappes de 40 capteurs) coupures de courant et les rétablissements. Les capteurs sont «proches» d’une panne s’ils se trouvent sur le même site qu’un capteur affecté par la coupure. La tension et la fréquence ne sont pas mesurées pour les capteurs subissant une coupure. Au fur et à mesure que la taille des grappes augmente, nous observons que les capteurs qui ne se trouvent pas à proximité d’une panne détectent des changements de fréquence et de tension en réponse à la variation de la demande associée à une panne ou à un rétablissement. Le changement du nombre de signaux WiFi à proximité est similaire : il diminue en cas de panne et augmente en cas de rétablissement. L’ensemble de ces signaux corrobore le fait que les pannes détectées par PowerWatch sont de vraies pannes.



nous observons des effets similaires sur l’ensemble du réseau de capteurs, ainsi qu’une augmentation de la fréquence sur l’ensemble du réseau juste après la panne. Ces changements de tension et de fréquence sont conformes à ce que nous nous attendons à observer lorsqu’un changement soudain (tel qu’une panne) se produit dans le réseau. Pour les petites et grandes pannes, nous observons une baisse du nombre de réseaux WiFi au moment de la panne et une augmentation du nombre de réseaux WiFi au moment du rétablissement du courant, ce qui correspond à la perte d’alimentation des points d’accès WiFi situés à proximité.

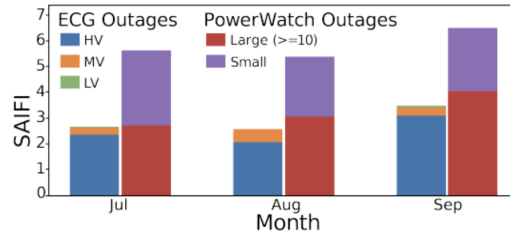
6.1.5 Résumé de l’extraction des pannes.

Bien que nous ne disposions pas de données de terrain pour connaître la précision exacte de nos algorithmes de détection des pannes, la présence de relations spatiales et temporelles entre les rapports collectés par les capteurs individuels sur le terrain, ainsi que la présence de changements de tension et de fréquence attendus à proximité d’une panne, s’explique le plus facilement par une défaillance du réseau. En utilisant la présence de ces relations pour renforcer l’hypothèse que les événements de panne détectés par PowerWatch sont de vraies pannes, nous évaluons la capacité de PowerWatch à échantillonner suffisamment le réseau pour estimer S-SAIFI et S-SAIDI.

6.2 Évaluation de l’échantillonnage

Une stratégie d’échantillonnage optimale consisterait à placer les capteurs de manière à ce qu’ils saisissent un ensemble représentatif du réseau. Malheureusement, cela n’est pas facile à réaliser à Accra, étant donné qu’il existe peu d’observations à haute résolution de la performance du réseau et que les cartes d’infrastructure disponibles

FIGURE 9 – Comparaison du S-SAIFI de PowerWatch avec le SAIFI rapporté par le fournisseur d’électricité (ECG) au cours du troisième trimestre 2018. Nous constatons que nos grandes grappes de pannes se comparent étroitement au SAIFI moyenne et haute tension rapporté par ECG, tandis que les pannes à basse tension (petites grappes de pannes) détectées par PowerWatch dépassent largement le SAIFI basse tension rapporté par ECG. Cela confirme l’ampleur du sous-échantillonnage par le fournisseur d’électricité au niveau de la basse tension du réseau.



sont incomplètes. Par conséquent, nous évaluons notre méthodologie de déploiement a posteriori, en tentant de répondre aux deux questions suivantes : (1) avons-nous déployé suffisamment de capteurs pour détecter et saisir correctement l’étendue de la plupart des pannes à haute et moyenne tension, et (2) avons-nous déployé un sous-échantillon suffisant pour avoir confiance en nos calculs S-SAIFI et S-SAIFI ?

Pour répondre à ces questions, nous comparons d’abord nos mesures de S-SAIFI à la meilleure mesure de SAIFI, recueillie par le système ECG SCADA, puis nous utilisons des méthodes numériques et statistiques pour évaluer le pouvoir prédictif de notre sous-échantillon pour l’estimation de S-SAIFI.

6.2.1 Comparaison avec la vérité de terrain

Nous comparons S-SAIFI au SAIFI rapporté par Electric Company of Ghana (ECG) au troisième trimestre 2018, la seule information SAIFI que nous avons pu collecter au moment de la publication. Le rapport de ECG comprend peu de pannes à basse tension car il n’y a pas de surveillance automatisée à basse tension. Certaines lignes de distribution moyenne tension ne sont pas non plus surveillées par le système SCADA d’ECG. Les rapports ECG sont agrégés par district, ce qui nous permet d’effectuer une comparaison directe avec le district que nous avons étudié au moment de cette analyse. Enfin, le calcul du SAIFI par ECG dépend de leur connaissance des raccordements au service clientèle dans chaque district, mais nous ne disposons pas de ces données.

Pour analyser le rapport ECG du troisième trimestre, nous comparons (Figure 9) le SAIFI mesuré par ECG à l’échelle du district avec le S-SAIFI mesuré par PowerWatch. La mesure est divisée en contributions provenant de petites grappes de moins de dix capteurs, ainsi que de grappes de plus de dix capteurs. Nous nous attendons à ce que la taille des grandes grappes corresponde aux pannes de haute et moyenne tension incluses dans le rapport ECG, et à ce que la taille des petites grappes corresponde aux pannes de basse tension.

Lorsque le S-SAIFI de PowerWatch est calculé pour des pannes plus importantes, nous constatons qu’il correspond étroitement au SAIFI rapporté par ECG. Nous observons également que PowerWatch détecte un nombre important de petites pannes qui ne sont pas détectées par ECG. Ces données suggèrent que ECG sous-échantillonne le réseau et ne signale pas les petites pannes qui affectent les clients.

6.2.2 Méthode combinatoire pour évaluer la couverture

Bien que les similitudes avec les chiffres SAIFI de ECG renforce la confiance dans notre méthode d’échantillonnage, nous explorons encore un peu plus la puissance de notre échantillon à l’aide de méthodes statistiques et

numériques. Nous commençons par vérifier que notre déploiement peut couvrir une partie importante du réseau haute et moyenne tension.

Nous ne disposons pas de cartes précises de l'infrastructure dans nos zones de déploiement, mais nous connaissons le nombre relatif d'éléments d'infrastructure à chaque niveau du réseau. Ces informations sont suffisantes pour construire un modèle simple qui estime la probabilité que notre déploiement observe une défaillance donnée à haute ou moyenne tension.

Pour créer ce modèle, nous supposons que : (1) dans chaque district, les transformateurs sont répartis uniformément entre les postes électriques, et (2) pour chaque site, chaque transformateur pas encore étudié a une chance égale d'être choisi. Nous présentons ensuite la question de la couverture comme un problème d'urne qui donne la probabilité que nous ayons choisi au moins un site de chaque poste électrique (haute tension) et/ou de chaque ligne d'alimentation (moyenne tension) après avoir sélectionné x sites, sans remplacement, dans un district donné.

La probabilité que notre site exclue un ou plusieurs postes électriques est de $1 - P(\bigcup \bar{S}_2 \bigcup \dots \bigcup \bar{S}_n)$, où S_x représente la proposition que l'échantillon inclut le x^{ieme} site, et \bar{S}_x l'exclut. L'union de n propositions pouvant être exprimée comme une somme de leurs intersections [10], nous pouvons écrire :

$$P\left(\bigcup_{i=1}^n A_i\right) = X_1 - X_2 + X_3 - \dots + (-1)^{n+1} X_n$$

où X_k , dans notre cas, est la somme, pour toutes les combinaisons de k postes, des probabilités que tous ces k postes aient été exclus de l'échantillon. Puisque nous supposons que chaque poste électrique est lié au même nombre de sites, et que les sites sont indépendants et identiquement distribués, cela se réduit à :

$$\binom{n}{1} P(\bar{S}_1) - \binom{n}{2} P(\bar{S}_1 \cap \bar{S}_2) + \dots + (-1)^{n+1} \binom{n}{n} P(\bar{S}_1 \cap \dots \cap \bar{S}_n),$$

où

$$P(\bar{S}_n) = \binom{\text{transformateur}}{x}^{-1} \binom{\frac{\text{transformateur}}{\text{poste électrique}} \times (\text{poste électrique} - n)}{x}$$

Nous constatons que parmi les nombreuses façons de choisir 85 sites à Achimota ($\binom{461}{85} \approx 2,4 \times 10^{94}$), Dansoman ($\binom{157}{38} \approx 4,0 \times 10^{36}$) et Kaneshie ($\binom{343}{28} \approx 1,0 \times 10^{41}$), dans les trois districts, moins de 0,01 % ont exclu des postes électriques, ce qui nous conforte dans l'idée que nous devrions pouvoir observer toutes les coupures de courant à haute tension. En remplaçant les postes électriques par des lignes d'alimentation dans l'équation ci-dessus, nous constatons qu'un tirage au sort de 85 sites à Achimota inclura toutes les lignes d'alimentation à moyenne tension avec une probabilité de 44 %.

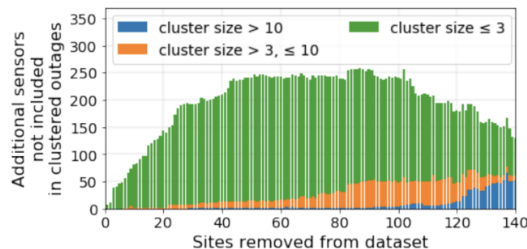
6.2.3 Étude d'exclusion pour l'évaluation de la couverture

Nous vérifions notre modèle de couverture à partir de nos données en notant que, si la couverture des capteurs est suffisante à un certain niveau du réseau, la suppression d'un petit nombre de capteurs de notre ensemble de données ne devrait pas avoir d'impact significatif sur le nombre ou la taille des pannes détectées à ce niveau. Nous testons cette hypothèse en réalisant une étude d'exclusion : nous retirons des sites de notre ensemble de données et observons l'impact de ces retraits sur le nombre et l'étendue des pannes détectées par PowerWatch.

Lorsque nous retirons un site dans cette étude, nous nous attendons à l'un des trois résultats suivants : une panne n'est plus détectée, la taille de la grappe diminue, ou une panne plus importante est divisée en deux ou plusieurs pannes plus petites. Dans chaque cas, sans une couverture suffisante, nous nous attendons à ce que la suppression d'un seul site entraîne une réduction de la taille de la panne au-delà du site supprimé pour l'étude, ce qui indique que le site supprimé a joué un rôle non redondant dans notre couverture. Nous supprimons de plus en plus de sites,

nous comptons les capteurs additionnels impactés, et nous présentons les résultats dans la figure 10. Nous constatons que pour toutes les pannes, sauf les plus petites, nous devons supprimer plus de 20 sites avant que les sites supprimés ne soient plus redondants, ce qui suggère que nous disposons d’une couverture suffisante pour partitionner les pannes de haute et moyenne tension dans nos zones de déploiement. Pour les petites pannes avec trois capteurs ou moins, les signes d’une couverture insuffisante apparaissent immédiatement - dès qu’un seul site est retiré de notre ensemble de données, les petites pannes qui se produisaient en dehors de ce site ne sont plus détectées. Cela n’a rien de surprenant puisque, comme le montre la figure 11, il arrive souvent que des pannes n’affectent que certaines parties d’un site. Nous savons que notre déploiement n’est pas assez dense pour détecter toutes les pannes de basse tension ; cependant, nous pouvons montrer qu’il est suffisant pour estimer S-SAIDI et S-SAIFI.

FIGURE 10 – Étude d’exclusion de la couverture de juin à août 2019. Pour évaluer la couverture de détection des pannes par PowerWatch, nous effectuons une étude d’exclusion, en retirant des sites de notre ensemble de données et en observant l’impact de ces suppressions. Plus précisément, nous examinons le nombre de «capteurs supplémentaires» qui faisaient partie d’un groupe de pannes avant l’exclusion, mais qui n’en faisaient plus partie après l’exclusion d’un site. Intuitivement, si la suppression d’un site entraîne l’absence de formation de nombreuses pannes ou une réduction significative de leur taille, cela indique que le site est nécessaire pour détecter l’étendue exacte d’une panne et que nous sommes peut-être en train de sous-échantillonner. Au cours de cette période, sans qu’aucun site ne soit supprimé, il y a 1,383 rapports de capteurs impliqués dans des pannes de taille ≤ 3 ; 1,030 rapports de capteurs impliqués dans des pannes de taille > 3 et ≤ 10 ; et 3,969 rapports de capteurs impliqués dans des pannes de taille > 10 . Nous observons que pour les pannes contenant plus de trois capteurs, près de 20 sites peuvent être retirés de notre ensemble de données avant qu’on ne commence à manquer de rapports provenant de capteurs supplémentaires. Cela indique que nous avons déployé suffisamment de capteurs pour détecter les pannes de moyenne et haute tension, mais, comme prévu, nous n’avons pas un degré élevé de couverture sur le réseau basse tension et nous devons nous appuyer sur l’échantillonnage pour estimer sa fiabilité.

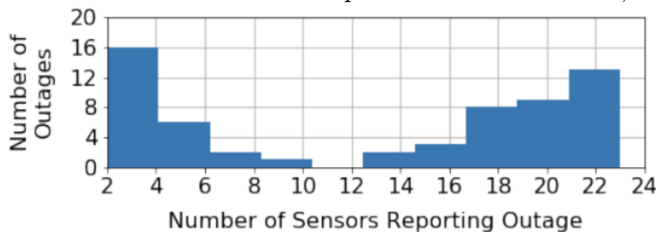


6.2.4 Étude d’exclusion pour l’évaluation de S-SAIDI.

Pour déterminer si notre échantillon est suffisamment grand pour saisir la plage des performances du réseau à Accra et estimer le S-SAIDI, nous menons une étude d’exclusion similaire à celle de la section 6.2.3. Nous effectuons trente cycles d’exclusion, en sélectionnant au hasard des ensembles de sites à retirer et en observant l’effet de chaque retrait de site sur le S-SAIDI.

Si la distribution des durées d’interruption avait changé de manière significative au fur et à mesure que nous retirions des sites, nous aurions vu S-SAIDI varier considérablement, ce qui suggérerait que nous étions peut-être en train de sur- ou sous-échantillonner des parties du réseau dont les expériences d’interruption seraient différentes. Cependant, comme le montre la figure 12, ce n’est pas le cas. Au fur et à mesure que nous supprimons des sites, nous observons une tendance à la baisse de la moyenne S-SAIDI, qui reflète que la distribution globale des durées d’interruption est

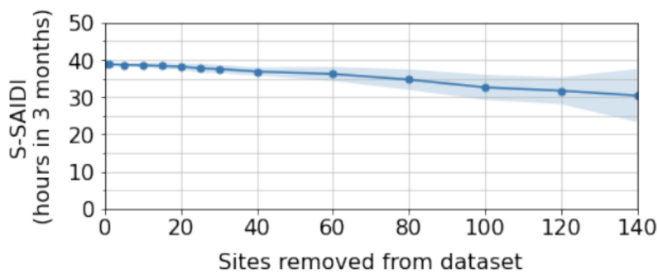
FIGURE 11 – Nombre de capteurs signalant des pannes dans un site densément instrumenté. Pour mieux comprendre les limites de notre échantillonnage basse tension, nous déployons 25 capteurs sur un seul site (sous un seul transformateur) pendant deux mois et observons les résultats. Nous observons deux groupes de pannes : les pannes plus importantes, qui touchent l’ensemble ou une partie importante du site, et les pannes plus petites, qui peuvent concerner une seule phase ou moins. Les pannes plus importantes représentent environ 60 % des pannes sur ce site, tandis que les pannes plus petites représentent environ 40 %. Cela suggère que notre stratégie de déploiement primaire de trois capteurs par site détecte de nombreuses pannes de basse tension, mais pas toutes.



asymétrique, avec une queue contenant quelques longues interruptions de basse tension ; néanmoins, même lorsque nous supprimons plus de 60 sites, nous ne voyons pas S-SAIDI s’écarter de manière significative de notre ensemble de données non modifiées (39 heures). Cela indique que le déploiement de PowerWatch a échantillonné de manière adéquate la plage de fiabilité présente sur nos sites de déploiement.

Au fur et à mesure que nous supprimons des sites, nous comparons également les distributions des durées d’interruption de notre ensemble de données complet et de notre sous-échantillon, en utilisant une distance statistique [17]. Nous avons utilisé l’implémentation du test de permutation pour une distribution égale [6, 18] disponible depuis la bibliothèque EUGENE [12]. Nous constatons que les distributions ne diffèrent pas de manière significative, ce qui signifie que la distribution combinée ne change pas en fonction des sites supprimés, ce qui confirme que notre échantillon est adéquat.

FIGURE 12 – S-SAIDI calculé \pm un écart-type au fur et à mesure que des sites sont retirés de l’ensemble de données entre juin et août 2019. Pour évaluer si PowerWatch couvre un échantillon suffisant du réseau pour calculer un S-SAIDI représentatif, nous retirons des sites de l’ensemble de données en 30 itérations et nous observons l’effet sur le S-SAIDI. Nous constatons qu’au fur et à mesure que des sites sont retirés, l’écart-type de S-SAIDI reste relativement faible, et la valeur moyenne de S-SAIDI diminue légèrement.



Références

- [1] Kobina AIDOO et Ryan C BRIGGS. “Underpowered : Rolling blackouts in Africa disproportionately hurt the poor”. In : *African Studies Review* 62.3 (2019), p. 112-131.
- [2] Jean ARLET. “Electricity tariffs, power outages and firm performance : a comparative analysis”. In : *Proceedings of the DECRG Kuala Lumpur Seminar Series, Kuala Lumpur, Malaysia*. T. 23. 2017.
- [3] Millennium Development AUTHORITY. “Power Distribution Feasibility Studies, Ghana Phase II Feasibility Assessment. (2014)”. In : (2014). URL : https://www.mida.gov.gh/pages/view/90/Phase_II_Report_%20ECG_Final_112414-1.pdf.
- [4] Derya BIRANT et Alp KUT. “ST-DBSCAN : An algorithm for clustering spatial-temporal data”. In : *Data & knowledge engineering* 60.1 (2007), p. 208-221.
- [5] Millenium Challenge CORPORATION. “Policy for Monitoring and Evaluation.” In : (2017). URL : <https://www.mcc.gov/resources/doc/policy-for-%20monitoring-and-evaluation>.
- [6] Bradley EFRON et Robert J TIBSHIRANI. “Permutation tests”. In : *An introduction to the bootstrap* (1993), p. 202-219.
- [7] Joseph H ETO et al. “Distribution-level electricity reliability : Temporal trends using statistical analysis”. In : *Energy policy* 49 (2012), p. 243-252.
- [8] Dimitry GERSHENSON, Brandon ROHRER et Anna. LERNER. “A new predictive model for more accurate electrical grid mapping.” In : (2019).
- [9] GHANAWEB. “Dumsor hits parts of Ghana. (2019).” In : (2019). URL : <https://www.ghanaweb.com/GhanaHomePage/NewsArchive>.
- [10] John HAIGH. *Probability models*. Springer Science & Business Media, 2013.
- [11] “IEEE Guide for Electric Power Distribution Reliability Indices”. In : *IEEE Std 1366, 2001 Edition* (2001), p. i-. DOI : 10.1109/IEEESTD.2001.94438.
- [12] Benjamin JANTZEN. *The EUGENE Project : A toolbox for detecting, comparing, and characterizing dynamical kinds*. 2021.
- [13] Noah KLUGMAN et al. “Hardware, apps, and surveys at scale : insights from measuring grid reliability in Accra, Ghana”. In : *Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies*. 2019, p. 134-144.
- [14] PARTICLE.IO. “Electron Datasheet.” In : (2019). URL : <https://docs.particle.io/%20datasheets/cellular/electron-datasheet/>.
- [15] Ghana. Statistical SERVICE. *Ghana Living Standards Survey Round 6 (GLSS 6). : Poverty profile in Ghana (2005-2013)*. Ghana Statistical Service, 2014.
- [16] “SurveyCTO”. In : (2020). URL : <https://www.surveycto.com/>.
- [17] Gábor J SZÉKELY et Maria L RIZZO. “Energy statistics : A class of statistics based on distances”. In : *Journal of statistical planning and inference* 143.8 (2013), p. 1249-1272.
- [18] Gábor J SZÉKELY, Maria L RIZZO et al. “Testing for equal distributions in high dimension”. In : *InterStat* 5.16.10 (2004), p. 1249-1272.
- [19] Prasad TALLAPRAGADA VSN et al. “Monitoring performance of electric utilities : indicators and benchmarking in sub-Saharan Africa”. In : (2009).

- [20] TIMESCALE. “TimescaleDB documentation.” In : (2019). URL : <https://docs.timescale.com/latest/main>.
- [21] International Telecommunication UNION. “Mobile Cellular 2000 - 2018.” In : (2019). URL : https://www.itu.int/en/ITU-D/Statistics/Documents/statistics/202019/Mobile_cellular_2000-2018_Jun2019.xls.
- [22] Rui YAO et al. “A multi-timescale quasi-dynamic model for simulation of cascading outages”. In : *IEEE Transactions on Power Systems* 31.4 (2015), p. 3189-3201.