



# **Pratiques de normalisation des données sur le SRAS-CoV-2 dans l'épidémiologie des eaux usées**

Rapport technique préparé pour l'ASPC

Thomas Maere, Jean-David Therrien et Peter Vanrolleghem

Université Laval



FACULTÉ DES SCIENCES ET DE GÉNIE

Département de génie civil et de génie des eaux

Cité universitaire  
Québec, Canada, G1V 0A6

## Rapport technique

# Pratiques de normalisation des données sur le SRAS-CoV-2 dans l'épidémiologie des eaux usées

AUTEURS/COAUTEURS	ÉQUIPE/GROUPE
Thomas Maere	modelEAU
Jean-David Therrien	modelEAU
Peter Vanrolleghem	modelEAU

RÉDIGÉ PAR		DERNIÈRE RÉVISION PAR	
<b>Nom</b>	Thomas Maere	<b>Nom</b>	Peter Vanrolleghem
<b>Date</b>	31/03/2022	<b>Date</b>	31/03/2022

<b>NOM DU FICHIER DU DOCUMENT</b>	Pratiques de normalisation des données sur le SRAS-CoV-2 en épidémiologie des eaux usées - Rapport technique
<b>NOMBRE DE PAGES</b>	18
<b>NUMÉRO DE RÉVISION</b>	1

## INDEX

1.	ÉNONCÉ DU PROBLÈME ET OBJECTIFS .....	1
2.	NATURE DU SIGNAL VIRAL DANS LES EAUX USÉES .....	1
3.	LE CONCEPT DE NORMALISATION DANS L'ÉPIDÉMIOLOGIE DES EAUX USÉES .....	1
4.	DESCRIPTION DES FACTEURS D'INFLUENCE LE LONG DU PARCOURS D'EEU.....	3
4.1.	DYNAMIQUE D'EXCRÉTION .....	3
4.2.	ORIGINE, TRANSPORT ET PROCÉDÉS PROPRES AUX EAUX USÉES .....	3
4.3.	ÉCHANTILLONNAGE .....	4
4.4.	ANALYSE.....	5
5.	DIFFÉRENCES ENTRE LES DONNÉES DES EAUX USÉES ET LES DONNÉES CLINIQUES SUR LA COVID-19.....	5
6.	LA PRATIQUE DE LA NORMALISATION DANS LE CONTEXTE CANADIEN .....	7
7.	UTILISATION DE BIOMARQUEURS ET DE PARAMÈTRES AUXILIAIRES DANS LE CONTEXTE INTERNATIONAL.....	10
8.	CONSIDÉRATIONS SUPPLÉMENTAIRES RELATIVES À LA PROCÉDURE DE NORMALISATION.	12
9.	CONCLUSIONS.....	14
10.	RÉFÉRENCES.....	14

## LISTE DES TABLEAUX

Tableau 1 : Aperçu de la façon dont les données d'EEU sont présentées sur les tableaux de bord partout au Canada .....	8
--	---

## **1. ÉNONCÉ DU PROBLÈME ET OBJECTIFS**

La quantité de matériel génétique du SRAS-CoV-2 trouvée dans les échantillons d'eaux usées, généralement exprimée en copies de gènes par ml, pourrait ne pas toujours permettre une évaluation fiable de la situation épidémiologique dans les égouts. Il est essentiel de comprendre les différents types de facteurs qui pourraient influencer sur les mesures du SRAS dans les eaux usées et les façons possibles de composer avec ces facteurs. Le présent rapport vise à donner un aperçu structuré de ces facteurs d'influence et à examiner les diverses stratégies et techniques d'adaptation qui sont généralement désignées dans la littérature comme « normalisation ».

## **2. NATURE DU SIGNAL VIRAL DANS LES EAUX USÉES**

Les données cliniques de santé publique sur les cas de COVID-19 dans les grandes villes au cours d'une vague épidémiologique ont un comportement caractéristique conforme à la théorie épidémiologique et aux simulations de modèles d'infection. Les données de cas d'une vague suivent généralement une courbe en forme de cloche où la valeur d'un jour particulier est déterminée principalement par la valeur du jour précédent et la tendance actuelle des infections. En moyenne, le signal viral des eaux usées correspond raisonnablement bien aux tendances d'infection à long terme. Néanmoins, il présente un comportement particulièrement bruité à court terme qui n'est généralement pas observé dans les données cliniques. Il reste à déterminer si ces pics de bruit sont réels (p. ex., en raison de la variabilité de l'excrétion fécale de matière virale par personne, de la variabilité du moment et de l'emplacement des cas dans les égouts, de la mobilité de la population, des événements de propagation excessive) ou de variations indésirables (p. ex., en lien avec des processus qui se produisent dans l'égout, des erreurs d'échantillonnage ou des problèmes de mesure). Dans le cas des grandes villes et d'une vague épidémiologique bien établie, on peut raisonnablement s'attendre à ce que les causes réelles de la variabilité soient pour la plupart égales selon la loi des grands nombres pour les événements stochastiques, laissant surtout la variabilité indésirable liée aux divers processus d'égout et aux erreurs d'échantillonnage et de mesure. Pour les petits égouts ou les faibles nombres de cas cliniques, la distinction entre la variation réelle et la variation indésirable est toutefois beaucoup plus difficile à faire. D'autant plus que dans ces conditions, avec des quantités infimes de matériel génétique dans les eaux usées, les techniques biomoléculaires employées sont poussées à leur limite et produisent des résultats moins fiables.

## **3. LE CONCEPT DE NORMALISATION DANS L'ÉPIDÉMIOLOGIE DES EAUX USÉES**

En science des données, la normalisation consiste à transformer ou à mettre à l'échelle toutes les valeurs des ensembles de données dans la plage [0, 1]. La normalisation des écarts réduits rééchelonne les données pour qu'elles aient les propriétés d'une distribution normale ( $\mu=0$  et  $\sigma=1$ ). En épidémiologie standard, le terme normalisation n'existe pas. Par contre, le terme standardisation est utilisé pour désigner un ensemble de techniques fondées sur la moyenne pondérée, utilisées pour éliminer autant que possible les effets des différences d'âge ou d'autres variables confusionnelles en comparant deux populations ou plus. Dans la littérature actuelle sur l'épidémiologie des eaux usées (EEU), l'utilisation du terme normalisation est ambiguë et couvre divers sujets. Elle comprend généralement l'utilisation d'un certain type de paramètre auxiliaire (p.

ex., taille des égouts, débit, ...) ou de biomarqueur (p. ex., charge des eaux usées, produits pharmaceutiques, virus crAssphage, virus PMMoV, ...) qui est mesuré parallèlement à la variable d'intérêt (p. ex., SRAS-CoV-2) et utilisé pour corriger ou évaluer la valeur de cette variable.

La mesure du SRAS-CoV-2 comporte de nombreuses étapes (échantillonnage, transport, extraction, amplification), et des erreurs peuvent se produire à chaque étape de la chaîne. L'utilisation de différentes techniques pour chaque étape de cette chaîne de mesure a également un impact significatif sur les résultats obtenus. La mesure d'un marqueur censée être constant dans les rejets d'eaux usées pour la population desservie et qui est soumis à toutes les mêmes étapes de mesure que le SRAS-CoV-2, peut être utilisée comme norme de contrôle interne et donner une première indication de la validité des valeurs du SRAS-CoV-2. Selon le résultat réel et la plage de valeurs attendues pour ce marqueur, on peut décider d'accepter ou de rejeter la valeur mesurée du SRAS-CoV-2. L'application d'un biomarqueur pour vérifier la validité des mesures du SRAS-CoV-2 en laboratoire n'est généralement pas considérée comme un type de normalisation, mais plutôt comme une partie de la procédure d'assurance et de contrôle de la qualité (AQ/CQ; Chik et coll., 2022) nécessaire pour produire des données fiables. Cependant, lorsqu'on applique une procédure de correction au signal du SRAS-CoV-2 en fonction des valeurs attendues d'un biomarqueur, la littérature sur l'épidémiologie des eaux usées parle souvent de normalisation. Dans ces cas-là, le terme normalisation est défini comme une technique pour corriger une variabilité indésirable dans le signal viral des eaux usées en comparant le matériel viral détecté à une autre substance excrétée par les humains dont la charge quotidiennement rejetée dans les égouts est connue ou réputée stable dans le temps.

Un biomarqueur est choisi parce qu'il peut décrire un facteur ou une combinaison de plusieurs facteurs qui influencent le signal d'intérêt. Le comportement du biomarqueur doit être bien connu tout au long de son parcours dans l'environnement, du moment de l'excrétion jusqu'à sa mesure en laboratoire. En outre, son comportement devrait ressembler au comportement de la variable d'intérêt en ce qui a trait au facteur d'influence que l'on souhaite corriger. Cela aide si la charge du biomarqueur dans le bassin versant des eaux usées est connue avec précision, mais ce n'est pas une condition préalable si la charge semble constante (c.-à-d., une charge constante est attendue et les données le confirment). Le biomarqueur peut être de la même nature que la variable d'intérêt (p. ex., matériel génétique d'un autre virus dans le cas du SRAS-CoV-2), mais ce n'est pas nécessaire si le facteur qui influence le signal viral affecte d'autres types de composés de façon égale. (p. ex., concentration des eaux usées pour caractériser la dilution).

La littérature montre qu'il est difficile de comparer les données sur les eaux usées du SRAS-CoV-2 provenant de différentes études. Non seulement y a-t-il des différences dans la situation épidémiologique et sa gestion dans différentes villes et pays, mais chaque égout possède aussi ses particularités, et les techniques utilisées par chaque étude sont différentes. Dans ce contexte, plusieurs approches sont envisageables pour mieux comparer les données entre les études. La comparaison de diverses techniques d'échantillonnage et de mesure a fait l'objet de plusieurs études et donne une idée de l'équivalence entre certaines méthodes. En revanche, les différences entre les laboratoires utilisant les mêmes méthodes pourraient être évaluées au moyen d'essais interlaboratoires d'aliqots du même échantillon. Actuellement, il n'existe aucune technique de mesure standard ni procédure d'AQ/CQ standard pour la quantification d'un signal viral dans les eaux usées. L'influence des caractéristiques des égouts locaux (p. ex., type, taille, âge) est encore assez méconnue, mais on suppose que les valeurs mesurées devraient être exprimées en équivalents par personne pour rendre toute comparaison possible. La littérature propose le terme « normalisation de la population » lorsque des estimations dynamiques de la population sont

utilisées pour tenir compte des variations de la population dans les bassins versants d'eaux usées - ce qui peut également être fait à l'aide de biomarqueurs. La combinaison des résultats de plusieurs régions en une seule mesure est appelée agrégation.

## **4. DESCRIPTION DES FACTEURS D'INFLUENCE LE LONG DU PARCOURS D'EEU**

### **4.1. Dynamique d'excrétion**

Une quantité considérable de personnes vont excréter du matériel viral par leurs excréments une fois infectés par le virus SRAS-CoV-2. Cependant, le pourcentage exact de personnes qui le font est inconnu, tout comme on ignore qui excrétera le matériel, en quelle quantité, et pendant combien de temps. On peut supposer que cela dépend de la génétique individuelle, de l'âge, de l'état vaccinal, du variant de la COVID-19 en question et de la gravité de l'infection (Hoffmann et Alsing, 2021). Une source supplémentaire d'incertitude est liée au moment et au lieu de chaque excrétion (c.-à-d. une chasse d'eau de toilette de 5 à 10 secondes), qui peuvent être considérés comme des charges ponctuelles discrètes de virus injectées dans le réseau d'égout à quelques moments par semaine pour un patient donné.

Il n'existe pas de technique de normalisation connue pour corriger l'incertitude liée à la dynamique d'excrétion. Par conséquent, cette source d'incertitude nuira à tous les autres types d'analyse des données sur les eaux usées, comme la recherche de procédures de correction pour d'autres facteurs influençant la variabilité des mesures. Grâce à la modélisation de l'excrétion et des égouts, il est possible de comprendre l'étendue de l'incertitude que les dynamiques d'excrétion entraînent et dans quelles conditions ses effets sont critiques (Wade et coll., 2022). Comme mentionné précédemment, avec la loi des grands nombres, on s'attendrait à moins d'influence de la dynamique d'excrétion pour les grandes villes où il y a beaucoup de cas et une dispersion et un mélange considérables dans les tuyaux d'égout.

### **4.2. Origine, transport et procédés propres aux eaux usées**

La quantité et la composition des eaux usées varient dans le temps selon des dynamiques diurnes, quotidiennes, saisonnières et annuelles. Les eaux usées sont constituées d'une fraction liquide et d'une fraction particulaire, chacune comportant une myriade de composants biologiques et chimiques. La dynamique de transport et de dispersion de ces deux fractions dans l'égout sont différentes, car les particules peuvent se déposer et se remettre en suspension selon le débit. Les égouts et les bassins versants d'eaux usées diffèrent d'une ville à l'autre. Des distinctions peuvent être faites en fonction de l'origine des eaux usées (c.-à-d. domestiques, industrielles), du type d'égout (canalisations d'égout séparées, combinées ou hybrides et aménagement des canalisations d'égout), de la taille (affectant le temps de rétention hydraulique et particulaire), de l'emplacement (température), de la topographie (flux gravitationnels ou pompés), de l'âge (infiltration des eaux souterraines par des fuites), des règles de construction (débits, surverses dans les cours d'eau, réservoirs de rétention). Tous ces facteurs mènent à des niveaux d'impact variables des différents processus physiques, chimiques et biologiques dans les égouts sur le signal du SRAS-CoV-2 mesuré. Bien que les procédures de normalisation puissent atténuer l'influence de certains de ces processus (p. ex., dilution par la pluie, les eaux souterraines ou la fonte des neiges; dispersion et mélange), ce n'est peut-être pas nécessairement le cas pour d'autres facteurs (p. ex., dégradation du matériel

génétique) pour lesquels l'importance devrait être quantifiée au moyen d'études ciblées et d'une modélisation du réseau d'égouts.

### **4.3. Échantillonnage**

La représentativité des résultats obtenus dépend largement de la méthode d'échantillonnage choisie. Les études d'EEU pour la COVID-19 montrent que la gamme complète des types d'échantillonnage est utilisée (échantillonnage instantané, composite en fonction du temps et du débit, échantillonnage passif). D'après l'expérience du suivi des contaminants émergents préoccupants (CEP) dans les eaux usées (Ort et coll., 2010), il est reconnu qu'une stratégie d'échantillonnage composite proportionnelle au débit bien exécutée (c.-à-d. une prise d'eau bien placée, des sous-échantillons fréquents, un grand volume d'échantillons, réfrigération) donnera une vision plus réaliste de la charge virale totale qui traverse le réseau d'égouts. Bien que des échantillonnages instantanés fréquents ou des échantillonnages composites proportionnels au temps pourraient fournir une bonne indication de la situation pour le SRAS-CoV-2, ces stratégies comportent des biais structurels qui ne peuvent pas être résolus par la normalisation d'un biomarqueur, puisque ce biomarqueur serait sujet aux mêmes biais lors de l'échantillonnage. Plus le patron diurne du composé d'intérêt est erratique, dynamique ou prononcé, plus le biais peut être important. Par conséquent, les sites d'échantillonnage en amont de leur bassin hydrographique seront plus vulnérables aux erreurs d'échantillonnage étant donné le temps de mélange réduit et la moindre dispersion des charges ponctuelles de virus dans le débit total des eaux usées. La modélisation des égouts peut estimer l'étendue du biais et de la variabilité attendus dans les petits et grands bassins hydrographiques tout au long d'une vague épidémique.

Puisque la concentration fluctue tout au long de la journée, des échantillons instantanés peu fréquents pourraient indiquer des valeurs très élevées ou très faibles selon le moment d'échantillonnage. En amont des bassins hydrographiques avec seulement quelques cas, un signal viral pourrait passer complètement inaperçu. Dans ces situations, s'il est impossible d'installer des échantillonneurs composites, une stratégie d'échantillonnage passive pourrait être utile pour indiquer la présence ou l'absence de signal. Plusieurs études indiquent la nature semi-quantitative de l'échantillonnage passif pour le SRAS-CoV-2 avec des valeurs mesurées plus élevées liées à des concentrations d'eaux usées plus élevées et à des durées d'exposition plus longues. Cependant, la déduction de la concentration ou de la charge à laquelle l'échantillonneur a été exposé (rétrocalcul) s'avère difficile, car la cinétique de sorption et de désorption n'est pas (encore) bien définie. Les expositions longues (s'étendant sur plusieurs jours) augmentent également le risque de saturation du matériau de sorption passif (Schang et coll., 2021). La normalisation avec un biomarqueur pourrait aider à améliorer la nature quantitative de l'échantillonnage passif, mais seulement si le biomarqueur se comporte de façon similaire à la variable d'intérêt.

Un dernier aspect important de la stratégie d'échantillonnage est la nature de l'échantillon à prélever. Bien que la plupart des études ciblent l'ensemble de la matrice des eaux usées, certaines se concentrent sur les solides et cherchent ces solides dans les conduites de soutirage des boues de décanteurs primaires. Cette approche semble améliorer la limite de détection, car le SRAS-CoV-2 a tendance à être principalement associé aux solides. Cependant, elle complique le rétrocalcul du signal à la charge virale dans le bassin hydrographique. Les décanteurs primaires parviennent généralement à recueillir environ la moitié des solides dans les eaux brutes en tête de station (à moins d'employer en adjuvant chimique), mais leur performance de séparation des particules et la concentration des boues qui en découle sont variables et dépendent des dynamiques du débit. La normalisation du signal d'intérêt au moyen de biomarqueurs associés aux solides et la modélisation

des décanteurs primaires devraient aider à tirer avantage des données issues de l'échantillonnage des boues primaires.

#### **4.4. Analyse**

Comme il a été mentionné précédemment, il n'existe actuellement ni technique de mesure standard du SRAS-CoV-2 ni procédure standard d'AQ/CQ pour la mesure du virus dans les eaux usées. Des différences dans les résultats ont été signalées pour différentes méthodes d'entreposage, d'extraction et d'analyse, ce qui complique la comparaison des résultats. L'une des sources de variation les plus importantes est la fraction des eaux usées qui est ciblée. Les échantillons sont centrifugés et filtrés à divers degrés dans diverses études, ce qui donne lieu à différentes combinaisons des fractions originales présentes dans les eaux usées (solutés, petites particules, particules plus grosses). Ces fractions variables peuvent modifier la relation entre les résultats obtenus et la teneur virale réelle dans les échantillons originaux et avoir une incidence sur l'efficacité et la répétabilité des étapes d'extraction et d'amplification de l'ARN.

De plus, il faut garder à l'esprit le degré variable auquel les facteurs le long de la voie de l'EEU (excrétion, égout, laboratoire) peuvent influencer sur chaque fraction d'eau usée (p. ex., sédimentation des particules et remise en suspension dans les égouts). L'équivalence et la normalisation des procédures devraient faire l'objet d'études spécifiques et peuvent être appliquées par des essais interlaboratoires et la certification. L'utilisation d'un biomarqueur comme substance de contrôle interne pourrait aider à accroître la comparabilité si le partitionnement de cette dernière dans la matrice des eaux usées est connu et stable, et si elle réagit de manière similaire à la cible d'intérêt tout au long de la procédure de mesure. La confiance dans les données de chaque laboratoire peut être accrue par une application méticuleuse de la méthodologie sélectionnée, par la production de nombreux réplicas techniques et biologiques et par la production de mesures à haute fréquence (quotidienne) qui facilitent l'analyse des séries temporelles.

## **5. DIFFÉRENCES ENTRE LES DONNÉES DES EAUX USÉES ET LES DONNÉES CLINIQUES SUR LA COVID-19**

Pour évaluer la qualité des données recueillies sur les eaux usées et la validité des procédures de normalisation appliquées, les données brutes ou corrigées sur les eaux usées sont généralement comparées aux données cliniques disponibles sur la COVID-19. Cette comparaison est faite soit visuellement, soit en utilisant une sorte de mesure de corrélation. Il y a plusieurs différences dans la nature des données sur les eaux usées et les données cliniques qu'il faut garder à l'esprit. Premièrement, l'ensemble de données cliniques est de type binaire. Autrement dit, un point de données cliniques n'a que deux valeurs possibles : « infecté » ou « non infecté ». En revanche, les données sur les virus dans les eaux usées peuvent être considérées comme semi-continues (c.-à-d. comportement continu à de grandes concentrations et comportement discret à de faibles concentrations). Une partie des personnes infectées excrètent périodiquement de grandes quantités de virus dans les eaux usées. Cette excrétion entraîne une étendue semi-continue de concentrations aux points d'échantillonnage situés plus loin dans le réseau d'égouts par mélange et par dispersion. Lorsque de nombreux cas de COVID-19 sont présents dans un grand bassin d'eaux usées urbain, les valeurs des séries chronologiques des deux sources de données montrent un comportement continu et relativement régulier. Cependant, avec seulement un petit nombre de cas de COVID-19, la nature binaire des données de santé publique deviendra apparente, et les résultats des eaux usées commenceront à montrer une variabilité plus considérable. Cette dernière n'est pas

seulement liée à la nature stochastique du processus d'excrétion (quantité, moment, emplacement), mais aussi celle du processus d'échantillonnage. Le risque d'avoir un certain nombre de copies de gènes viraux dans un échantillon est distribué selon la loi de Poisson. La stochasticité devient donc non négligeable à l'approche de la limite de quantification (LQ) et de la limite de détection (LD) des techniques biomoléculaires employées, ce qui entraîne une erreur dans l'estimation du signal viral des eaux usées inversement proportionnelle à sa valeur absolue (Wade et coll., 2022).

Lorsqu'on compare le signal viral des eaux usées aux données de cas cliniques, on peut choisir parmi plusieurs paramètres de santé publique, comme les nouveaux cas, les cas actifs, le taux de positivité, les hospitalisations et les décès. Il n'y a pas de norme de référence qui puisse être considérée comme la bonne dans tous les cas. La normalisation vise à relier les données sur les eaux usées à la situation réelle sur le terrain, mais il est parfois impossible de parfaitement mesurer cette réalité. Dans certains cas, le signal des eaux usées pourrait même être un meilleur indicateur de la situation épidémiologique actuelle de la population que des mesures épidémiologiques classiques. En premier lieu, les tests de dépistage de la COVID-19 en santé publique sont rarement effectués au hasard. Le processus dépend des recommandations des organismes de santé publique locaux (qui varient au fil du temps) et se concentre généralement sur les personnes présentant des symptômes, les personnes en contact avec des cas confirmés de COVID-19, et les personnes à risque. Le taux de positivité peut donner une indication du caractère aléatoire de la stratégie de dépistage, selon quoi l'OMS suggère que des valeurs de positivité supérieures à 10 % indiquent un dépistage insuffisamment complet. Les hospitalisations, les admissions aux soins intensifs et les décès ne devraient pas, en principe, être touchés par le type de système de dépistage de santé publique ou par la volonté des gens de se faire tester. Toutefois, on peut s'attendre à des différences selon le statut immunitaire (vacciné et rétabli), la répartition de la population selon l'âge, la politique d'admission à l'hôpital (en raison de la COVID-19 ou de la gravité de la maladie) et la qualité des soins. Toutefois, l'information de ces paramètres montre nécessairement un retard par rapport à l'état réel de transmission de la COVID-19 au sein de la population et accuse un retard considérable par rapport aux données sur les eaux usées. L'ampleur du décalage dépend du variant de la COVID-19 et diffère pour chaque cas. L'écart entre les données des eaux usées et les données épidémiologiques a, par conséquent, une nature stochastique particulière qui s'accroît lorsque le taux d'infection est faible. Cela complique davantage la comparaison entre ces deux types de données. Le paramètre des cas actifs comporte des lacunes, car il nécessite une surveillance continue et des tests réguliers pour chaque cas. Lorsque cette capacité de test fait défaut, un nouveau cas est souvent considéré comme actif pendant une période standardisée. Le paramètre des nouveaux cas est le plus fréquemment utilisé dans la littérature pour comparer avec les données sur les eaux usées, ce qui est logique, car c'est le paramètre exhibant le plus petit décalage. L'apparition des symptômes devrait, en principe, être un indicateur plus précoce que les nouveaux cas détectés, mais c'est un paramètre moins fiable, car il dépend des déclarations des patients et parce qu'une quantité considérable de personnes est asymptomatique. En fait, des études ont montré qu'avec un effort de dépistage suffisamment important, le décalage entre les cas actifs et les données sur les eaux usées est négligeable. Étant donné que la documentation sur l'excrétion fécale indique une libération maximale du virus au début de l'infection, cela explique également pourquoi on peut s'attendre à une forte corrélation entre les eaux usées et les nouveaux cas (Hoffmann et Alsing, 2021).

Les problèmes mentionnés ci-dessus ne peuvent pas être résolus par une procédure de normalisation. Ils doivent plutôt être pris en considération lors de l'analyse des ensembles de

données sur l'EEU, c'est-à-dire en évaluant la pertinence des ensembles de données sélectionnés et des procédures d'analyse envisagées pour répondre à la question de recherche.

## **6. LA PRATIQUE DE LA NORMALISATION DANS LE CONTEXTE CANADIEN**

Lorsqu'on examine les tableaux de bord publics d'EEU à l'échelle du Canada disponibles sur le [site Web COVIDPops19](#) (dernière vérification le 22 mars 2022), on remarque une multitude de façons de présenter les résultats de mesure du SRAS-CoV-2 (tableau 1). Au début de la pandémie, la plupart des tableaux de bord offraient la possibilité de visualiser les résultats bruts quotidiens pour le gène N1 ou N2 ou les deux, exprimés en copies de gènes par ml d'eaux usées (gc/ml). Plus récemment, cependant, plusieurs tableaux de bord ont supprimé l'accès aux données brutes et ne montrent que les données transformées. Ce changement indique un consensus croissant selon lequel les données virales brutes (gc/ml) pourraient ne pas donner la meilleure image possible de la situation épidémiologique. Les transformations de données les plus courantes sont : prendre une moyenne sur 3 à 7 jours (mobile); faire la moyenne ou la somme des résultats des gènes N1 et N2; multiplier la concentration virale par le débit; la diviser par le signal de PMMoV, et représenter le signal sur une échelle log<sub>10</sub>. Parmi les transformations moins courantes, on trouve des algorithmes de lissage dédiés, des ajustements en fonction de l'efficacité de la procédure d'extraction de l'ARN, et l'agrégation de données de plusieurs sites d'échantillonnages. Les réponses par courriel de divers groupes de recherche canadiens interrogés sur leurs pratiques de normalisation mentionnaient l'utilisation des chiffres de population du bassin hydrographique pour effectuer la normalisation par la population et l'expression du signal viral par « masse humide » de solides.

Il est logique de lisser le signal viral lorsqu'on s'attend à un comportement lisse conforme à la théorie épidémiologique (c.-à-d. pour les grands bassins hydrographiques et la transmission communautaire bien développée). Cependant, on ne s'attend pas à un signal lisse a priori (c.-à-d. petits bassins hydrographiques, faible nombre de cas). Les méthodes utilisées au Canada incluent la simple moyenne hebdomadaire, la moyenne mobile sur des périodes différentes, ainsi que des algorithmes de lissage spécialisés. Le choix de la meilleure méthode et l'étalonnage de cette méthode sont probablement propres au cas, car ils dépendent des caractéristiques des bassins d'eaux usées, de la fréquence des mesures et d'autres facteurs. Cependant, la procédure de sélection pourrait être standardisée, comme dans Arabzadeh et coll. (2021), qui ont testé 13 techniques de filtrage de données différentes dans un cadre d'optimisation défini à l'aide de mesures de performance pour réduire le bruit dans le signal et améliorer sa corrélation avec les données de cas cliniques.

Cibler plusieurs gènes améliorera certainement la confiance dans les données virales. Les résultats de chaque gène peuvent être comparés afin de mettre en évidence tout problème d'analyse génétique spécifique. Lorsqu'on s'attend à ce que les résultats des essais sur différents gènes soient semblables, la moyenne ou la sommation des résultats pour ces gènes devrait réduire le bruit dans les données. Lorsque différents essais montrent des différences cohérentes, il est peut être approprié d'analyser les résultats pour chaque gène séparément. L'utilisation de l'efficacité d'extraction au moyen de l'injection d'un virus de contrôle externe aux eaux usées pour corriger le signal viral n'est pas courante dans la documentation, puisque les taux de récupération obtenus dans les eaux usées sont généralement très faibles et variables. Il en résulterait des facteurs de correction importants et incertains, qui devraient être traités avec prudence. Il est donc plus commun de désigner un interval de taux de récupération acceptables pour détecter des mesure erronées qui devront être reprises.

L'utilisation de l'échelle log<sub>10</sub> est conforme aux lignes directrices des CDC, tout comme l'utilisation du débit, des chiffres de population et des biomarqueurs. L'utilisation de concepts comme la « masse humide » devrait probablement être évitée dans le contexte de la normalisation des méthodes de mesure, car elle n'est pas clairement définie et elle varie probablement d'un laboratoire à l'autre et d'un site à l'autre. La multiplication des concentrations virales à partir d'un échantillon proportionnel au débit sur 24 h avec le débit quotidien moyen correspondant donne la charge quotidienne de virus qui passe à ce point d'échantillonnage. Remarque : tout autre type d'échantillonnage pour ce calcul donnerait des charges virales quotidiennes biaisées. La recommandation selon laquelle il est préférable d'utiliser les charges virales quotidiennes au lieu des concentrations pour la comparaison avec

**Tableau 1 : Aperçu de la façon dont les données d'EEU sont présentées sur les tableaux de bord partout au Canada**

<u>Information sur la COVID-19, Territoires du Nord-Ouest (T. N.-O.)</u>	Moyenne (gènes non spécifiés) (gc/ml)
<u>Centre for Health Informatics Alberta (Alb.)</u>	Moyenne mobile quotidienne et sur 3 jours de : N1 (copies/jour), N2 (copies/jour), moyenne N1 et N2 (copies/jour). La charge quotidienne est calculée avec le débit.
<u>Terre-Neuve-et-Labrador (T.-N.-L.)</u>	SRAS (gène non spécifié) (gc/mg), SRAS (gène non spécifié) (gc/ml)
<u>CentrEau-COVID (Qc)</u>	N1 (gc/ml), N1 (gc/ml) divisé par PMMoV (gc/ml)
<u>Simcoe-Muskoka Health District (Ont)</u>	Moyenne mobile sur 7 jours de : N1 (gc/ml), N1 (gc/ml) divisé par PMMoV (gc/ml)
<u>613covid (Ont)</u>	Moyenne quotidienne et mobile sur 7 jours de : somme de N1 et N2 (gc/ml) divisée par PMMoV (gc/ml)
<u>Leeds, Grenvill and Lanark District (Ont)</u>	Moyenne mobile sur 7 jours (signal viral non spécifié)
<u>KFL&amp;A Public Health (Ont)</u>	Moyenne géométrique quotidienne et mobile (plage non précisée) : SRAS (gène non spécifié) (gc/ml)
<u>Suivi des données sur la COVID-19 de Durham (Ont)</u>	Moyenne quotidienne et lissée (régression pondérée localement par LOWESS, plage non précisée) du SRAS (gène non spécifié) (gc/ml) sur l'échelle log <sub>10</sub>
<u>COVID-19 dans la région de York (Ont)</u>	Moyenne mobile sur 7 jours de : SRAS (gène non spécifié) (gc/ml) divisé par PMMoV (gc/ml), SRAS (non spécifié) (gc/ml)
<u>TABLEAU DE BORD DE L'ONTARIO (Ont)</u>	Moyenne pondérée à l'échelle de la province des résultats normalisés du SRAS (N1 et N2) normalisés par biomarqueurs (PMMoV) provenant de plusieurs emplacements (unités inconnues).
<u>Région de Waterloo (Ont)</u>	Moyenne mobile sur 7 jours pour : SRAS (gène non spécifié) (gc/ml); SRAS (gène non spécifié) (gc/ml) divisé par PMMoV (gc/ml)

<u>Thunder Bay District Health Unit (Ont)</u>	Moyenne quotidienne et mobile sur 7 jours pour : N1 (gc/ml) divisé par PMMoV
<u>Global Institute for Water Security (Sask.)</u>	Moyenne sur 7 jours pour : SRAS ajusté en fonction de l'efficacité de l'extraction virale (gène non spécifié) (gc/100 ml)
<u>Région métropolitaine de Vancouver (C.-B.)</u>	SRAS (gène non spécifié) (gc/ml), SRAS (gène non spécifié) (copies/jour). La charge quotidienne est calculée en fonction du débit.

les cas cliniques quotidiens est intuitive et n'a pas besoin d'explications supplémentaires. La multiplication par le débit est une procédure standard en génie des eaux usées pour calculer les charges quotidiennes pour divers paramètres. Cette pratique est souvent appelée « normalisation du débit » en EEU, car elle aide à tenir compte des effets de dilution. Elle ne corrige cependant pas les effets secondaires du débit dans les égouts (p. ex., décantation et remise en suspension des particules). L'analyse des données de débit doit donc être étendue et amenée à un autre niveau comme le décrit la section §7. Pour comparer les concentrations virales entre les lieux d'échantillonnage, il est courant dans la documentation d'effectuer une mise à l'échelle de la population en tenant compte du nombre de personnes dans les bassins d'eaux usées et d'exprimer les données en charge virale par personne et par jour. La normalisation de la population se produit lorsque des estimations dynamiques de la population sont utilisées pour corriger les variations de la population dans les bassins d'eaux usées. Il est possible d'obtenir ces estimations dynamiques de plusieurs façons en utilisant les données sur la population et le débit, ou des biomarqueurs liés aux humains, comme nous le verrons dans la section §7. L'agrégation des signaux viraux provenant de plusieurs emplacements en une seule mesure dépasse la portée du présent rapport.

Il est intéressant de noter que la gamme de paramètres auxiliaires utilisés pour la transformation des données dans le contexte canadien s'est jusqu'à présent limitée au débit et au PMMoV. Les expériences de différents groupes de recherche canadiens avec le PMMoV ont révélé un degré de satisfaction variable, de « grande satisfaction » à « je ne l'utilise plus ». Selon le bassin d'eaux usées, on mentionne des anomalies quotidiennes sporadiques, des biais continus sur des périodes moyennes à longues (fonte des neiges) et des variations saisonnières persistantes. Pour certains groupes de recherche, la normalisation avec le PMMoV n'élimine aucune variabilité du signal viral. Cependant, il demeure utile pour vérifier l'intégrité de l'échantillon et le rendement de l'extraction (c.-à-d. comportement anormal, valeurs aberrantes). Les valeurs du PMMoV observées semblent être spécifiques à l'emplacement et tendent à montrer un comportement plus erratique pour les petits bassins hydrographiques où les choix alimentaires individuels ne sont probablement pas suffisamment égalisés suivant la loi des grands nombres. Compte tenu de ces résultats mitigés pour le PMMoV, il est logique de commencer à examiner plus attentivement divers paramètres communs de qualité des eaux usées comme le  $\text{NH}_4$ , la DBO, les MeS, etc., pour effectuer la normalisation. Ces paramètres ont un lien clair avec les rejets fécaux humains; leurs méthodes de mesure sont bien définies et donnent des résultats reproductibles avec des barres d'erreur plus petites; et leur méthode d'échantillonnage est parfaitement compatible avec les exigences requises pour le SRAS-CoV-2 (p. ex., réfrigération et analyse en quelques jours). De plus, il a déjà été démontré qu'une combinaison de ces paramètres communs de la qualité de l'eau a contribué à améliorer les prédictions de cas de SRAS-CoV-2 d'un modèle d'apprentissage automatique comparativement à l'utilisation de mesures virales du SRAS-CoV-2 et du PMMoV uniquement (Therrien et coll., 2022).

## 7. UTILISATION DE BIOMARQUEURS ET DE PARAMÈTRES AUXILIAIRES DANS LE CONTEXTE INTERNATIONAL

Les études sur les pratiques exemplaires de normalisation n'ont été publiées que récemment. Dans une étude exhaustive des incertitudes liées aux mesures du SRAS-CoV-2 pour l'EEU au Royaume-Uni, Wade et coll. (2022) explorent l'utilisation du débit, de l'ammoniac, de l'orthophosphate, du virus crAssphage et des antidépresseurs pour normaliser le signal viral. Bien que le débit soit considéré comme le paramètre le plus critique pour corriger les variations temporelles des signaux dues à la dilution, il n'est pas jugé suffisant pour normaliser la population et obtenir des estimations dynamiques de la population. Pour ce faire, il est suggéré d'utiliser des marqueurs biochimiques tels que l'ammoniac et l'orthophosphate en comparant les valeurs quotidiennes réelles et prévues des rejets par habitant par site. Lorsque les données sur le débit ne sont pas disponibles pour corriger la variabilité temporelle, on peut utiliser des techniques de normalisation indirecte utilisant des marqueurs biochimiques comme substitut du débit en supposant qu'ils sont présents dans le bassin d'eaux usées à charge constante. On a constaté que la dilution avait une incidence mineure sur les charges de SRAS-CoV-2 dans les eaux usées, les changements importants ne se produisant que lors de fortes précipitations ou de rejets provenant d'autres sources. En ce qui concerne l'utilisation d'antidépresseurs, les métabolites (et non les composés d'origine) ont donné des résultats prometteurs en tant que marqueurs de population. Les biomarqueurs fécaux comme le virus crAssphage sont considérés comme utiles, mais ne sont pas le premier choix pour la normalisation.

Dans le programme anglais, lorsque les données sur le débit n'étaient pas disponibles, les données sur le virus crAssphage ont d'abord été utilisées pour aider à normaliser les résultats du SRAS-CoV-2 afin de tenir compte de la dilution par les eaux usées industrielles et les précipitations. Cependant, cela a créé une charge de travail supplémentaire et a retardé le travail de laboratoire. Il a donc été abandonné au profit d'autres indicateurs de charge fécale. Une approche multi-biomarqueurs (p. ex., azote ammoniacal et orthophosphate) a été proposée pour fournir des estimations de la dilution plus robustes, puisqu'il est impossible de faire la distinction entre une diminution du débit et une augmentation de la charge de marqueurs (par exemple, en raison d'un rejet industriel ou agricole unique) en utilisant un seul marqueur. Le modèle multi-biomarqueurs peut identifier des valeurs aberrantes comme les rejets ponctuels et estimer la variabilité du débit si au moins un marqueur est quantifié, mais avec des barres d'erreur plus grandes. Lorsqu'aucun marqueur n'est quantifié (données manquantes), le modèle prédit des flux moyens avec des barres d'erreur importantes. Il est essentiel, avec cette approche, que la variabilité naturelle des charges des biomarqueurs soit prise en compte en attribuant une importance variable à différents marqueurs. Par exemple, une plus grande priorité devrait être accordée à l'azote ammoniacal puisque les concentrations de copies du gène du virus crAssphage présentent une plus grande variabilité naturelle, bien que les deux permettent d'estimer la dilution. Dans le programme écossais, lorsque les données de débit ne sont pas disponibles pour un site particulier, on utilise un modèle linéaire mixte basé sur le débit, le  $\text{NH}_4$  et les données démographiques de plusieurs sites. Si aucun débit ni  $\text{NH}_4$  n'est disponible (données manquantes), une fonction spline utilisant les tendances récentes de l'ammoniac est utilisée (ajustée selon les tendances nationales en plus des effets spécifiques au site). Dans le cas d'importants débordements d'égouts unitaires (DEU), il est préférable de normaliser les données sur le  $\text{NH}_4$  plutôt que les données sur le débit.

Langeveld et coll. (2021) ont étudié trois méthodes de normalisation utilisant le débit, la conductivité électrique (CE) et le virus crAssphage pour plusieurs sites. Ils proposent une normalisation du débit soutenue par un contrôle de qualité utilisant la surveillance de la conductivité

comme méthode de normalisation à préconiser dans le cas où la surveillance détaillée du débit est connue ou peut l'être. Les charges du virus CrAssphage se sont révélées très constantes dans le temps et dans l'espace, et une normalisation directe fondée sur le virus crAssphage a pu être appliquée de façon fiable pour les populations d'environ 5 000 personnes et plus. Contrairement à ce que rapporte Wade et coll. (2022), le débit est jugé suffisant pour effectuer la normalisation de la population en estimant la proportion d'eaux usées domestiques dans le débit total en supposant un débit domestique fixe par temps sec par personne. La normalisation selon la CE et crAssphage nécessite initialement une analyse approfondie des données historiques sur les flux, la CE et crAssphage, mais ne nécessite pas de nouvelles mesures quotidiennes des flux par la suite. On a constaté d'importantes différences dans les facteurs de correction, ce qui indique la nécessité de normaliser les résultats pour les différentes régions. La méthode de la CE est sensible au dégivrage des routes et devrait être utilisée avec prudence dans un contexte canadien. L'utilisation combinée des trois paramètres fournit un niveau supplémentaire de contrôle de la qualité et pourrait détecter des problèmes occasionnels ou persistants propres à un site en ce qui concerne le débit des eaux usées (défaillances de la pompe, DEU, etc.) ou leur qualité (sels, composés industriels, etc.). En l'absence de données sur le débit (actuelles et historiques) pour déterminer la proportion d'eaux usées domestiques au moyen de l'une des méthodes ci-dessus, la normalisation directe du SRAS-CoV-2/crAssphage est suggérée comme option valide pour générer les mêmes tendances. Les résultats obtenus avec le virus crAssphage semblent saisonnièrement stables et sont comparables entre les différents sites même si la variation entre les échantillons de selles individuels est considérable. Le virus crAssphage est persistant sur le plan environnemental et ne montre aucune accumulation significative dans les sédiments dans le réseau d'égouts, contrairement à d'autres paramètres comme la DBO ou les MeS.

Une partie des efforts de normalisation de Wade et coll. (2022) est décrite plus en détail dans Sweetapple et coll. (2021). Lorsque des données sur les flux et la population sont disponibles, le  $\text{NH}_4$  et le  $\text{PO}_4$  peuvent être utilisés pour obtenir des estimations dynamiques de la population et la normaliser. En l'absence de données sur les flux et la population, la normalisation directe (division par  $\text{NH}_4$  ou  $\text{PO}_4$ ) peut être utilisée pour obtenir les mêmes tendances, tout comme l'utilisation du virus crAssphage par Langeveld et coll. (2021). Cependant, le signal normalisé direct n'est valide que localement (c.-à-d. pour chaque site). On a constaté que la normalisation de la population tenait compte de la dilution et des variations de la population, mais dans l'ensemble, elle n'a pas été considérée comme ayant une incidence sur les sites étudiés et la période étudiée (en raison du confinement). Bien qu'elle ne soit pas négligeable à court terme, la normalisation n'a pas beaucoup modifié l'analyse des tendances à long terme. Les corrélations avec les données de santé publique étaient généralement plus faibles pour les sites plus petits et proches de la source. L'approche normalisée selon la population a permis de comparer les sites, mais elle a été jugée insuffisante comme seul facteur d'importance. Les données montrent un degré élevé de variation des charges quotidiennes par habitant de  $\text{NH}_4$  et  $\text{PO}_4$  entre les différents sites, possiblement en raison des rejets industriels et de l'incertitude quant à l'attribution des données démographiques du recensement aux bassins d'eaux usées. De plus, d'autres différences entre les sites qui pourraient avoir une incidence, comme le temps de rétention hydraulique ou la technique d'échantillonnage, pourraient devoir être considérées plus explicitement. On a tenté d'inclure les métadonnées de la campagne d'échantillonnage, de la qualité des eaux usées et du bassin d'eaux usées pour améliorer la prédiction de la prévalence fondée sur l'EEU au moyen d'une approche d'apprentissage automatique et on a obtenu des résultats prometteurs (Morvan et coll., 2021).

Alabzadeh et coll. (2021) ont employé le  $\text{NH}_4$  pour la normalisation de la population, mais n'ont pas abordé les avantages potentiels, car ce n'était pas l'objet de l'étude. Les auteurs ont noté un certain

seuil de charge virale pour les petits bassins hydrographiques avant de pouvoir constater une relation claire entre le signal des eaux usées et la dynamique de l'infection de la population, et ils affirment également que les grappes d'infection semblent avoir un effet plus important dans les petits bassins hydrographiques. Nagarkar et coll. (2022) ont étudié la normalisation des signaux N1/N2 SRAS-CoV-2 dans les eaux usées pour trois réseaux d'égouts de tailles différentes en utilisant l'efficacité de récupération matricielle (souche OC43 du bêtacoronavirus-1), le débit et trois marqueurs fécaux : crAssphage, PMMoV et HF183. La normalisation directe a été utilisée simplement en divisant les signaux respectifs. Les résultats ont montré que crAssphage et HF183 se comportent de manière mutuellement similaire, mais différente du virus PMMoV. De plus, on a constaté une nette différence dans le comportement de normalisation des différents bassins d'eaux usées. Cependant, les conclusions fermes sur les pratiques exemplaires de normalisation sont malheureusement entravées par la longueur limitée des ensembles de données et la fréquence d'échantillonnage hebdomadaire. Greenwald et coll. (2021) ont comparé crAssphage, PMMoV, l'ARN ribosomique 16S des bactéroïdes HF183 et l'ARN ribosomique 18S humain pour la normalisation directe. Les virus crAssphage et PMMoV étaient les biomarqueurs les plus cohérents; toutefois, ils n'ont pas renforcé la corrélation avec les données cliniques. La normalisation avec le virus PMMoV a même entraîné une détérioration significative de la corrélation. Malgré la procédure rigoureuse d'analyse des données qui a été suivie, la fréquence d'échantillonnage hebdomadaire limitée entrave malheureusement les conclusions solides. Schoen et coll. (2022) suggèrent une fréquence d'échantillonnage minimale de quatre jours par semaine pour l'EEU, fondée sur une analyse qui repose sur un modèle d'ensembles de données artificiels à fréquence réduite des valeurs du SRAS/PMMoV à partir d'échantillons instantanés quotidiens de boues de clarificateur primaire.

## **8. CONSIDÉRATIONS SUPPLÉMENTAIRES RELATIVES À LA PROCÉDURE DE NORMALISATION**

Comme c'est souvent le cas, la qualité des données brutes est primordiale pour parvenir à des conclusions significatives. Comme la correction des données après coup est lourde et souvent impossible, toute évaluation fiable de la prévalence de la COVID-19 fondée sur les données d'EEU devrait commencer par une campagne d'échantillonnage et une procédure de mesure bien conçues. Idéalement, les échantillons devraient être prélevés quotidiennement, avec des échantillonneurs composites en fonction du débit en 24 heures. Le type d'égout, le lieu d'échantillonnage dans l'égout et la cible d'échantillonnage (solides, liquides) détermineront dans une large mesure les limites de ce qui peut être réalisé avec les données. L'incertitude entourant les mesures en laboratoire peut être réduite au minimum en utilisant des procédures rigoureuses d'AQ/CQ (Chik et coll., 2022), et les estimations de la qualité des données peuvent être obtenues par une note pondérée de plusieurs paramètres d'AQ/CQ (Kantor et coll., 2022). Ces estimations pourraient faire l'objet d'une analyse poussée des données de façon directe ou indirecte.

Avant l'analyse des données, plusieurs auteurs transforment leurs ensembles de données sur le SRAS-CoV-2 en log10. Cette procédure est souvent utilisée lorsque l'on travaille avec des données fortement biaisées vers la droite et modifie l'importance relative des petites et des grandes valeurs. Pour que cette transformation de données fonctionne, des valeurs nulles dans l'ensemble de données d'EEU doivent être remplacées par une constante. En général, la moitié de la valeur de la LD est utilisée. Toutefois, il est probablement préférable d'omettre ces points de données lors de la vérification des procédures de normalisation (Wade et coll., 2022). Avec l'EEU, on suppose que les données sur le SRAS-CoV-2 sont distribuées normalement. Toutefois, ce pourrait ne pas être le cas,

en particulier pour les petites valeurs proches de la LD et de la LQ, comme nous l'avons vu plus haut dans la section §5. Les petites valeurs posent un problème particulier relativement à l'EEU, car elles sont assorties d'une grande incertitude, mais présentent un grand intérêt pour la santé publique lorsqu'on tente de détecter le début d'une nouvelle vague épidémiologique. Un problème supplémentaire spécifique aux milieux où le nombre de cas est faible est le masquage des données cliniques (c.-à-d. ne pas mentionner le nombre exact de cas en dessous d'un certain seuil, p. ex., cinq cas par jour) par des préoccupations éthiques sur la traçabilité des patients. Un autre problème survient lorsque la prévalence est faible, soit le comportement asymétrique entre les données cliniques et les données d'EEU (c.-à-d. la détection de cas positifs dans les eaux usées, mais pas cliniquement, ou l'inverse), ce qui entraîne des scores de corrélation très faibles. Pour toutes ces raisons, il pourrait être utile d'évaluer séparément les données des situations de faible incidence dans un cadre statistique dédié afin d'accroître la confiance dans les décisions qui peuvent être prises à partir de ces données.

Les algorithmes de lissage des données et de détection des valeurs aberrantes doivent être évalués pour chaque ensemble de données et chaque situation spécifique (voir section §6), comme c'est le cas pour les différentes procédures de normalisation possibles. Les améliorations en matière de lissage et de normalisation sont généralement évaluées sous la forme d'une meilleure corrélation entre les données ajustées en fonction de l'EEU et les données de cas cliniques. La relation entre les données d'EEU et les données cliniques sur les cas peut changer à la fois en termes d'ampleur et de délai (p. ex., en raison d'un autre variant du virus, de changements dans l'état de vaccination de la population ou de régime de tests cliniques). Il faut donc choisir soigneusement la période de données à utiliser (p. ex., une vague à la fois) et garder à l'esprit les limites inhérentes de l'ensemble de données (p. ex., les attentes en matière de comportement des données pour les grandes villes par rapport aux petites municipalités; section §5).

Dans la documentation, le degré de corrélation est très souvent vérifié à l'aide du coefficient de corrélation de Pearson ou d'une variante de celui-ci, comme le coefficient  $r$  bayésien Pearson (Nagarkar et coll., 2022) ou le test Tau-b de Kendall, un coefficient de corrélation non paramétrique adapté aux données censurées à gauche (c.-à-d. ensembles de données dont les données sont inférieures à la limite inférieure de détection; Greenwald et coll., 2021). Avec un coefficient de corrélation pondéré de Pearson, l'incertitude des données pourrait être incorporée. Le coefficient de corrélation de Pearson mesure la corrélation linéaire des variables et ne tient donc pas compte de nombreux autres types de relations ou de corrélations. C'est une statistique intuitive et facile à comprendre qui n'exige pas que ses variables aient la même unité ou la même plage de valeurs. Il est toutefois affecté par le bruit et les valeurs aberrantes, et son inférence statistique est sensible au fait que la distribution des données est normale, surtout avec de petits ensembles de données. En utilisant le coefficient de corrélation de Pearson pour les ensembles de données d'EEU, on ignore sciemment les propriétés statistiques des données (non choisies au hasard, autocorrélation, non normalement distribuées). Pourtant, l'utilisation du coefficient permet encore de réfléchir aux avantages de la normalisation (Arabzadeh et coll., 2021). L'analyse de régression linéaire est aussi souvent utilisée avec le coefficient de détermination (ajusté) indiquant la qualité de l'ajustement du modèle de régression linéaire. Les propriétés statistiques des données sont généralement ignorées ici aussi. Les modèles de régression statistique non linéaire ou d'autres types de modèles descriptifs et prédictifs (p. ex., apprentissage automatique) peuvent également être utilisés pour évaluer les avantages de la normalisation par les biomarqueurs. La qualité de l'ajustement peut être exprimée par une grande variété de scores. Inversement, les résidus peuvent être testés pour l'autocorrélation, l'homoscédasticité, leur distribution statistique, etc.

## 9. CONCLUSIONS

Un aperçu des motifs, des pratiques employées et des problèmes entourant la normalisation du signal viral du SRAS-CoV-2 à partir d'échantillons d'eaux usées a été présenté. Lors du choix des biomarqueurs et des paramètres auxiliaires pour corriger les influences indésirables dans les données d'EEU, il est logique d'évaluer d'abord l'incertitude liée au paramètre de normalisation utilisé et de considérer les processus par lesquels ce paramètre lui-même est influencé. Il est également logique d'aborder la normalisation par étapes. Par exemple, le débit peut être mesuré avec précision et est sans ambiguïté lié à la dilution, de sorte que son utilisation pour la correction de la dilution est évidente. Les estimations de population pour la normalisation de la population peuvent être obtenues avec de nombreux paramètres. Cependant, chacun présente un degré différent d'incertitude intrinsèque (p. ex., source, stabilité, transport par égout, mesure) qui permet de les classer selon leur pertinence : débit > paramètres communs de la qualité de l'eau > contaminants émergents préoccupants > indicateurs viraux comme crAssphage et PMMoV. L'utilisation de plusieurs paramètres à la fois (de manière combinée) pour normaliser un effet indésirable spécifique réduit l'incertitude dans la procédure de correction par rapport à l'utilisation d'un seul paramètre. Il faudrait étudier des combinaisons spécifiques de paramètres (p. ex., indicateur de solubilité, indicateur particulière, indicateur viral, etc.) pour déterminer leur capacité à normaliser d'autres effets indésirables qui n'ont pas été abordés à fond dans la littérature jusqu'à maintenant (processus de transport des égouts, erreur d'échantillonnage, qualité des mesures en laboratoire, etc.).

## 10. RÉFÉRENCES

- Arabzadeh, R., Grünbacher, D. M., Insam, H., Kreuzinger, N., Markt, R., & Rauch, W. (2021). Data filtering methods for SARS-CoV-2 wastewater surveillance. *Water Science and Technology*, 84(6), 1324–1339. <https://doi.org/10.2166/wst.2021.343>
- Chik, A., Abbey, A.-M., Flemming, C., Fletcher, T., Ho, J., Oke, M., Pileggi, V., Raby, M., Simhon, A., Thomas, J., Weir, S., D'Aoust, P., Brown, S., Chao, T.-C., Daigle, J., Delatolla, R., Dhiyebi, H., Donaldson, M., Edwards, E., & Xie, Y. (2022). Protocol for Evaluations of RT-qPCR Performance Characteristics: Technical Guidance. <https://doi.org/10.13140/RG.2.2.10627.02080>
- Chik, A. H. S., Glier, M. B., Servos, M., Mangat, C. S., Pang, X. L., Qiu, Y., D'Aoust, P. M., Burnet, J. B., Delatolla, R., Dorner, S., Geng, Q., Giesy, J. P., McKay, R. M., Mulvey, M. R., Prystajeky, N., Srikanthan, N., Xie, Y., Conant, B., & Hrudey, S. E. (2021). Comparison of approaches to quantify SARS-CoV-2 in wastewater using RT-qPCR: Results and implications from a collaborative inter-laboratory study in Canada. *Journal of Environmental Sciences (China)*, 107, 218–229. <https://doi.org/10.1016/j.jes.2021.01.029>
- Greenwald, H. D., Kennedy, L. C., Hinkle, A., Whitney, O. N., Fan, V. B., Crits-Christoph, A., Harris-Lovett, S., Flamholz, A. I., Al-Shayeb, B., Liao, L. D., Beyers, M., Brown, D., Chakrabarti, A. R., Dow, J., Frost, D., Koekemoer, M., Lynch, C., Sarkar, P., White, E., ... Nelson, K. L. (2021). Tools for interpretation of wastewater SARS-CoV-2 temporal and spatial trends demonstrated with data collected in the San Francisco Bay Area. *Water Research X*, 12. <https://doi.org/10.1016/j.wroa.2021.100111>
- Hoffmann, T., & Alsing, J. (2021). Faecal shedding models for SARS-CoV-2 RNA amongst hospitalised patients and implications for wastewater-based epidemiology. *MedRxiv*, 2021.03.16.21253603. <https://doi.org/10.1101/2021.03.16.21253603>

- Kantor, R. S., Greenwald, H. D., Kennedy, L. C., Hinkle, A., Harris-Lovett, S., Metzger, M., Thornton, M. M., Paluba, J. M., & Nelson, K. L. (2022). Operationalizing a routine wastewater monitoring laboratory for SARS-CoV-2. *PLOS Water*, 1(2). <https://doi.org/10.1371/journal.pwat.0000007>
- Langeveld, J., Schilperoort, R., Heijnen, L., Elsinga, G., Schapendonk, C. E. M., Fanoy, E., de Schepper, E. I. T., Koopmans, M. P. G., de Graaf, M., & Medema, G. (2021). Normalisation of SARS-CoV-2 concentrations in wastewater: the use of flow, conductivity and CrAssphage. *MedRxiv*, 2021.11.30.21266889. <https://doi.org/10.1101/2021.11.30.21266889>
- Morvan, M., lo Jacomo, A., Souque, C., Wade, M., Hoffmann, T., Pouwels, K., Singer, A., Bunce, J., Engeli, A., Grimsley, J., O'Reilly, K., & Danon, L. (2021). Estimating SARS-CoV-2 prevalence from large-scale wastewater surveillance: insights from combined analysis of 44 sites in England. <https://doi.org/10.21203/rs.3.rs-770963/v1>
- Nagarkar, M., Keely, S. P., Jahne, M., Wheaton, E., Hart, C., Smith, B., Garland, J., Varughese, E. A., Braam, A., Wiechman, B., Morris, B., & Brinkman, N. E. (2022). SARS-CoV-2 monitoring at three sewersheds of different scales and complexity demonstrates distinctive relationships between wastewater measurements and COVID-19 case data. *Science of the Total Environment*, 816. <https://doi.org/10.1016/j.scitotenv.2021.151534>
- Ort, C., Lawrence, M. G., Reungoat, J., & Mueller, J. F. (2010). Sampling for PPCPs in Wastewater Systems: Comparison of Different Sampling Modes and Optimization Strategies. *Environmental Science & Technology*, 44(16), 6289–6296. <https://doi.org/10.1021/es100778d>
- Schang, C., Crosbie, N. D., Nolan, M., Poon, R., Wang, M., Jex, A., John, N., Baker, L., Scales, P., Schmidt, J., Thorley, B. R., Hill, K., Zamyadi, A., Tseng, C.-W., Henry, R., Kolotelo, P., Langeveld, J., Schilperoort, R., Shi, B., ... McCarthy, D. T. (2021). Passive Sampling of SARS-CoV-2 for Wastewater Surveillance. *Environmental Science & Technology*, 55(15), 10432–10441. <https://doi.org/10.1021/acs.est.1c01530>
- Schoen, M. E., Wolfe, M. K., Li, L., Duong, D., White, B. J., Hughes, B., & Boehm, A. B. (2022). SARS-CoV-2 RNA wastewater settled solids surveillance frequency and impact on predicted COVID-19 incidence using a distributed lag model. *MedRxiv*, 2022.02.21.22270864. <https://doi.org/10.1101/2022.02.21.22270864>
- Sweetapple, C., Wade, M. J., Grimsley, J. M. S., Bunce, J. T., Melville-Shreeve, P., & Chen, A. S. (2021). Population normalisation in wastewater-based epidemiology for improved understanding of SARS-CoV-2 prevalence: A multi-site study. *MedRxiv*, 2021.08.03.21261365. <https://doi.org/10.1101/2021.08.03.21261365>
- Therrien, J.-D., Maere, T., Halle, S., Dallaire, P., & Vanrolleghem, P. A. (2022). Using the right wastewater characteristics for early COVID-19 pandemic warning and forecast using deep machine-learning. 12th Urban Drainage Modeling Conference, 10-12 January, Costa-Mesa, California, USA.
- Wade, M. J., lo Jacomo, A., Armenise, E., Brown, M. R., Bunce, J. T., Cameron, G. J., Fang, Z., Farkas, K., Gilpin, D. F., Graham, D. W., Grimsley, J. M. S., Hart, A., Hoffmann, T., Jackson, K. J., Jones, D. L., Lilley, C. J., McGrath, J. W., McKinley, J. M., McSparron, C., ... Kasprzyk-Hordern, B. (2022). Understanding and managing uncertainty and variability for wastewater monitoring beyond the pandemic: Lessons learned from the United Kingdom national COVID-19 surveillance programmes. *Journal of Hazardous Materials*, 424. <https://doi.org/10.1016/j.jhazmat.2021.127456>