

Progrès dans le domaine de la surveillance

Des données pour agir : le rôle de la science des données dans la surveillance de la santé publique

Lundi 28 novembre 2023 12 h – 13 h (HNC) / 13 h – 14 h (HNE)

Conférenciers

Professeur Jude Kong

Professeur adjoint, Université York
Directeur général, ACADIC, AI4PEP

Professeur
Nathaniel Osgood

Professeur, Université de la Saskatchewan
Directeur, Laboratoire d'épidémiologie informatique et d'informatique de santé publique



National Collaborating Centre
for Infectious Diseases
Centre de collaboration nationale
des maladies infectieuses



Public Health
Agency of Canada

Agence de la santé
publique du Canada

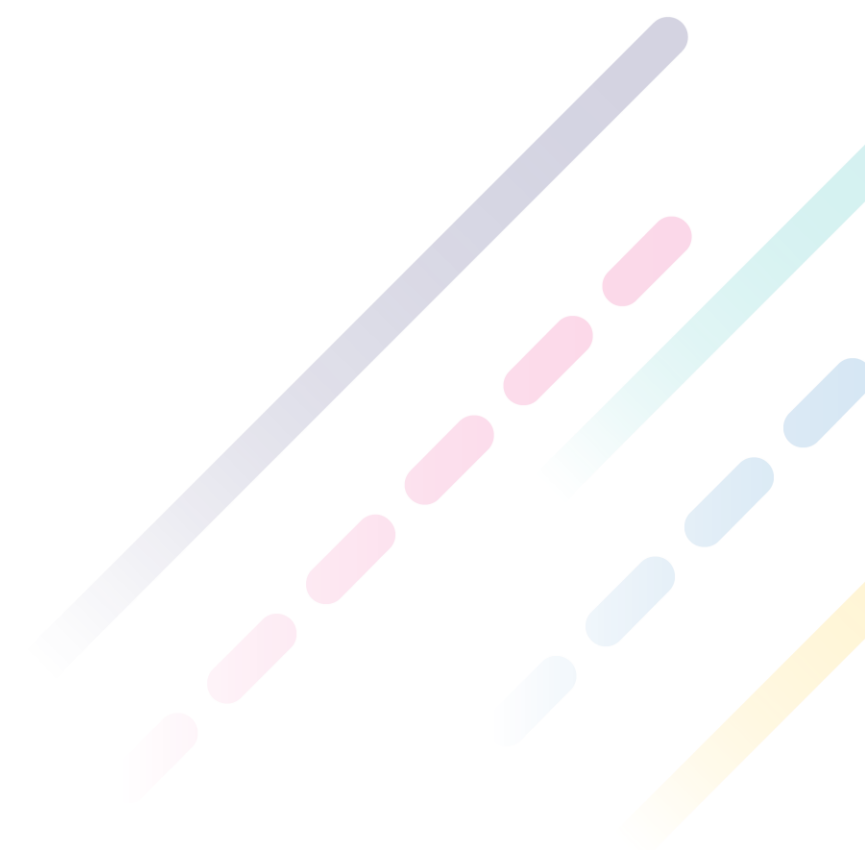


Reconnaissance du territoire : CCNMI



Le Centre de collaboration nationale des maladies infectieuses est hébergé à l'Université du Manitoba, sur les terres ancestrales des Anichinabés, des Cris, des Oji-Cris, des Dakotas et des Dénés, et la patrie de la nation métisse.

Au CCNMI, nous aspirons à honorer les terres et leurs gardiens originels dans notre travail. Nous reconnaissons que nous sommes sur la terre cédée en vertu du territoire numéro un. Nous reconnaissons que ce traité, comme d'autres, est issu du processus de colonisation visant à avantager certains et à nuire à d'autres. Nous sommes engagés à travailler avec nos partenaires vers la réconciliation.



Questions d'ordre administratif



- L'enregistrement et les diapositives de la présentation du séminaire seront accessibles sous peu sur le site Web du CCNMI : <https://ccnmi.ca/>.
- Si vous éprouvez des difficultés techniques avec Zoom, veuillez nous envoyer un courriel à nccid@umanitoba.ca.
- La messagerie instantanée pour les participants a été désactivée pour cette séance. Nous l'utiliserons pour transmettre des renseignements supplémentaires.
- Veuillez utiliser l'onglet des questions-réponses pour envoyer vos questions à nos conférenciers.
- Vous pouvez « aimer » les questions d'autres personnes pour les rendre prioritaires.



Agrément



« Progrès dans le domaine de la surveillance » est une activité de formation collective auto-approuvée agréée au titre de la section 1, conformément au programme de Maintien du certificat du **Collège royal des médecins et chirurgiens du Canada (CRMCC)**.

« Progrès dans le domaine de la surveillance » est également approuvée par le Conseil de Expérience Professionnelle pour les heures de développement professionnel des membres de **l'Institut Canadien des inspecteurs en santé publique (ICISP)**.

Si vous souhaitez obtenir une lettre de participation, veuillez répondre au sondage après le séminaire.

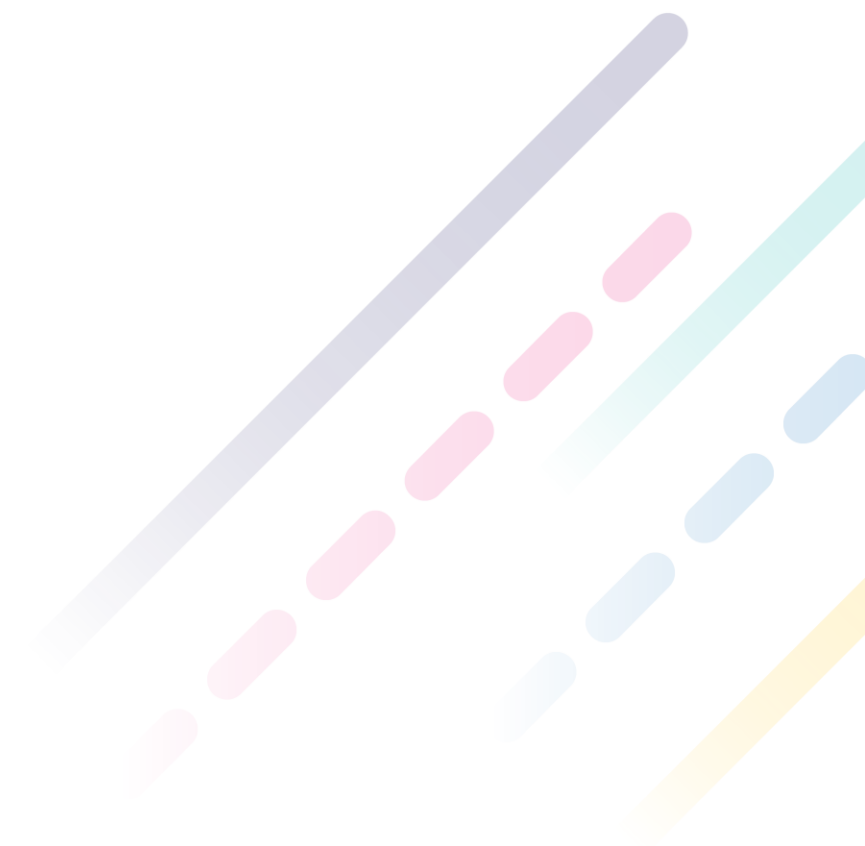


Reconnaissance des terres : ASPC



J'aimerais prendre le temps de reconnaître que le territoire sur lequel je vis et travaille est le territoire traditionnel des Wendat, des Anishnaabee, des Haudenosaunee, des Métis et de la Première Nation des Mississaugas de New Credit.

Elle abrite de nombreuses communautés métisses, inuites et des Premières Nations. Je suis reconnaissant d'avoir eu l'occasion de partager le lieu où elles habitent.



Conférenciers d'aujourd'hui



Dr. Jude Kong Phd

Directeur général, Africa-Canada Artificial Intelligence & Data Innovation Consortium, Global South AI for Pandemic and Epidemic Preparedness and Response Network (AI4EP)

Professeur adjoint, Université York

jdkong@yorku.ca

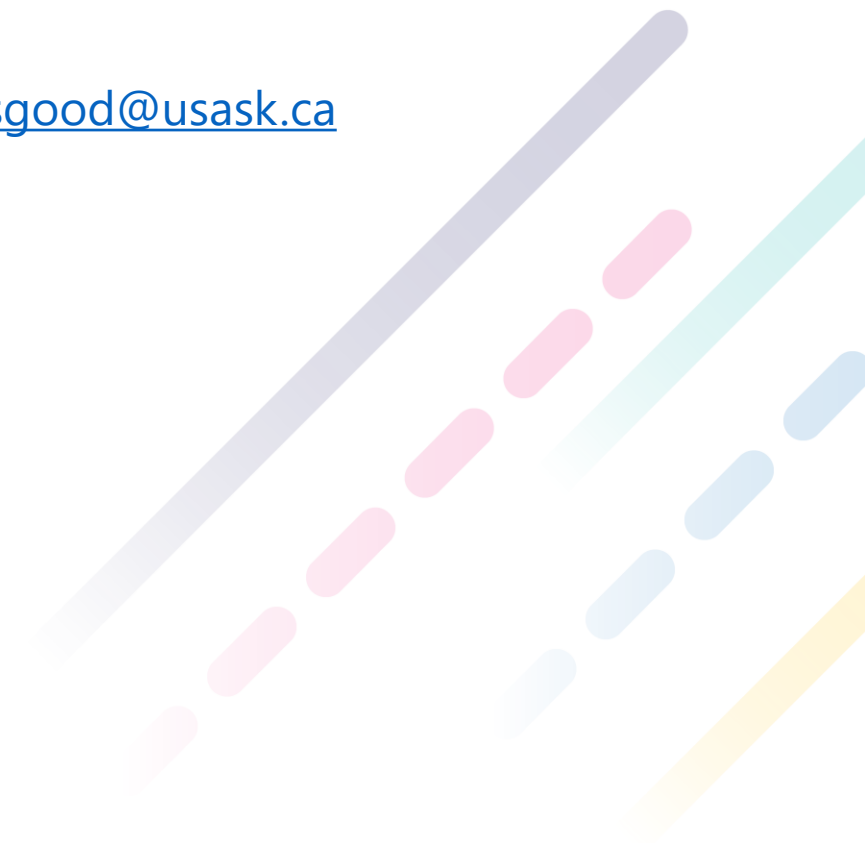


Dr. Nathaniel Osgood Phd

Professeur, Université de la Saskatchewan

Directeur, Laboratoire d'épidémiologie informatique et d'informatique de santé publique

nathaniel.osgood@usask.ca



Objectifs d'apprentissage



À l'issue de ce séminaire, vous serez en mesure de :

- Comprendre comment l'IA, la science des données et les solutions mathématiques peuvent être utilisées de manière responsable pour améliorer la surveillance de la santé publique et la réponse aux épidémies de maladies infectieuses émergentes et récurrentes
- Comprendre la faisabilité et les avantages de l'utilisation de l'IA pour permettre des mises à jour en temps réel de la modélisation de la transmission avec divers flux de données entrants
- Mettre en évidence les avantages en termes de précision obtenus en incorporant des sources de données sur les eaux usées parmi les ensembles de données réelles utilisées comme base pour les modèles
- Noter les fortes chances que la mise à jour conjointe des modèles de pathogènes multiples partageant des facteurs de risque communs avec les données de surveillance des pathogènes croisés offre des possibilités fort intéressantes



AI-Epidemix : Un outil de détection et de réponse aux épidémies soutenu par l'IA et une plateforme de collecte de données multisources en temps réel



Professeur Jude Kong, Ph. D.

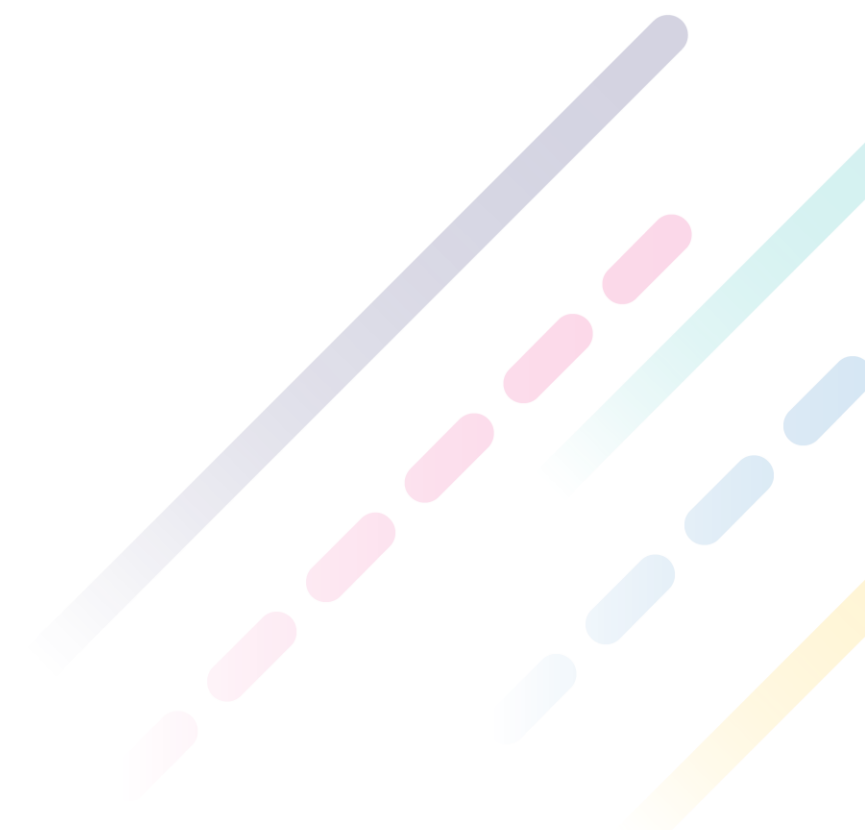
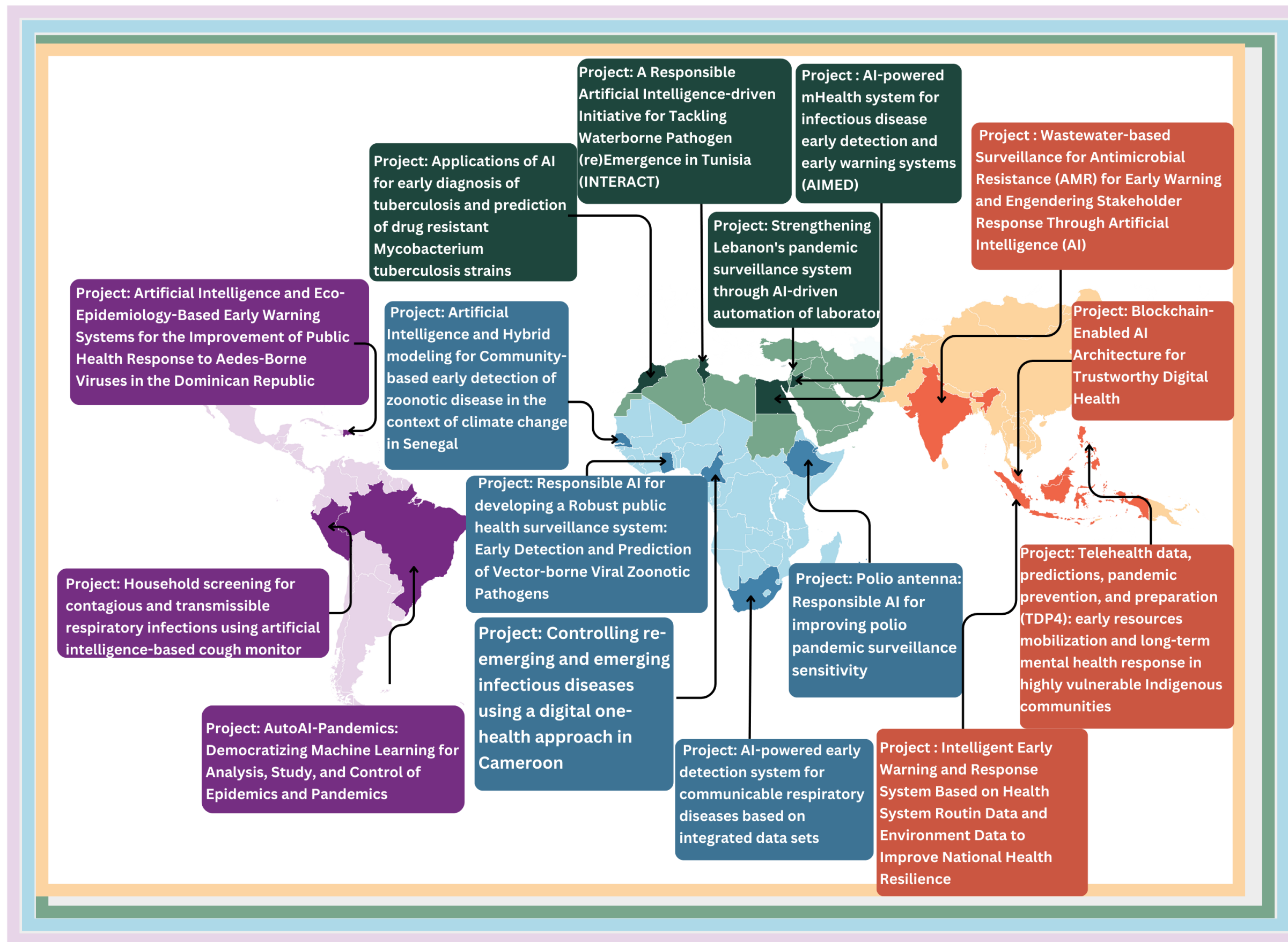
Directeur général, Africa-Canada Artificial Intelligence & Data Innovation Consortium

Directeur général, Global South AI for Pandemic and Epidemic Preparedness and Response Network (AI4EP)

Professeur adjoint, Département de mathématiques et de statistiques, Université York

Twitter : @dzevela



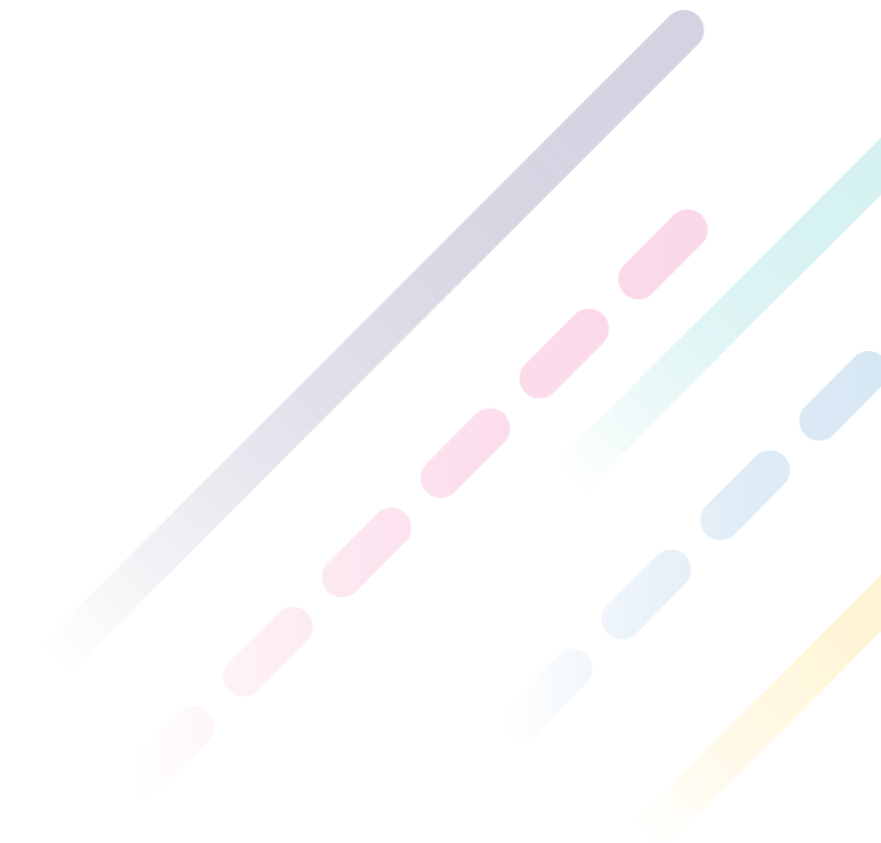




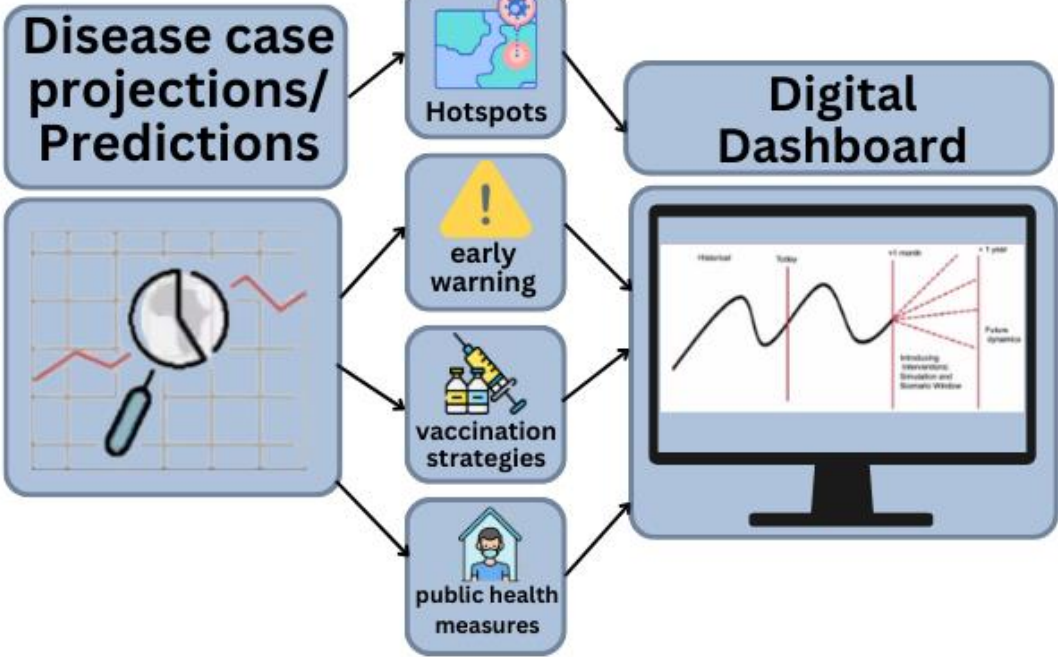
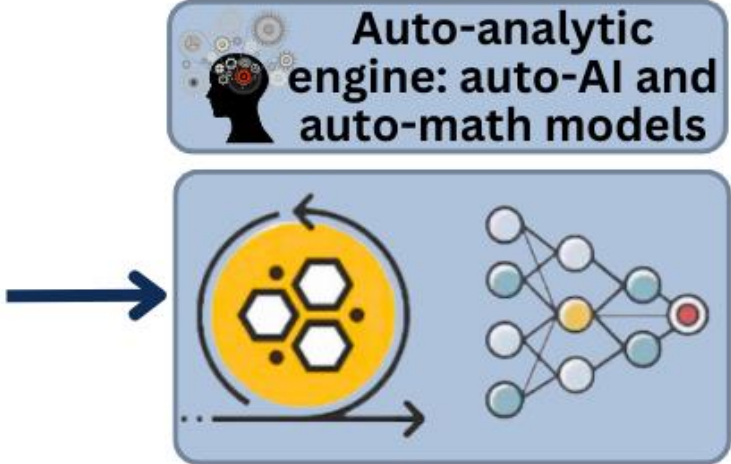
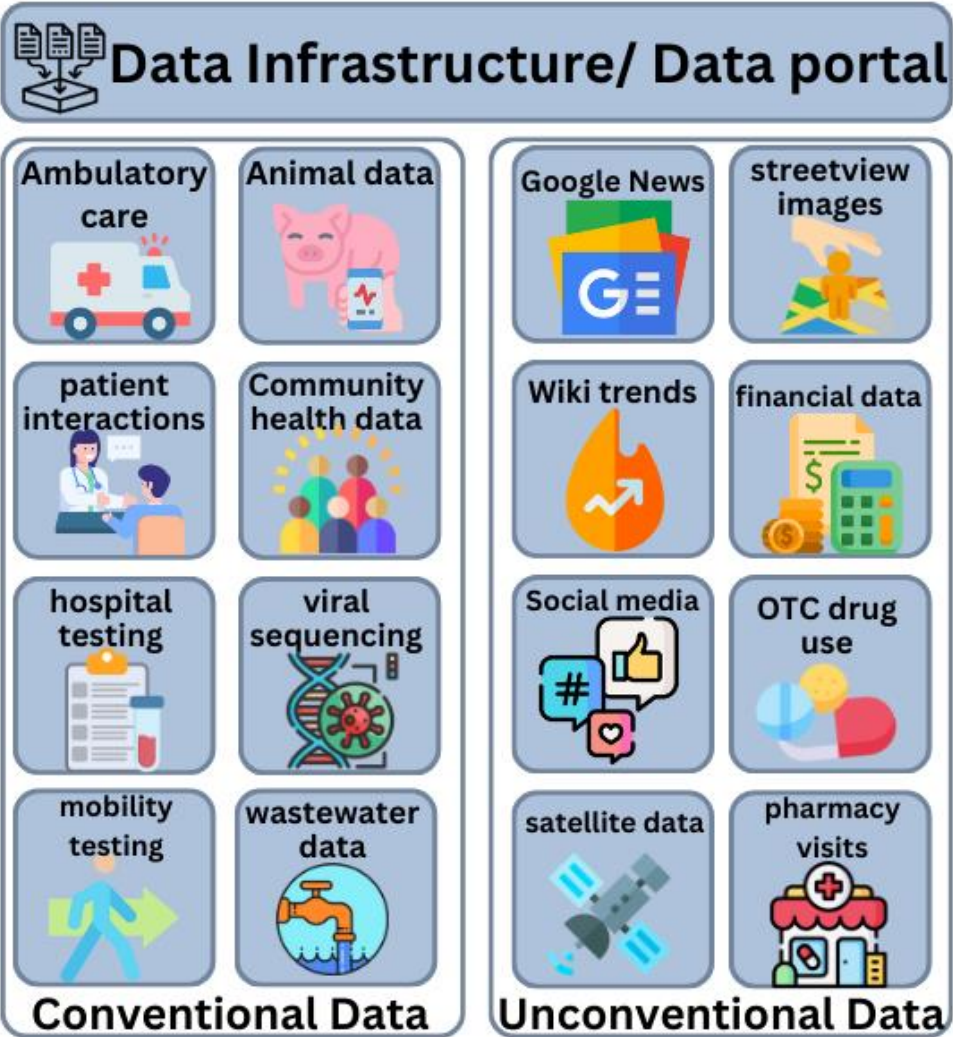
IDRC • CRDI

International Development Research Centre
Centre de recherches pour le développement international

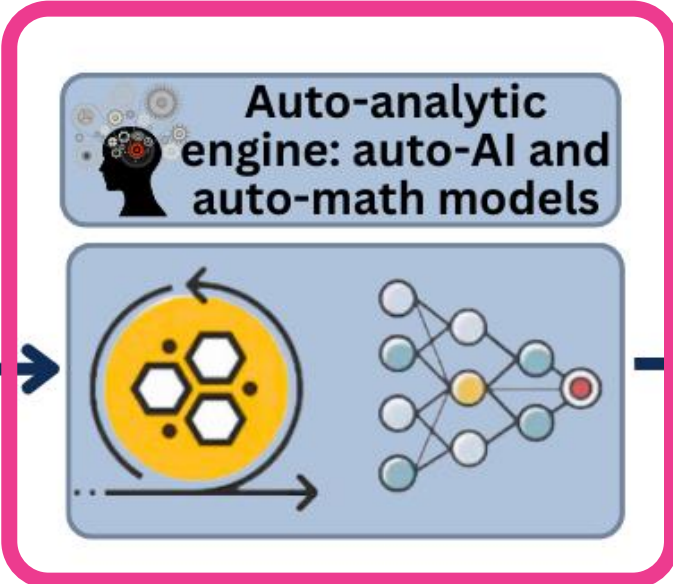
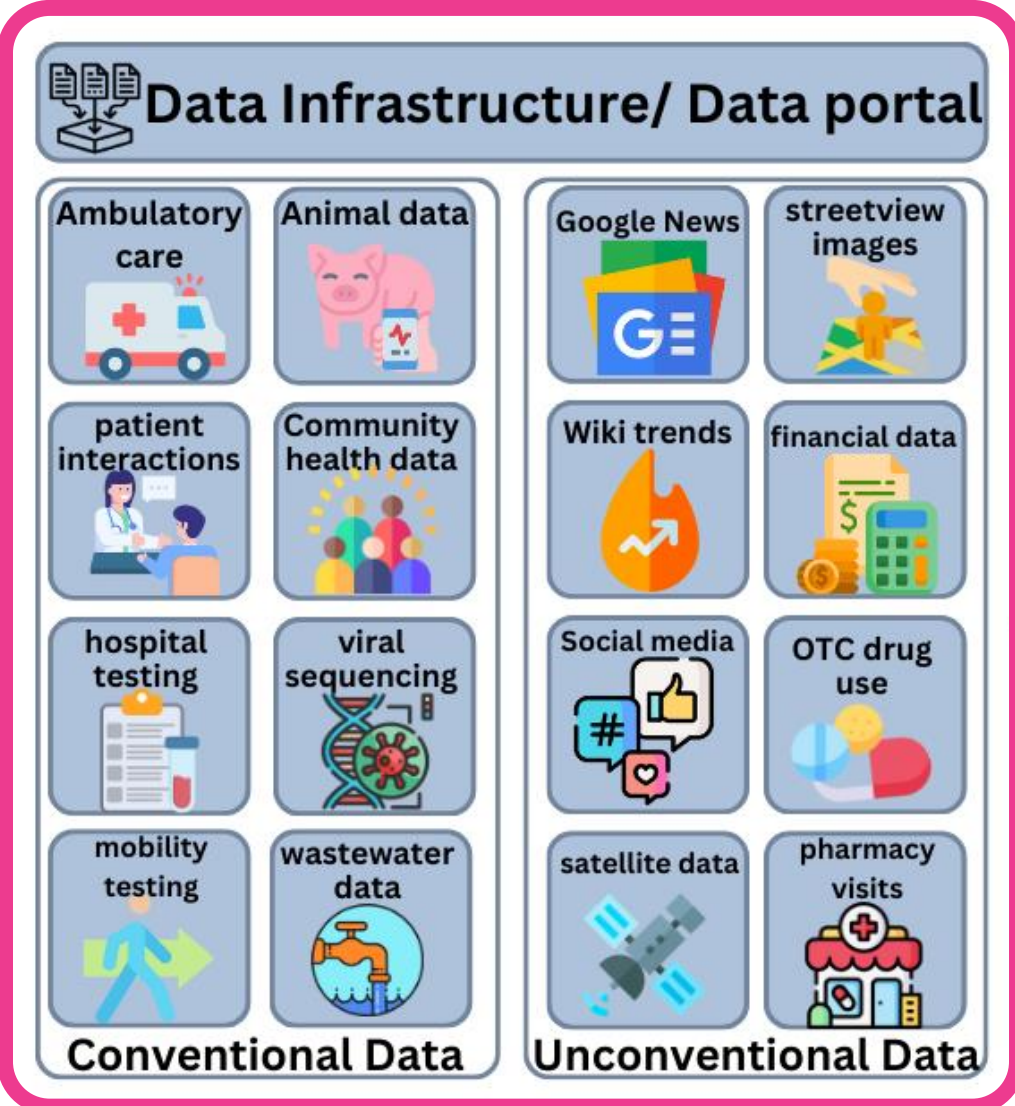
Canada 



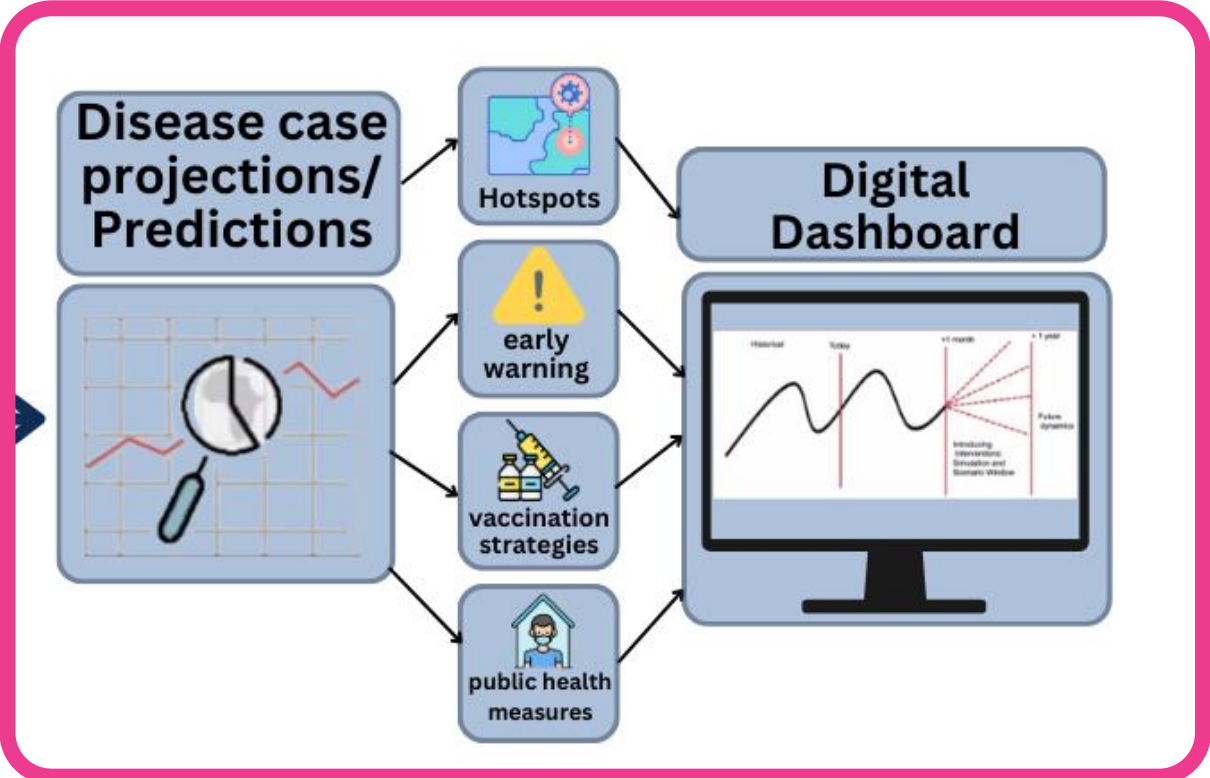
AutoAI Epidemix



AutoAI Epidemix



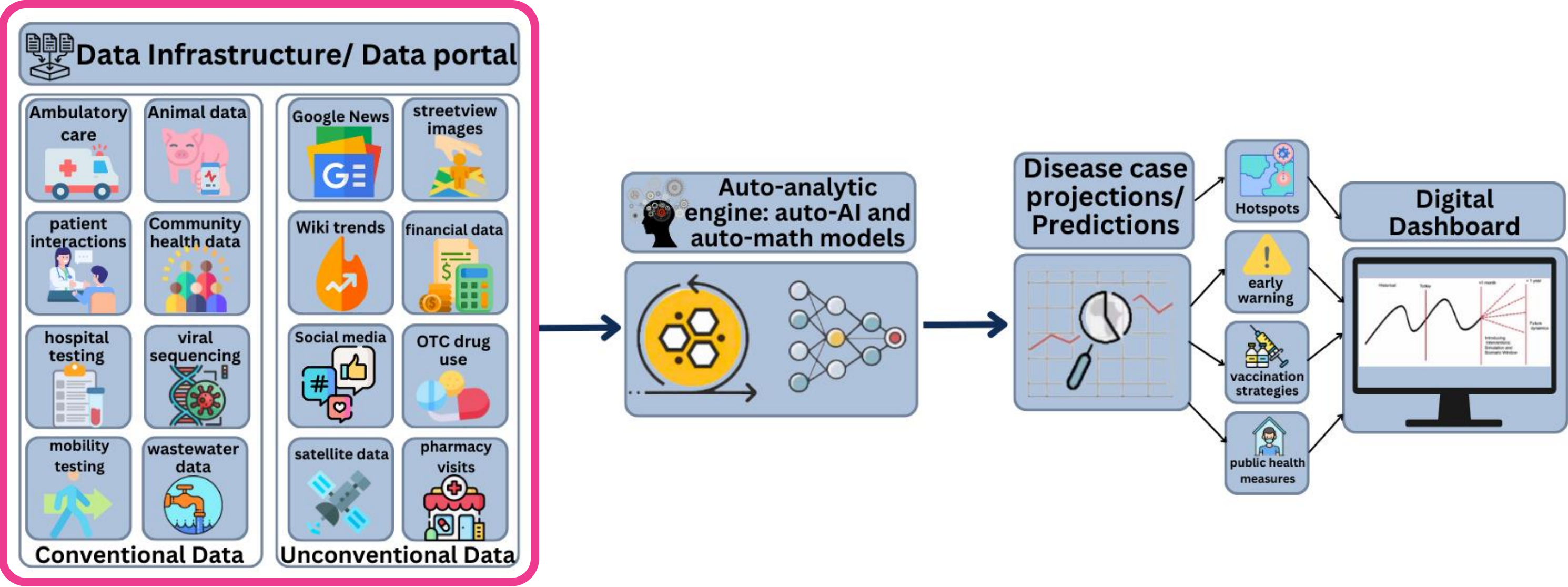
2^e partie :
Cadre du modèle



Conclusion :
Exemples (influenza, maladie de Lyme, COVID-19)

1^{re} partie de la conférence :
Sources de données et corrélation avec les maladies

AutoAI Epidemix

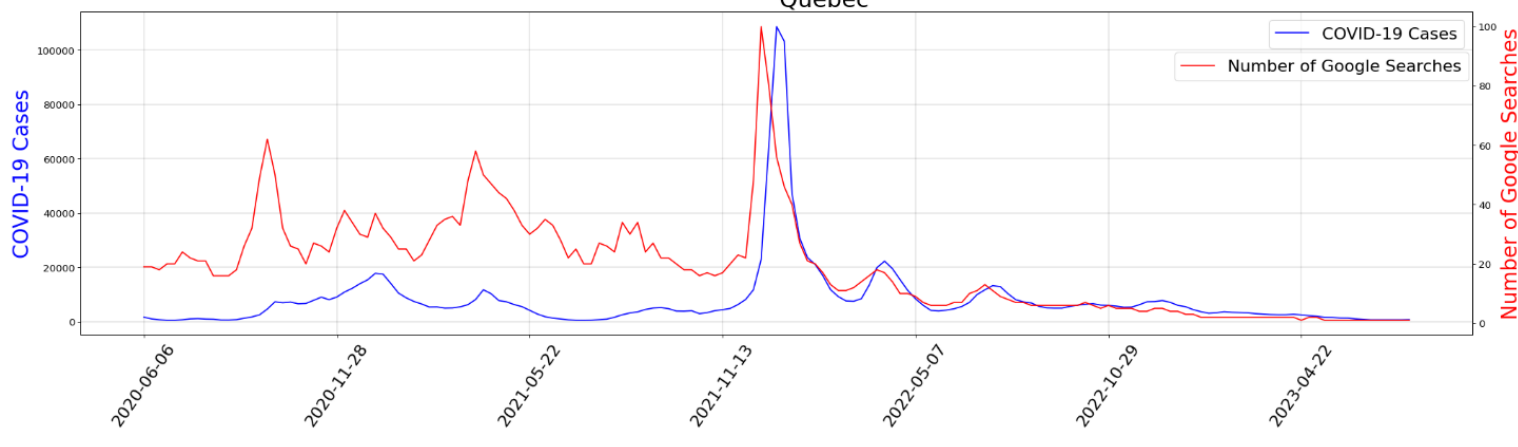
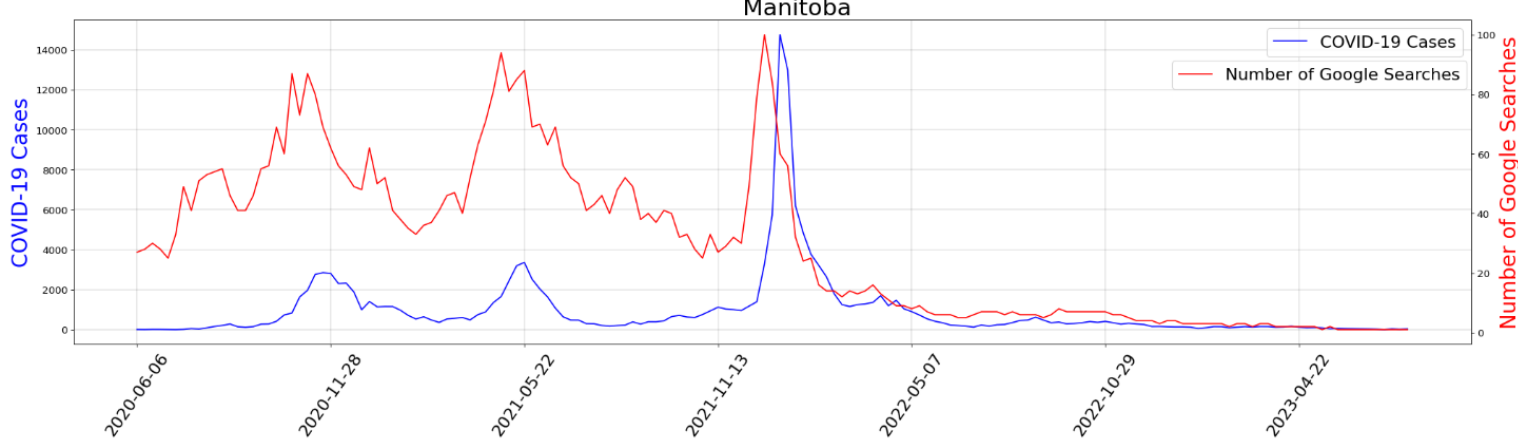
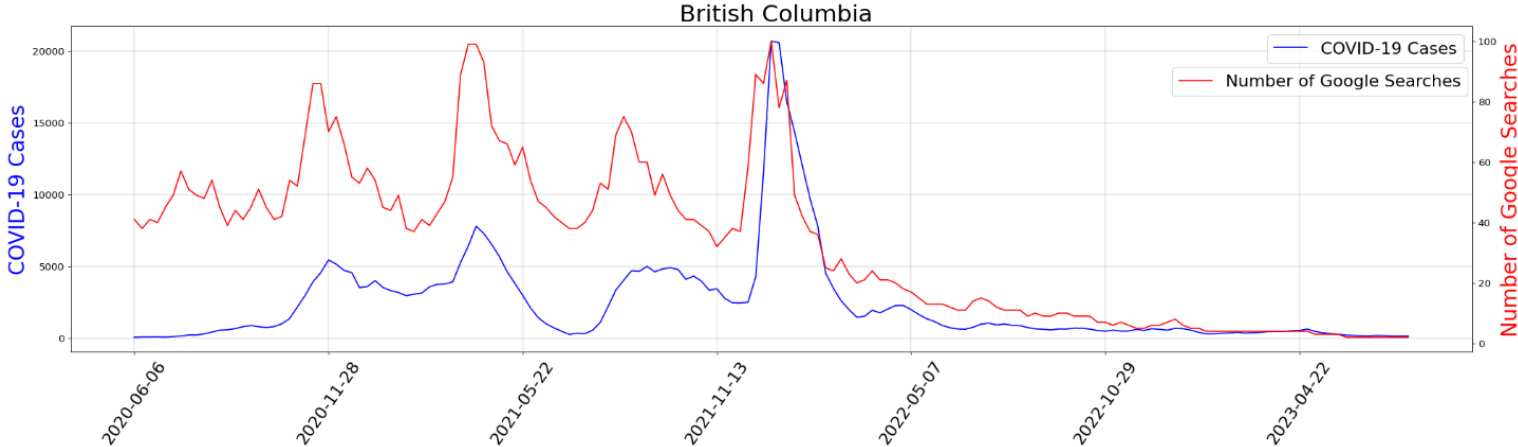
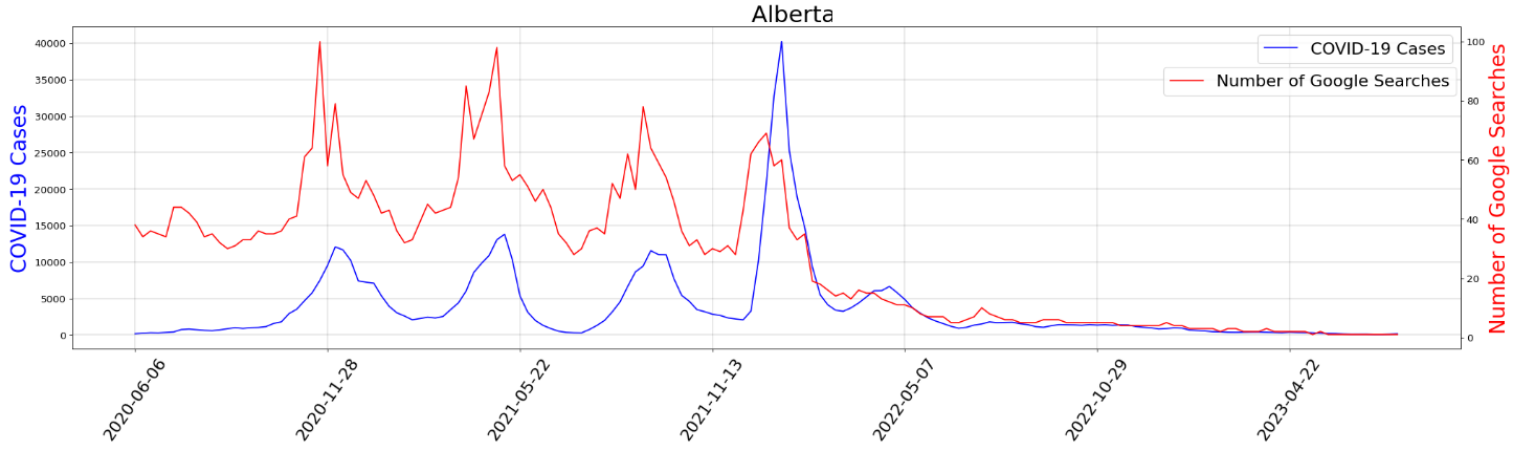
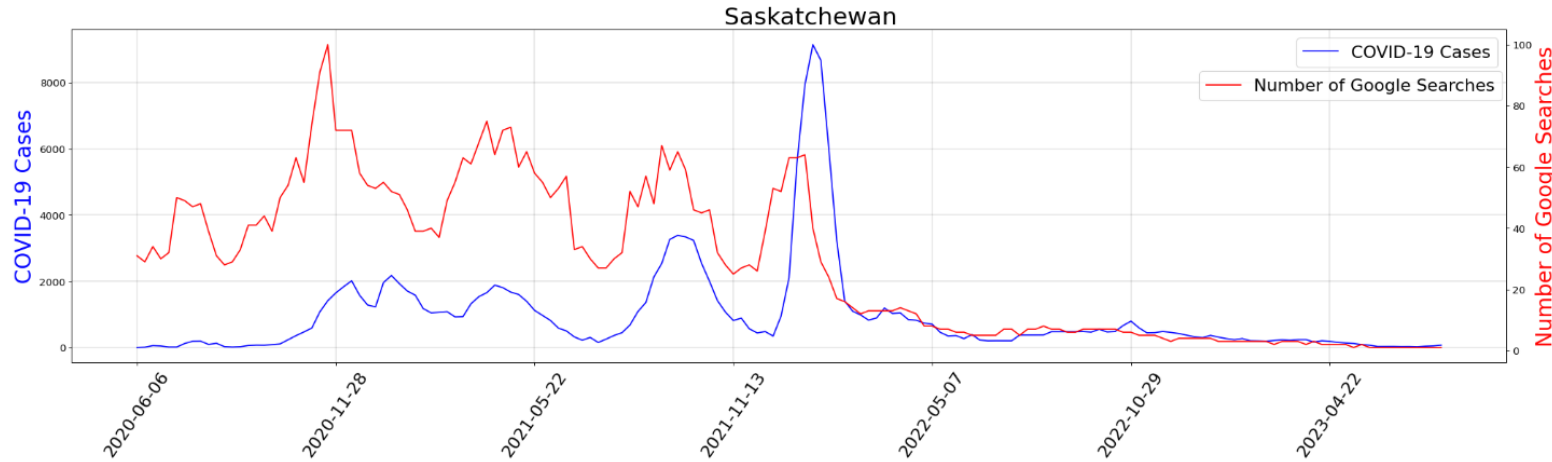


1^{re} partie de la conférence :
**Sources de données et
corrélation avec les maladies**

Google Trends



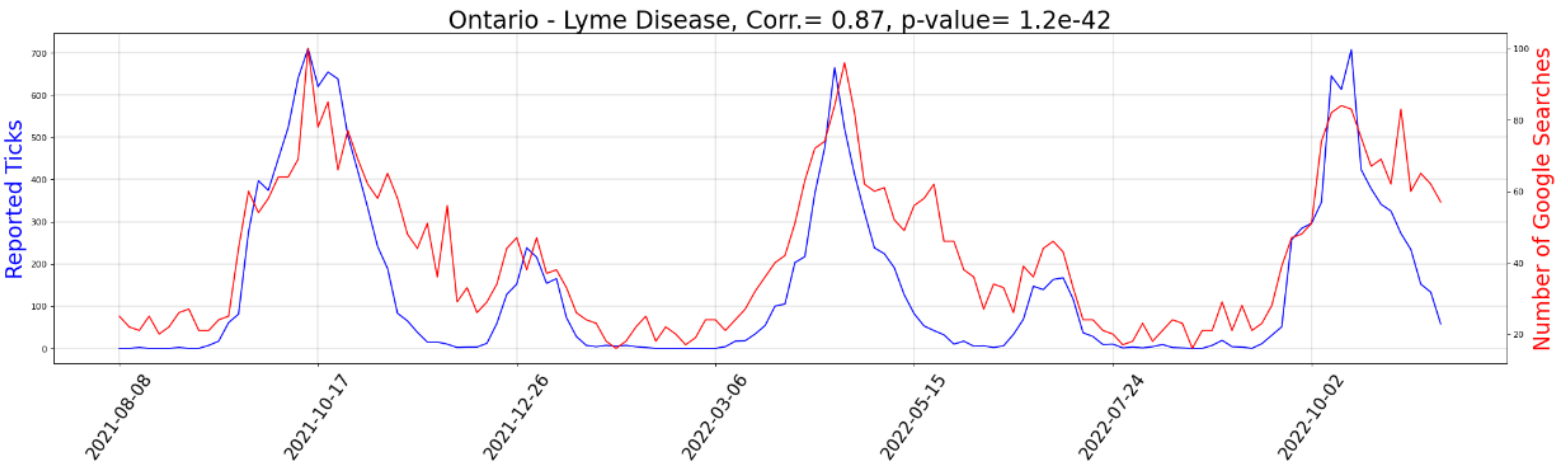
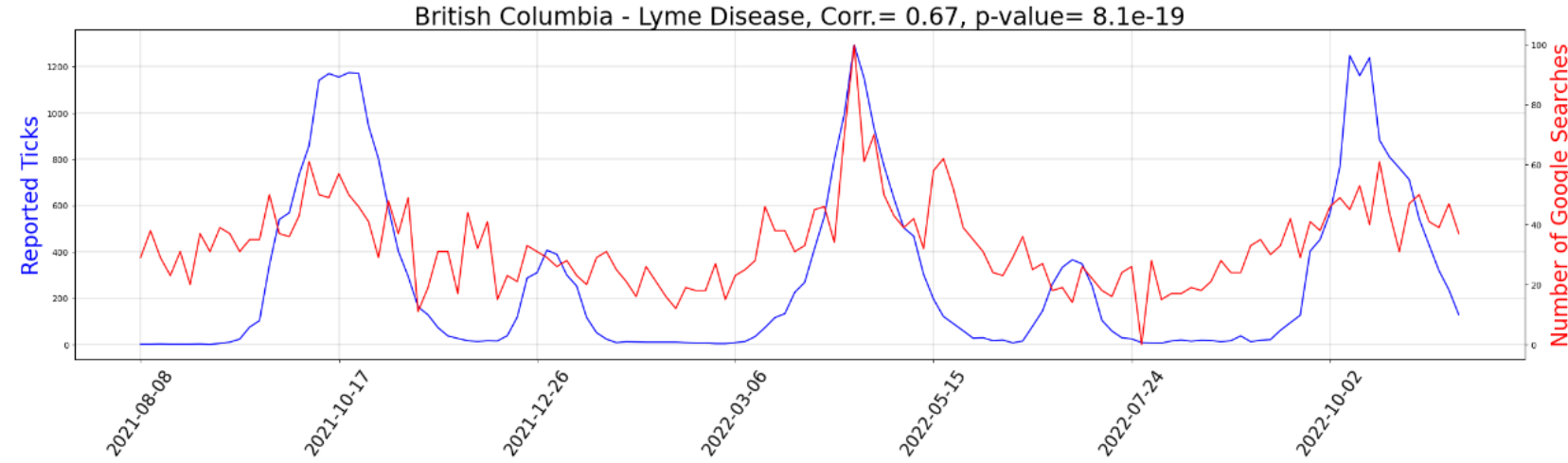
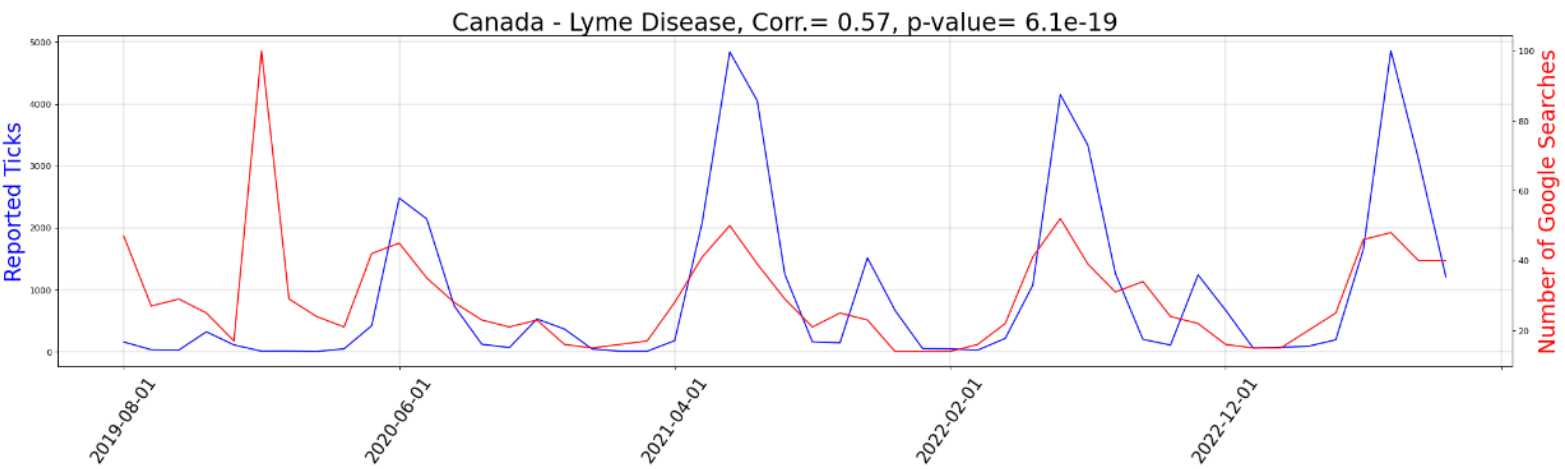
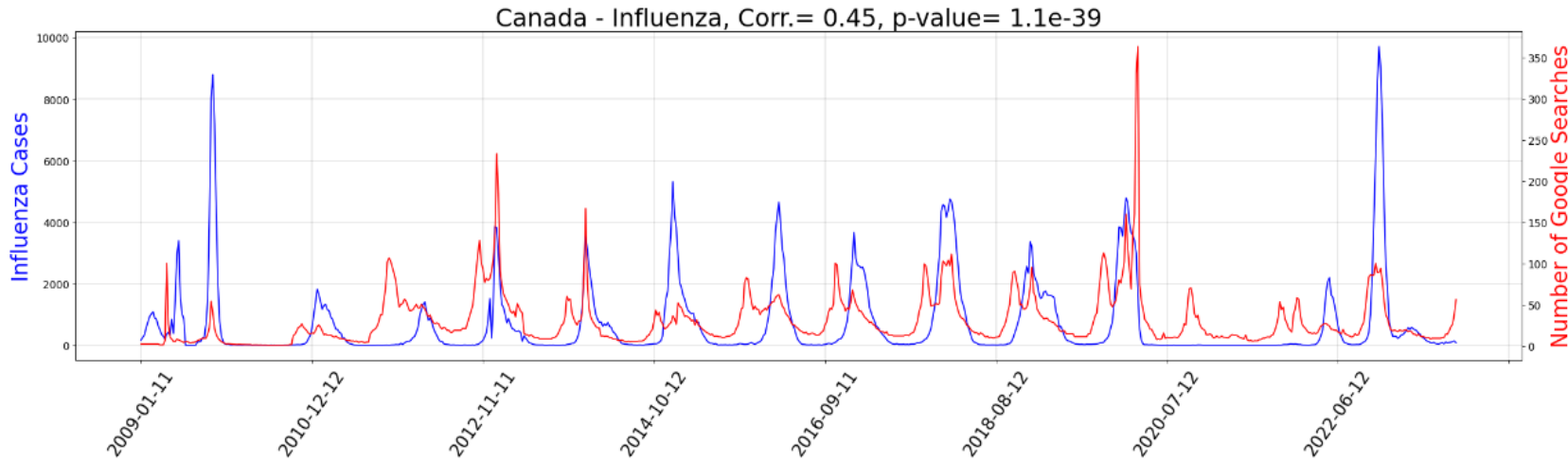
- Le nombre de fois où les personnes d'une province ou d'une communauté donnée ont effectué une recherche sur Google pour un sujet précis lié à une maladie constitue une excellente source de données.
- Dans la plupart des cas, il existe une forte corrélation entre les tendances de Google dans une région donnée et le nombre de cas de maladie dans cette région.
- Dans la plupart des cas, les tendances Google atteignent même un pic plus tôt que les cas réels de maladie.



Google Trends



- Le nombre de recherches Google sur une maladie donnée est presque toujours fortement corrélé avec le nombre de cas.



Google Trends



- Outre le nombre de recherches, le taux d'augmentation ou de diminution de ces dernières est un autre indicateur de l'augmentation ou de la diminution du nombre de cas.
- En outre, le taux de second ordre, ou la concavité et la convexité de la série chronologique, est un excellent indicateur de l'augmentation ou de la diminution du nombre de cas.

Intérêt au fil du temps 



Google Trends



- Outre le nombre de recherches, le taux d'augmentation ou de diminution de ces dernières est un autre indicateur de l'augmentation ou de la diminution du nombre de cas.
- En outre, le taux de second ordre, ou la concavité et la convexité de la série chronologique, est un excellent indicateur de l'augmentation ou de la diminution du nombre de cas.

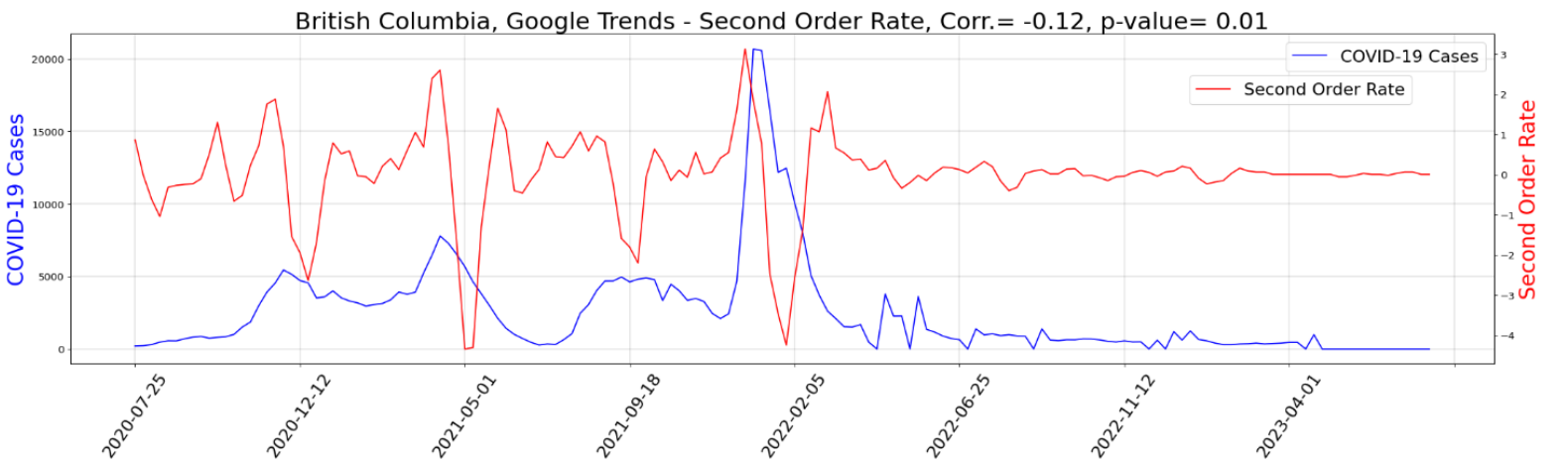
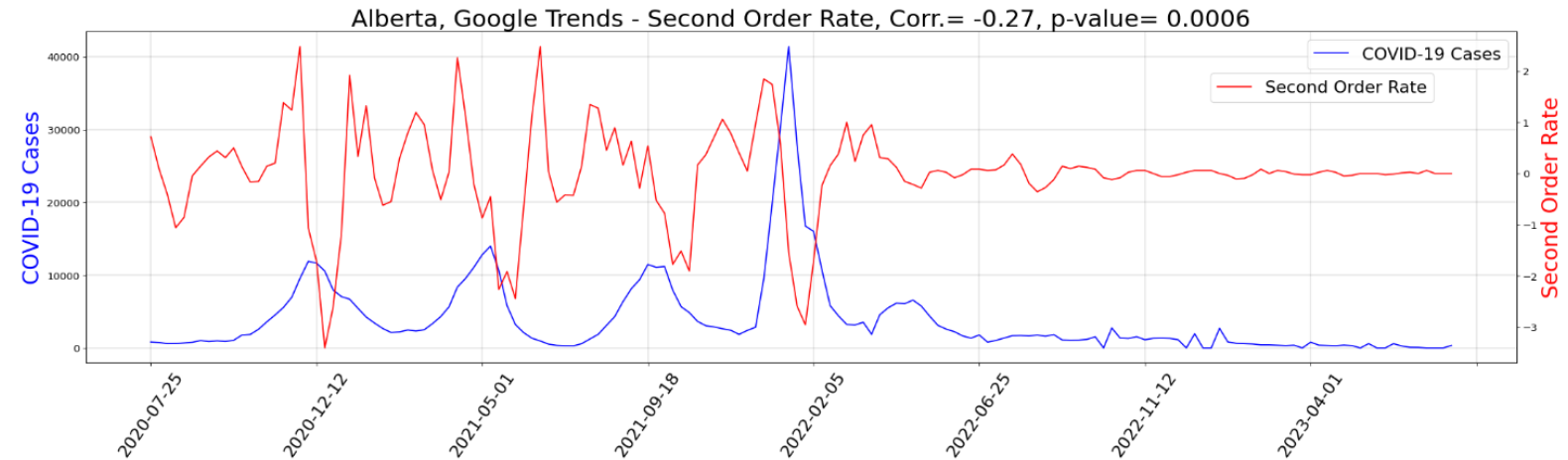
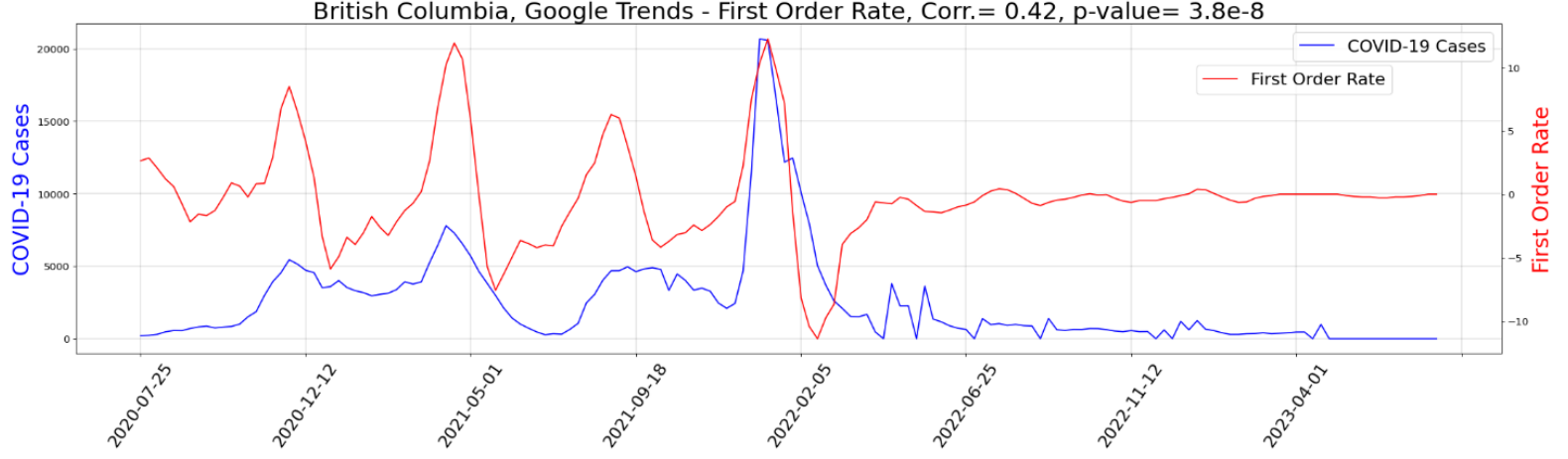
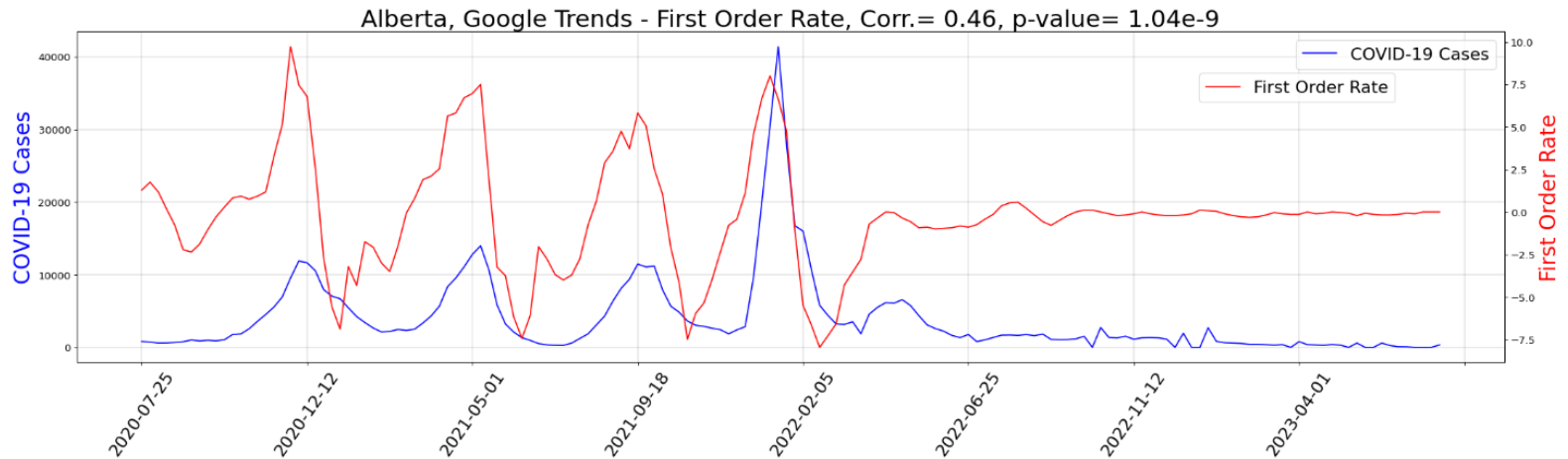
Intérêt au fil du temps 



Google Trends



- Les taux de premier et de second ordre des tendances Google sont également corrélés avec le nombre de cas.



Médias sociaux



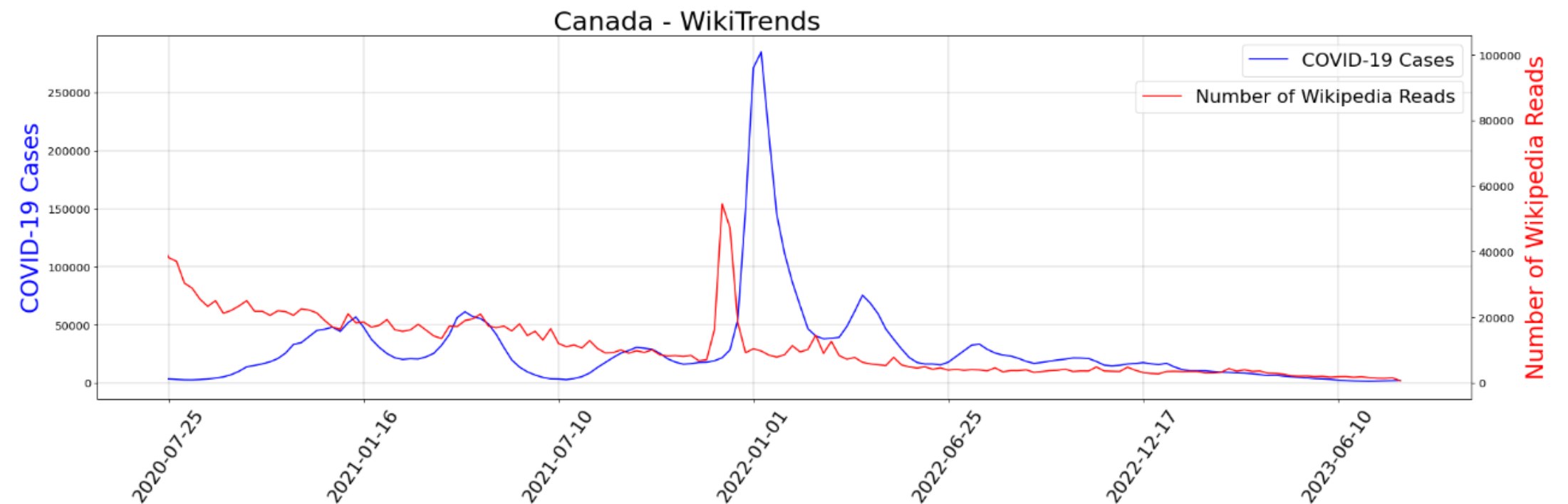
- Les gens utilisent de plus en plus les médias sociaux pour partager leurs opinions et leurs expériences.
- À mesure que le nombre de cas augmente, les gens discutent davantage de l'épidémie sur les médias sociaux. Par conséquent, les plateformes de médias sociaux telles que Reddit et X sont également une excellente source pour prédire les épidémies.



Wiki Trends



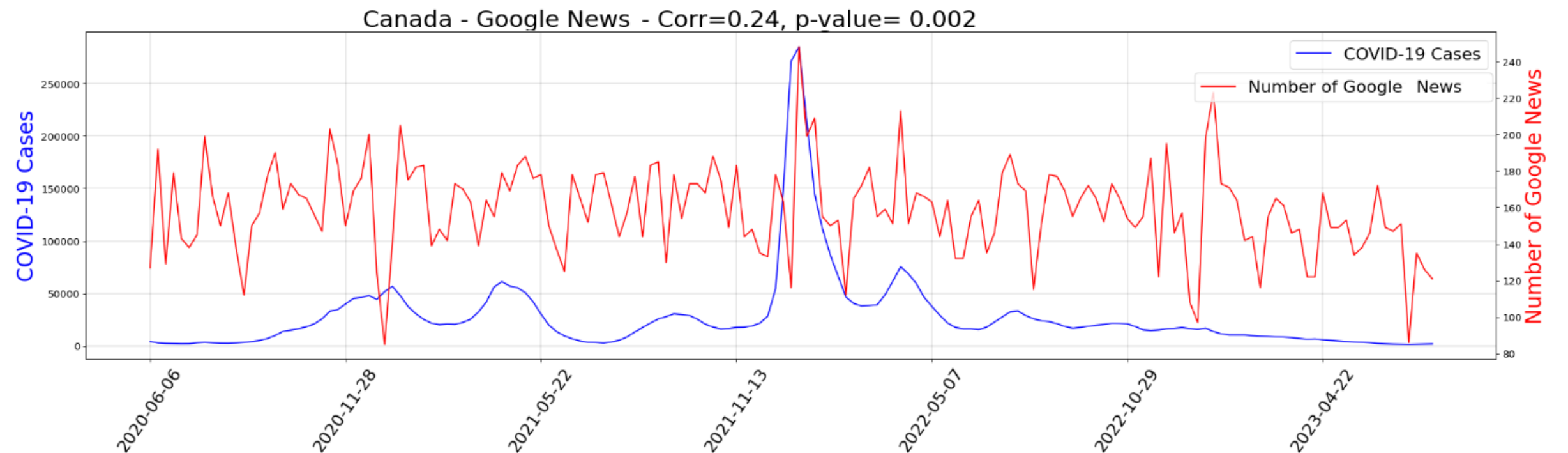
- Pour certaines maladies (p. ex. la COVID-19), une page Wikipédia distincte est disponible pour les différents pays.
- Le nombre de consultations d'une page précise qui a pu être récupéré à l'aide de l'API de Wiki Trends est également un indicateur du nombre de cas.
- Pour la plupart des pays, Wiki Trends est également bien corrélé avec le nombre de cas.
- Parfois, Wiki Trends atteint le pic plus tôt que le nombre de cas.



Actualités



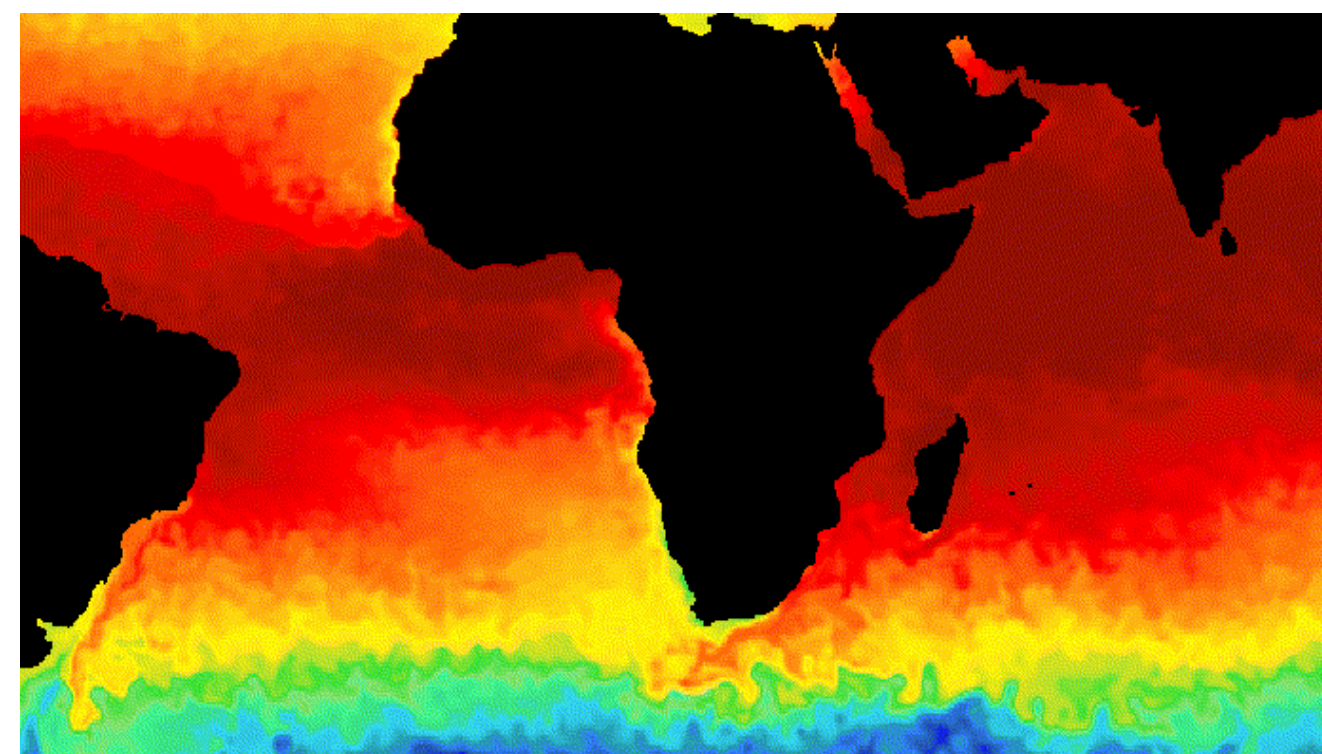
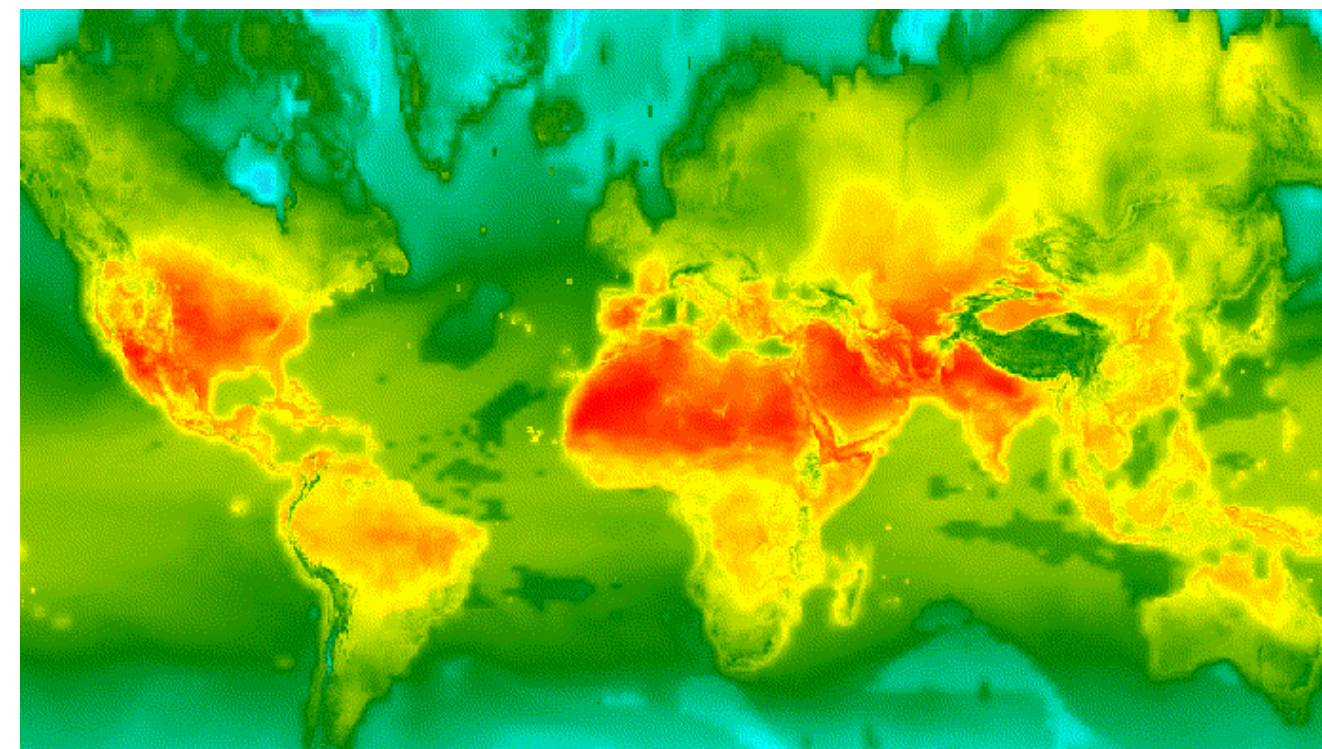
- Le volume de nouvelles publiées sur une maladie particulière pourrait également être utilisé comme indicateur du nombre de cas.
- Le nombre d'actualités Google publiées dans un pays donné pour un certain mot-clé, qui ont pu être récupérées à l'aide de l'API Google Actualités, est généralement bien corrélé avec le nombre de cas.



Données satellites



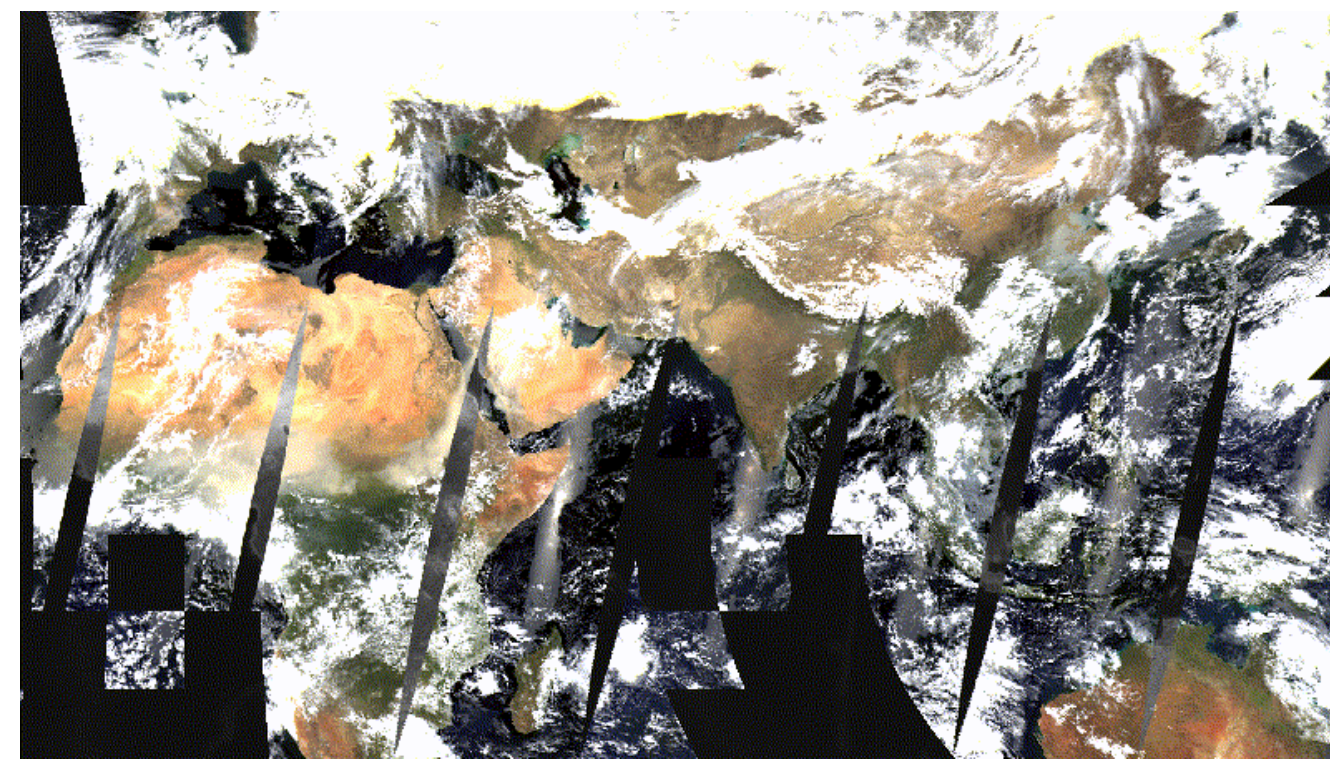
- De nombreux paramètres ont pu être obtenus à l'aide de données satellites, notamment (sans toutefois s'y limiter) :
 - **Climat** : des facteurs tels que la température annuelle, les précipitations annuelles, l'isothermie, l'amplitude diurne de la température, etc. peuvent être obtenus à l'aide de données climatiques.
 - **Température de surface** : les capteurs fournissent des renseignements sur la température de surface et l'émissivité. Les maladies à transmission vectorielle pourraient être bien corrélées avec les données relatives à la température de surface.



Données satellites



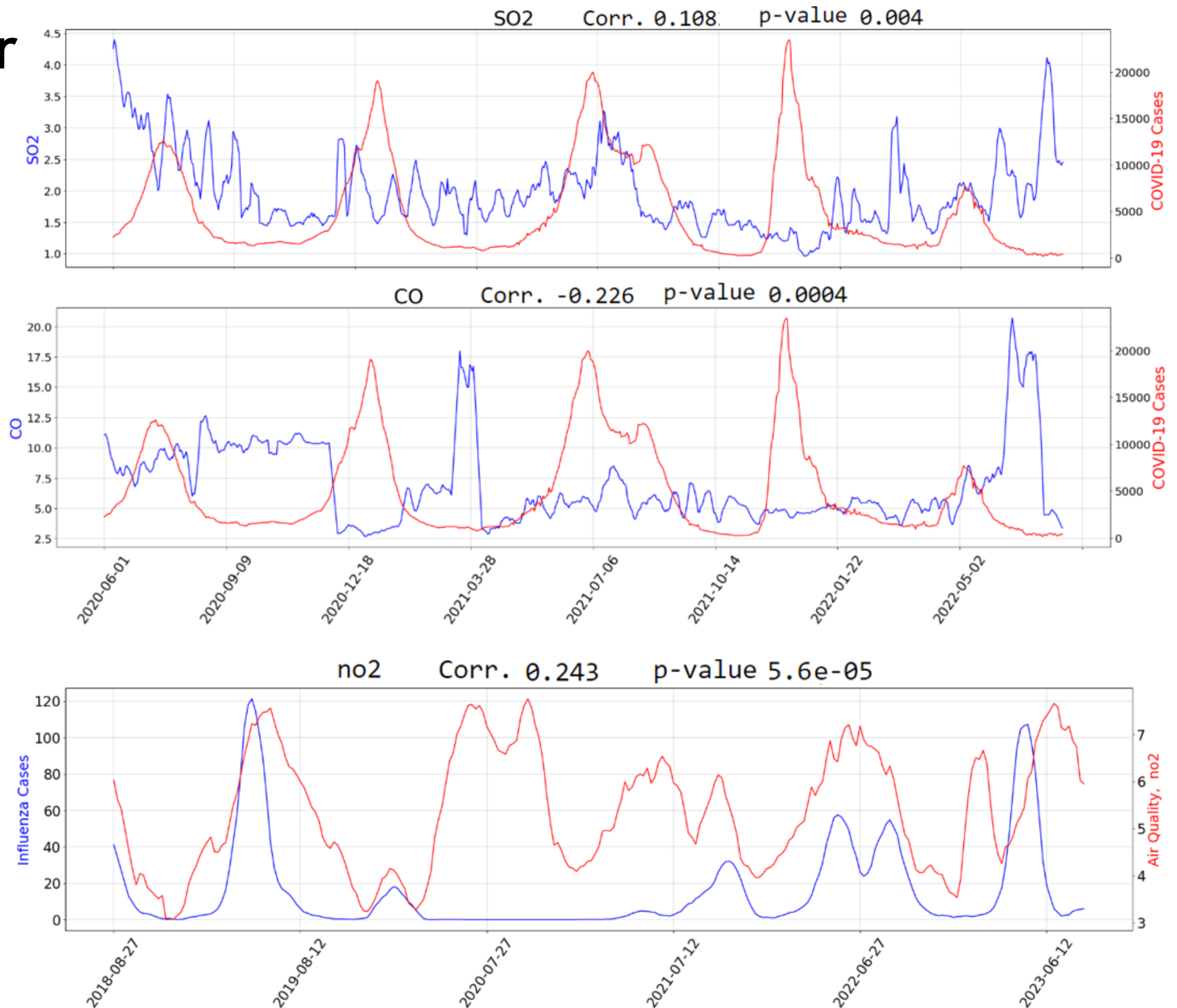
- De nombreux paramètres ont pu être obtenus à l'aide de données satellites, notamment (sans toutefois s'y limiter) :
- **Données météorologiques** : des facteurs tels que la température, les précipitations, les chutes de neige, la direction du vent, les rafales, l'humidité, etc. peuvent être obtenus à l'aide des données météorologiques.
- **Données sur la qualité de l'air** : des facteurs tels que la concentration de NO_2 , de CO , de SO_2 , d' O_3 et de CH_4 peuvent être obtenus grâce à la mission SENTINEL-5, qui fait partie du programme européen d'observation de la Terre.



Données sur la qualité de l'air



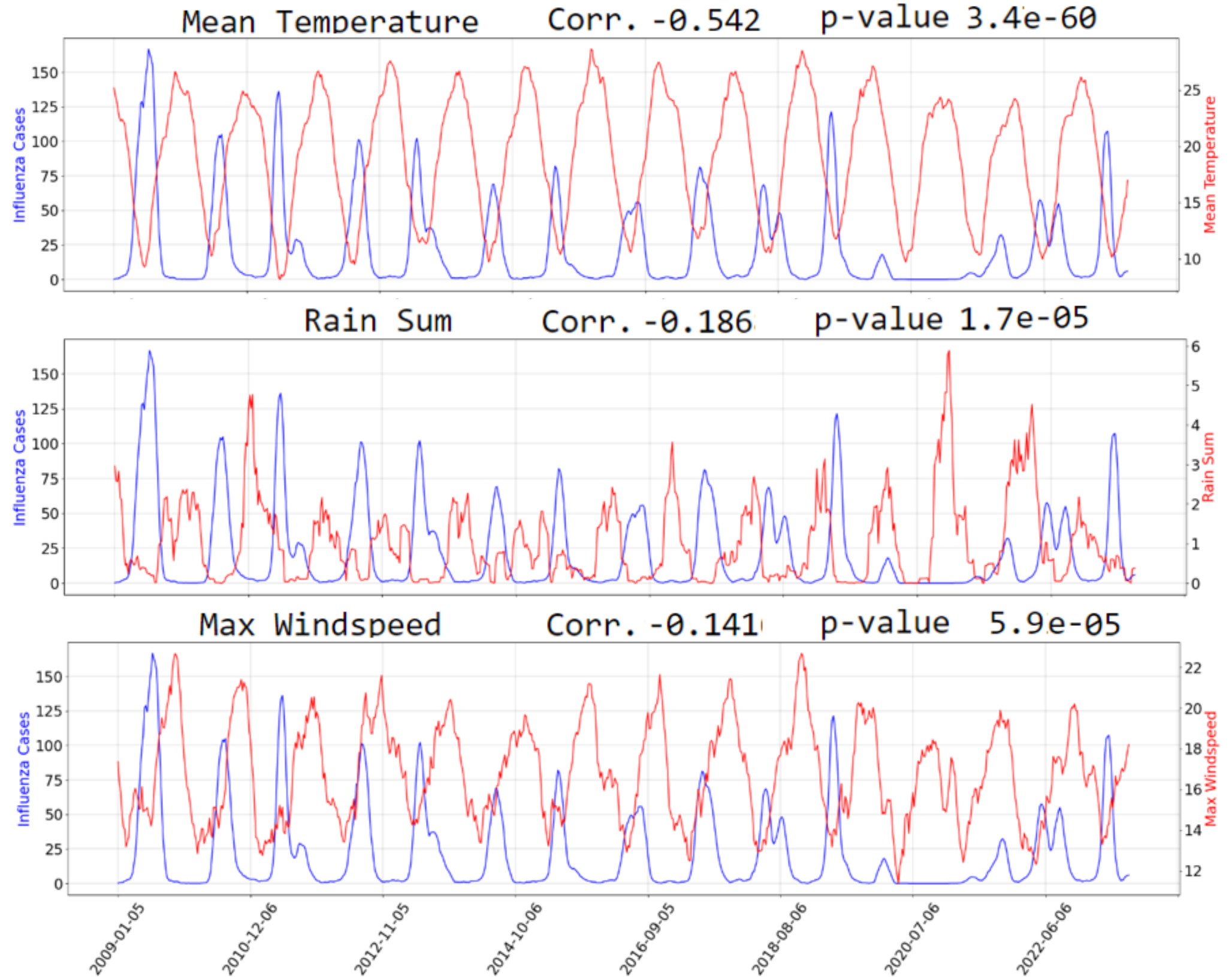
- Les paramètres de la qualité de l'air comprennent la concentration d'éléments tels que le CO, le CO₂, le NO₂, l'O₃ et le CH₄.
- Les paramètres de la qualité de l'air sont en corrélation avec les maladies respiratoires telles que la COVID-19 et l'influenza.



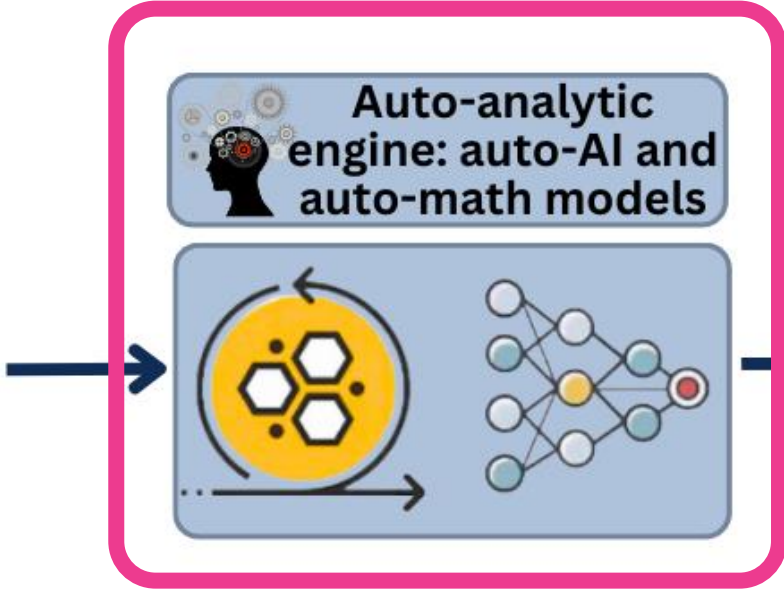
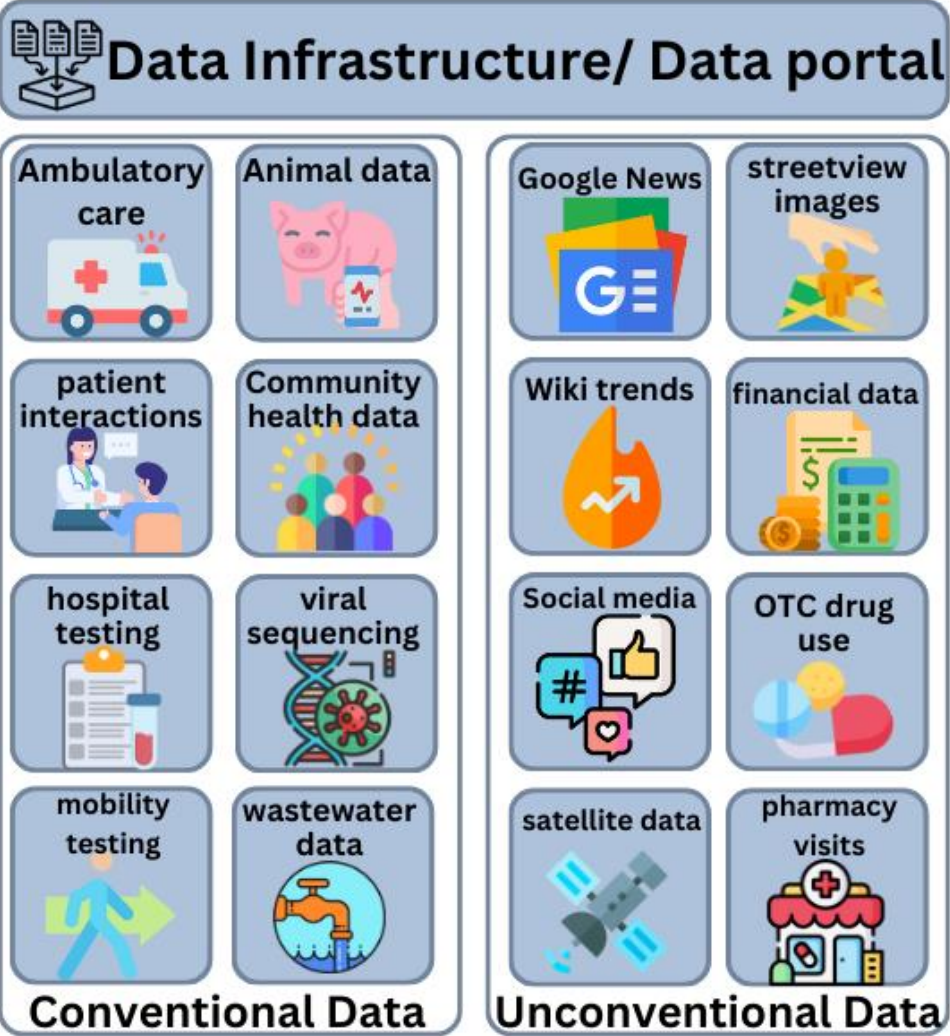
Données météorologiques



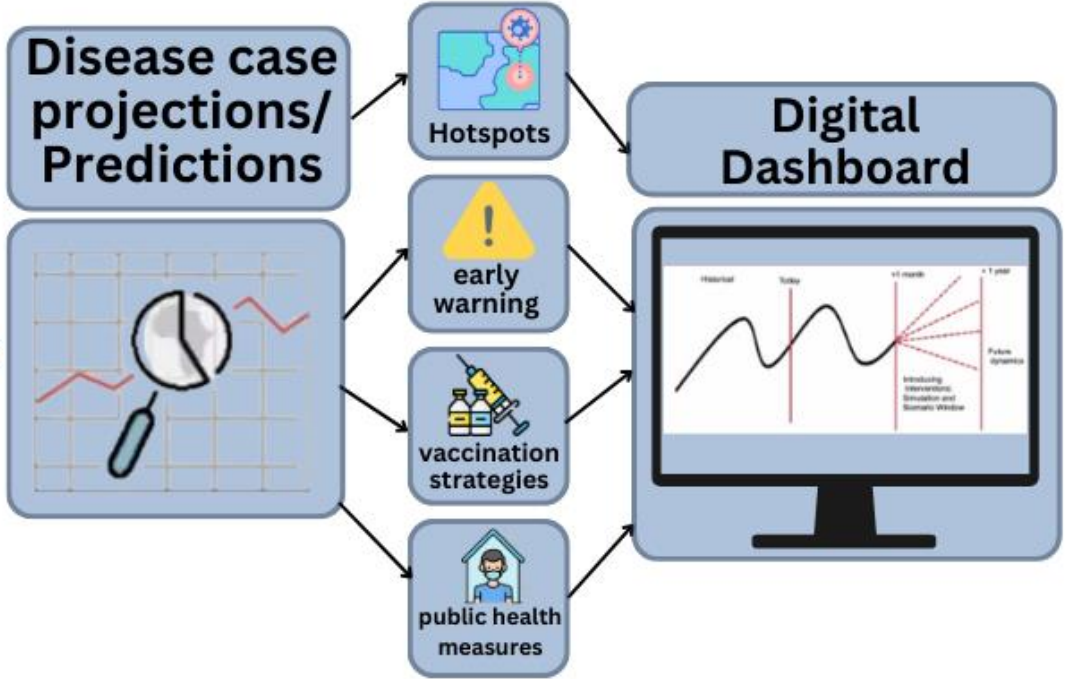
- Les données météorologiques comprennent des paramètres tels que la température, l'humidité, les précipitations, la vitesse du vent, les rafales, etc.
- Les données météorologiques sont en corrélation avec les cas d'influenza.



AutoAI Epidemix



2^e partie :
Cadre du modèle



Données sur la santé communautaire



Organigramme



L'organigramme comprend trois blocs :

1. Ingestion des données :

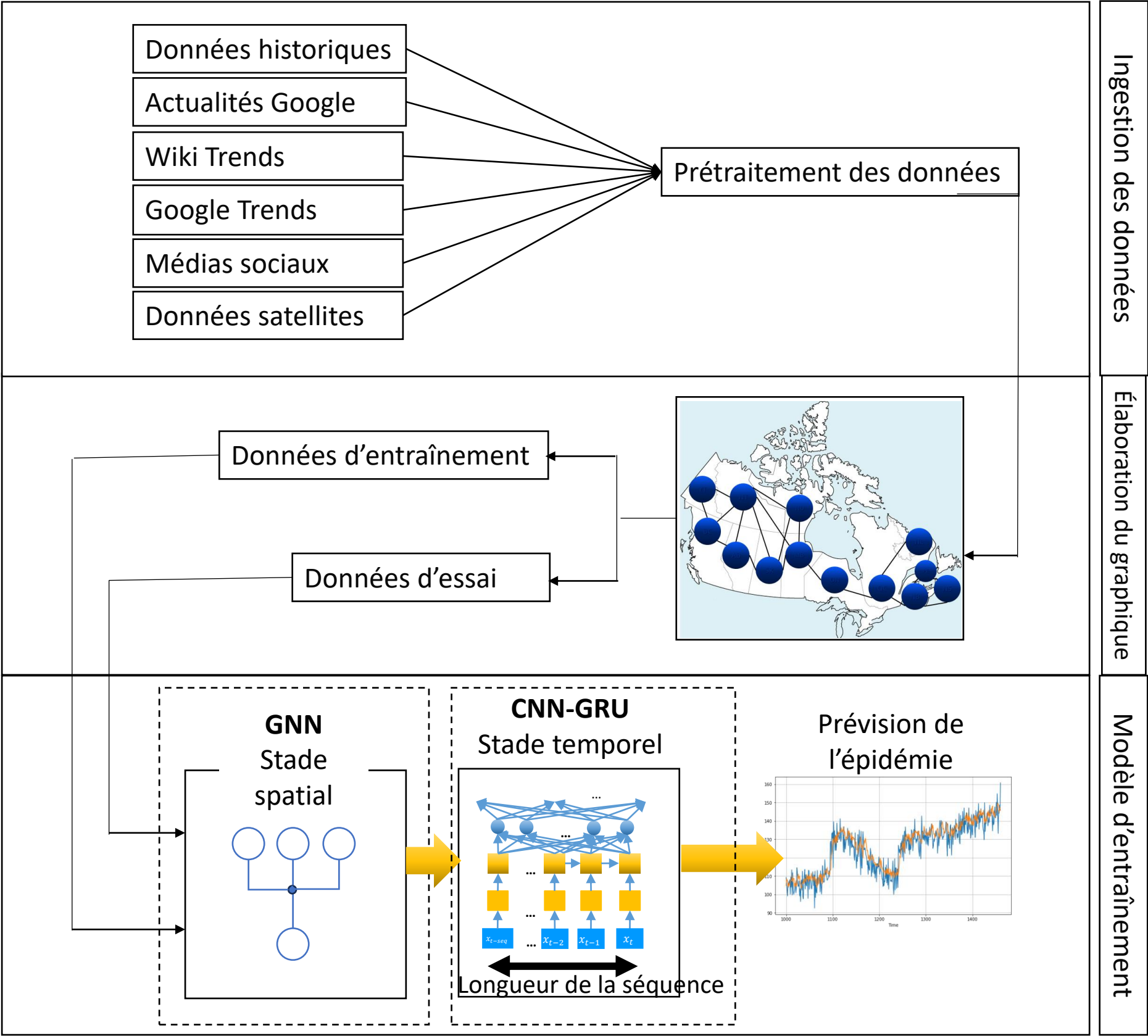
- Pour chaque province, plusieurs sources de données sont traitées comme des séries temporelles et stockées dans des tenseurs.
- Le prétraitement consiste à remplacer les données manquantes par zéro, à spécifier les caractéristiques et les étiquettes, puis à centrer et à mettre à l'échelle les valeurs.

2. Élaboration du graphique :

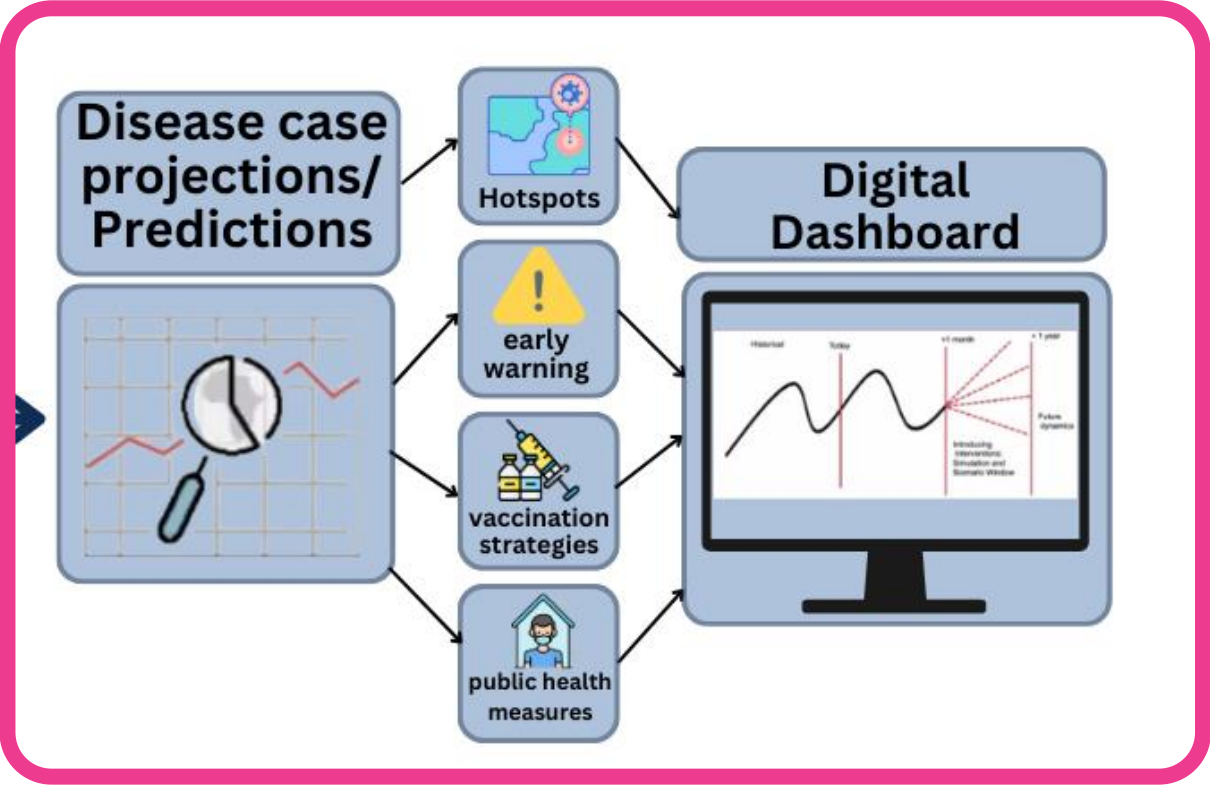
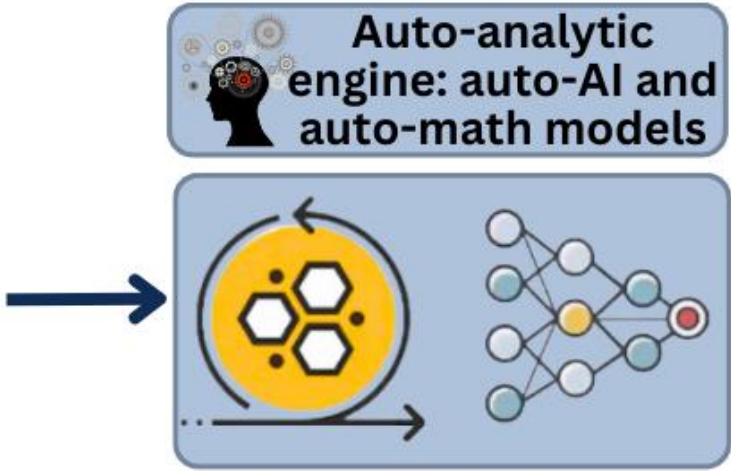
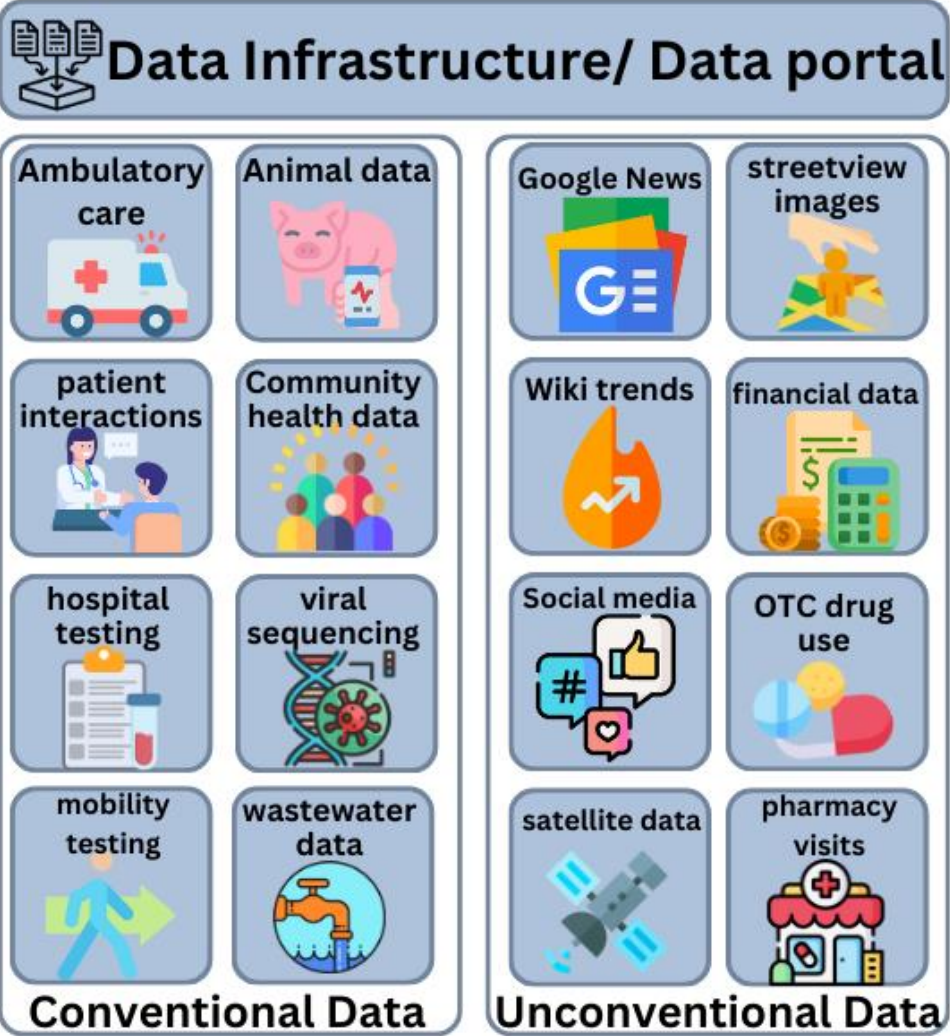
- Les nœuds et leurs données, les arêtes ainsi que les poids sont définis.
- La série chronologique d'entraînement et d'essai de chaque nœud est préparée.

3. Modèle d'entraînement :

- Dans chaque nœud, les données des voisins sont combinées et le résultat passe par un modèle RNN, qui est le CNN-GRU.
- Le modèle est entraîné, puis mis à l'essai.



AutoAI Epidemix

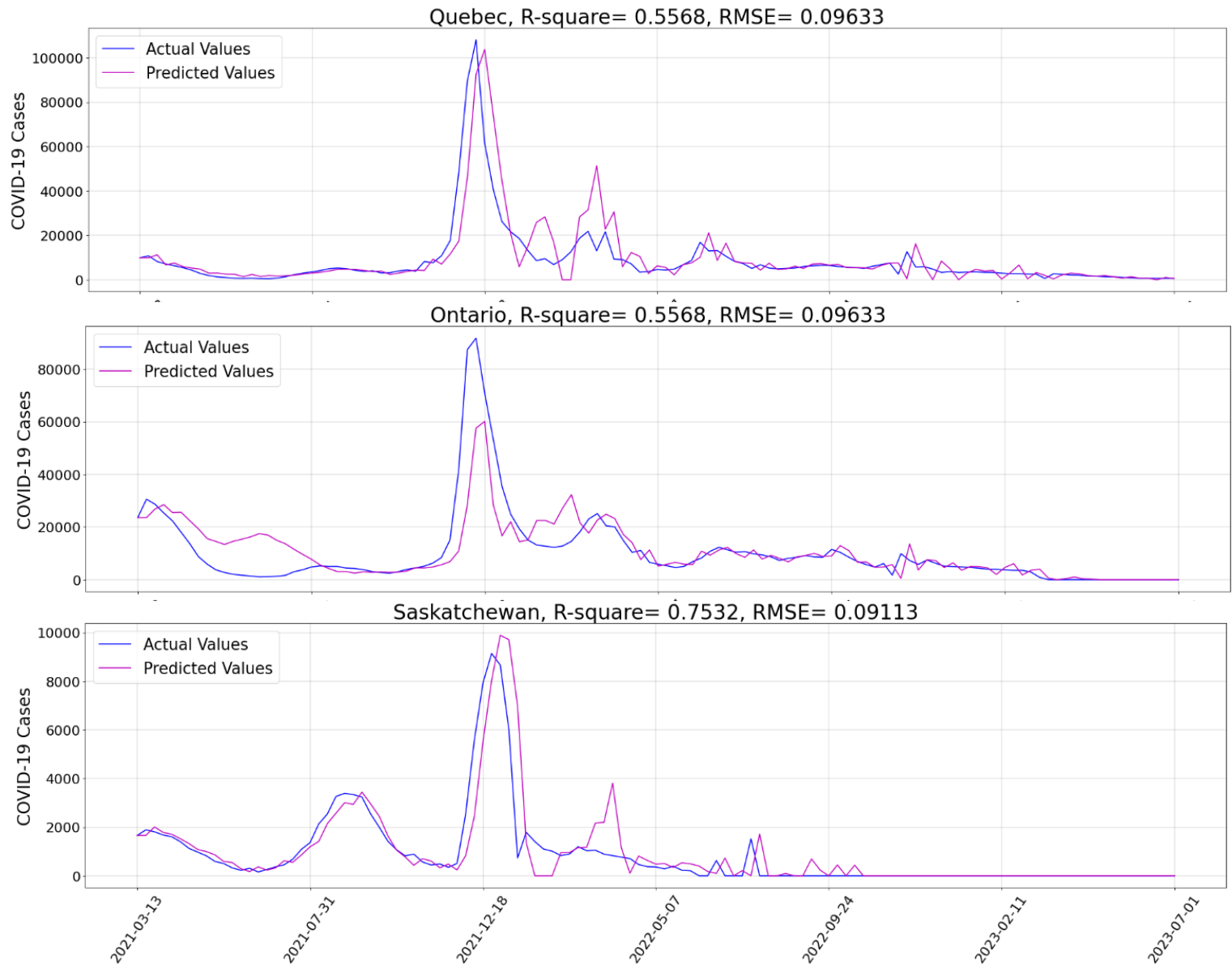
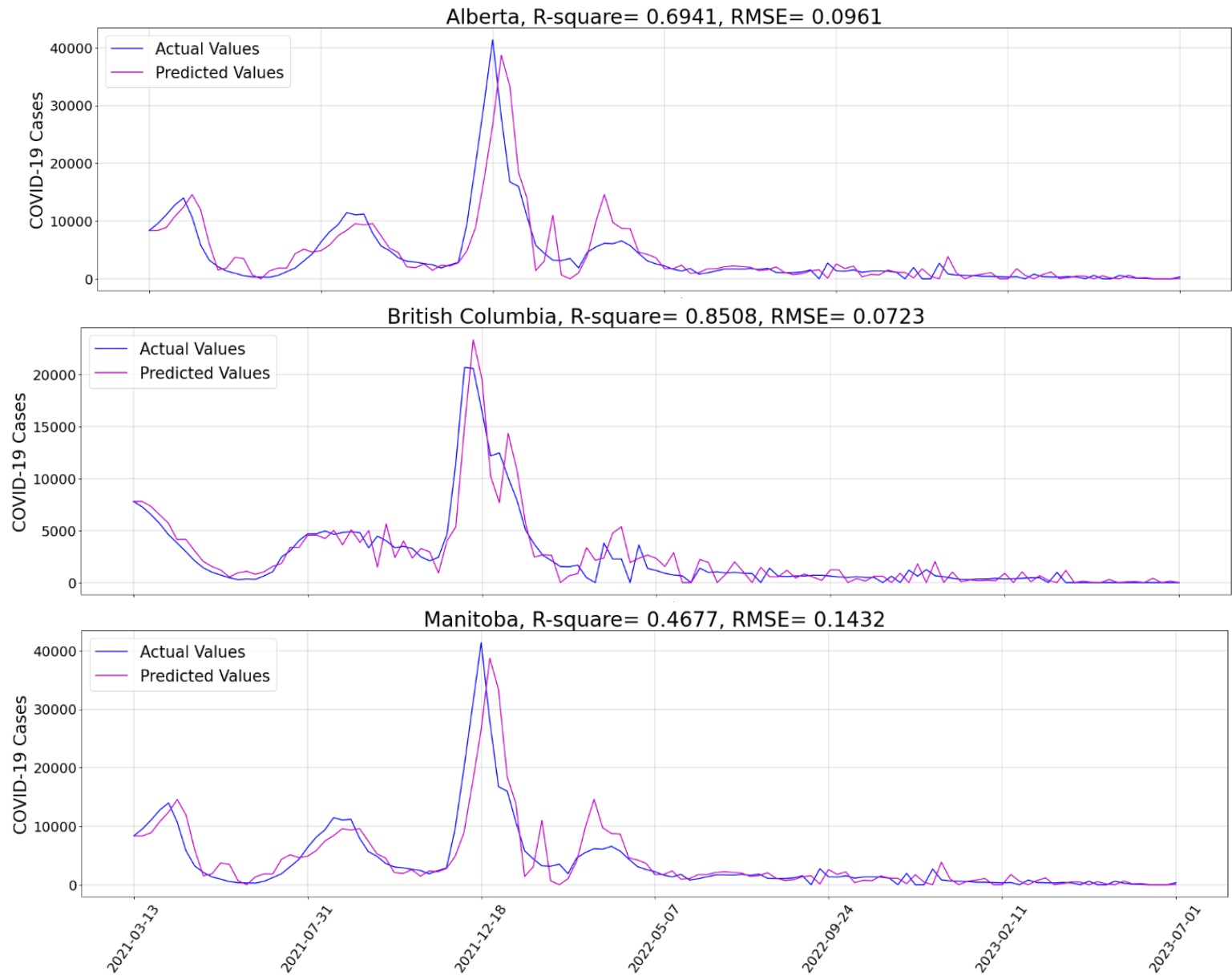


Conclusion :
Exemples (influenza, maladie de Lyme, COVID-19)

Canada : Prédiction 14 étapes à l'avance



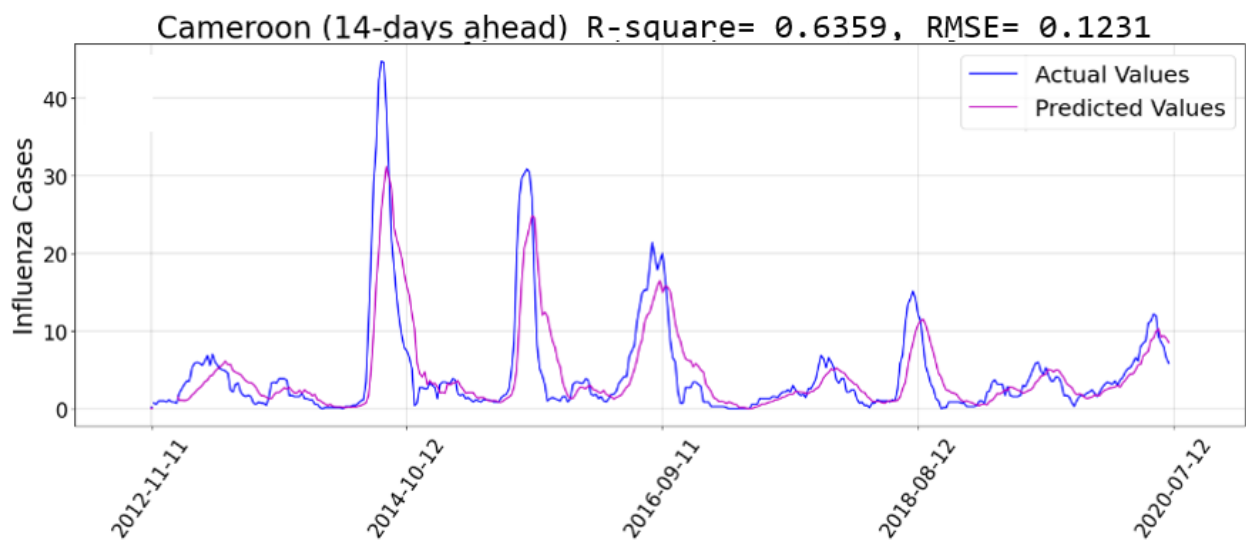
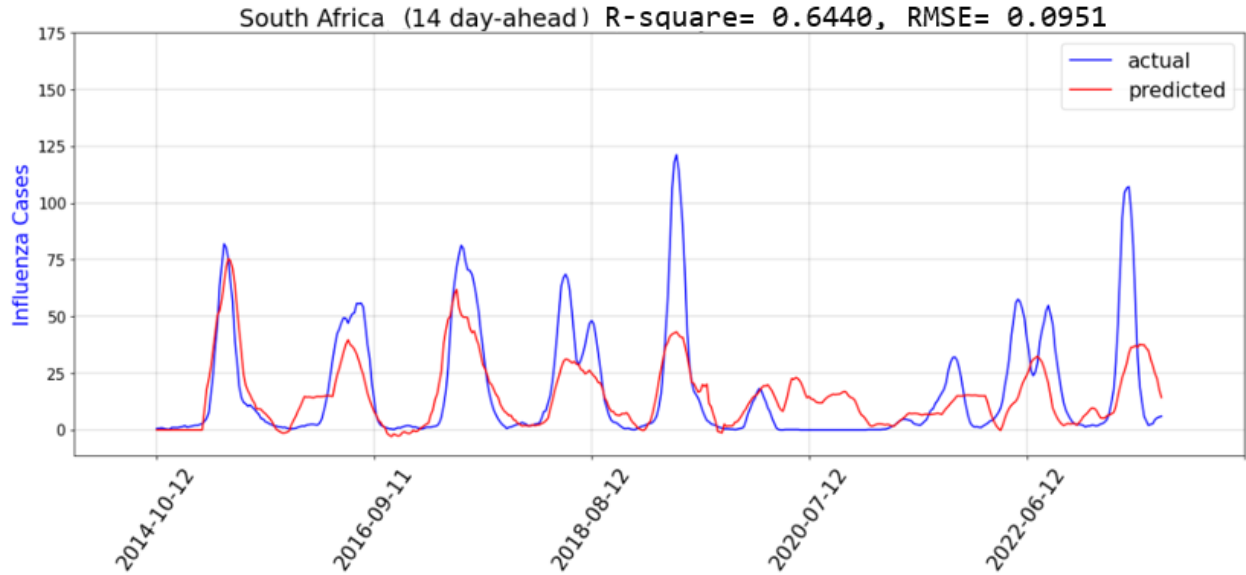
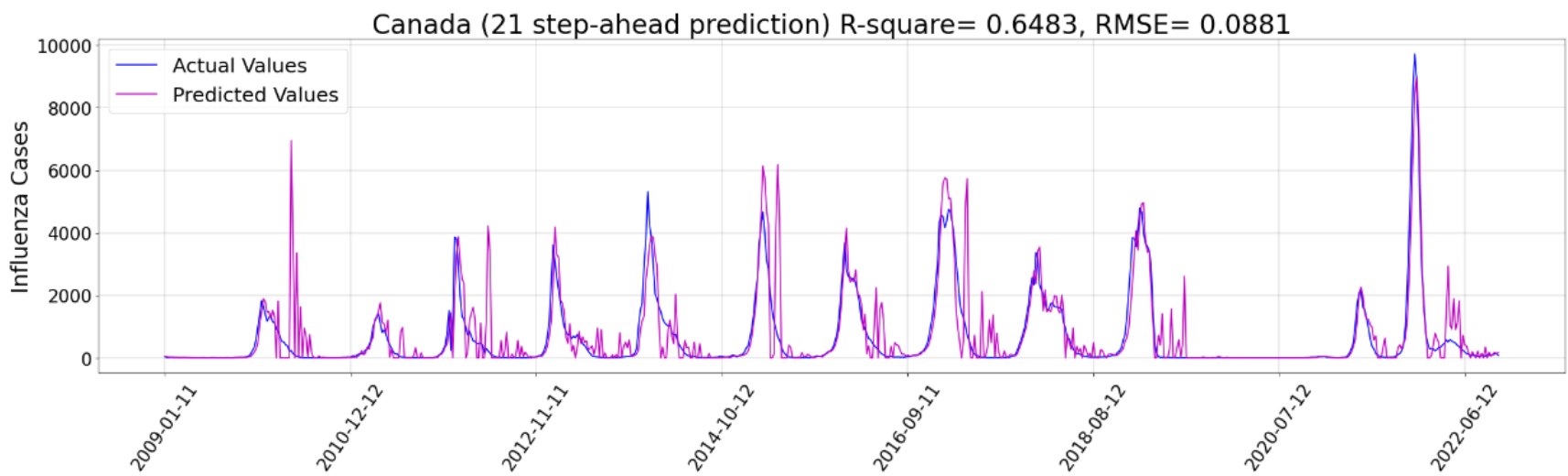
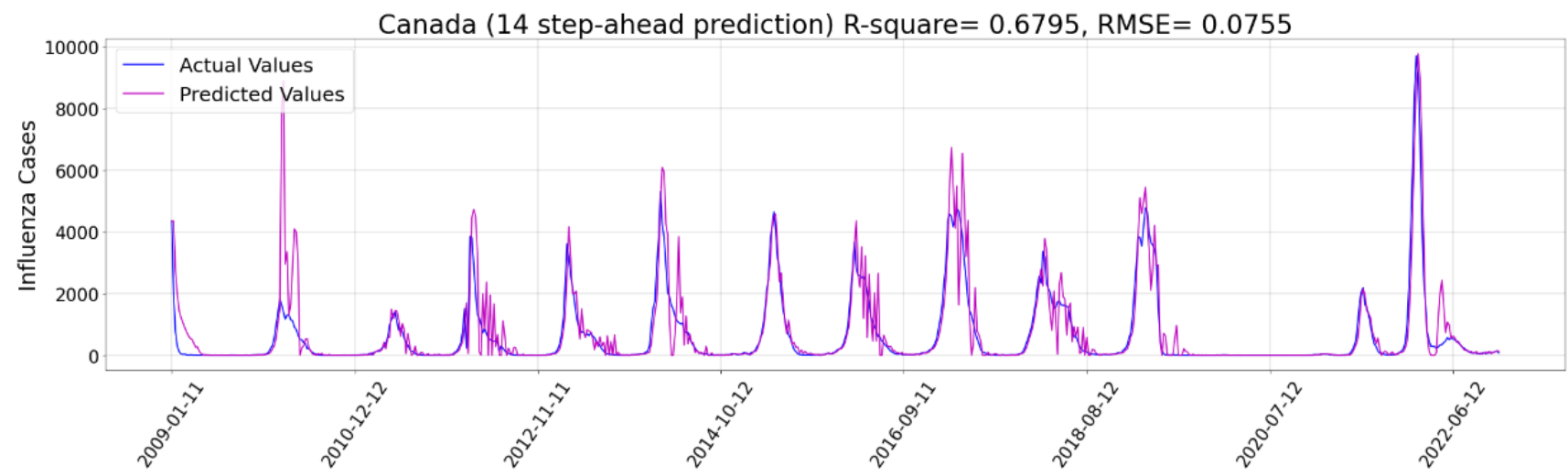
- Les cas de COVID-19 des différentes provinces canadiennes sont publiés sur une base hebdomadaire.
- Le modèle final prédit très bien les vagues de COVID-19 de toutes les provinces.



Prévision de l'influenza



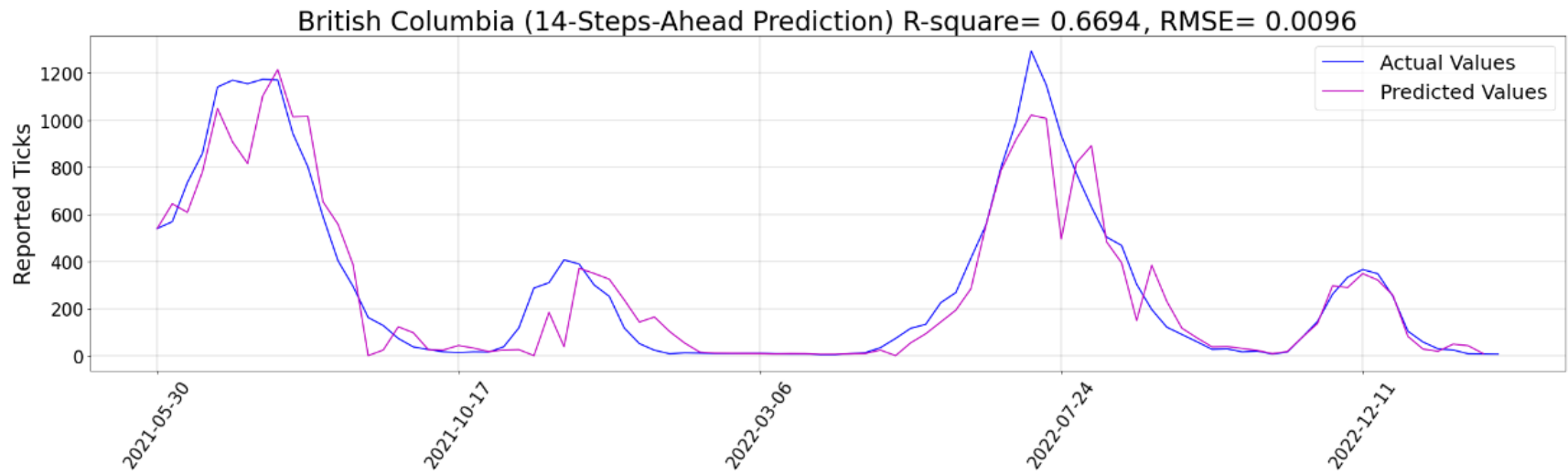
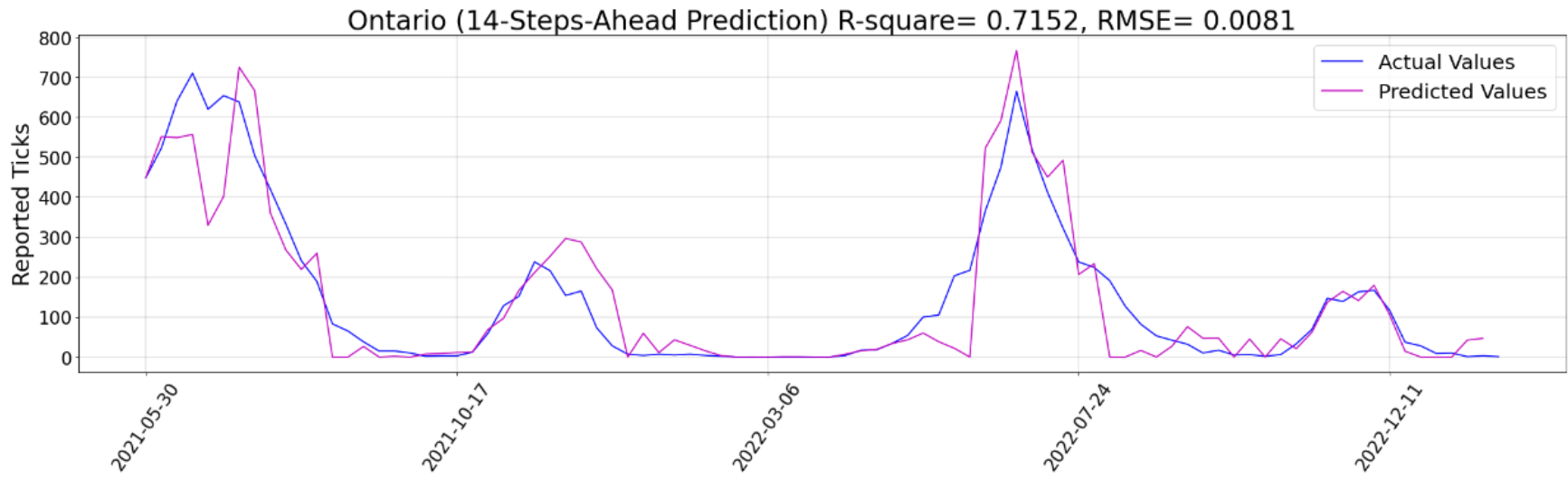
- Des vagues d'influenza ont été prévues pour différents pays avec une précision remarquable.



Prévision de la maladie de Lyme



- L'ensemble de données etick a été utilisé comme indicateur de la prévalence de la maladie de Lyme dans les différentes provinces.
- Notre modèle est capable de prédire le volume de tiques pour différentes provinces avec une précision exceptionnelle.



Ressemblance entre le colonialisme historique et le colonialisme basé sur les données actuelles

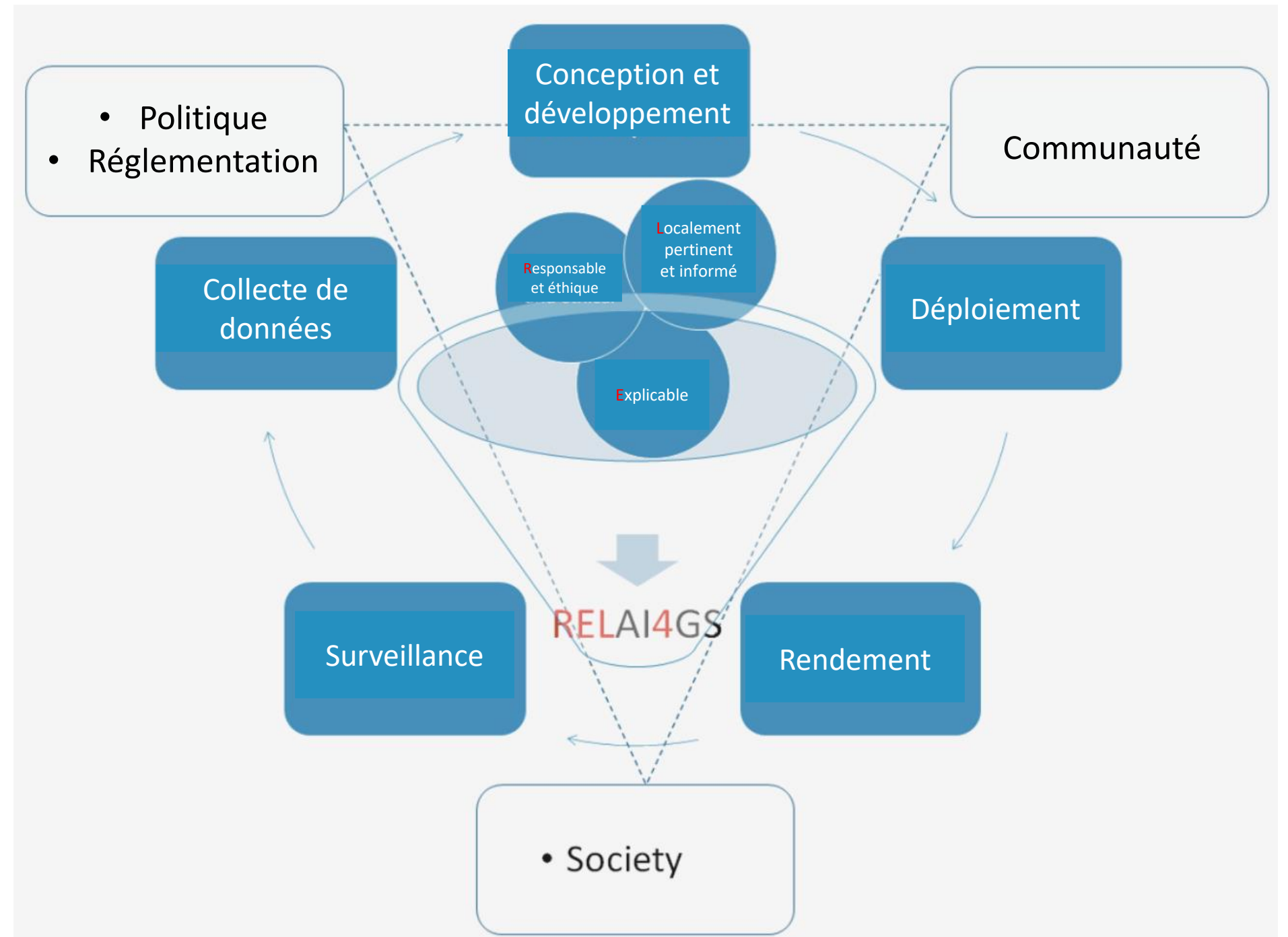


Colonialisme historique	Colonialisme basé sur les données actuelles
Appropriation des ressources naturelles	Appropriation [et quantification] de la vie humaine (par le biais de la mise en données)
Expropriation de terres, de ressources, de personnes	Expropriation de la vie sociale (p. ex. les médias sociaux) et des personnes (p. ex. l'émergence de l'IdO) <ul style="list-style-type: none">▪ Les gens sont « juste là » pour que le capital les « découvre » et les exploite
Exploitation par le capitalisme industriel	Exploitation par le capitalisme de l'IA (marchandisation de la vie humaine)

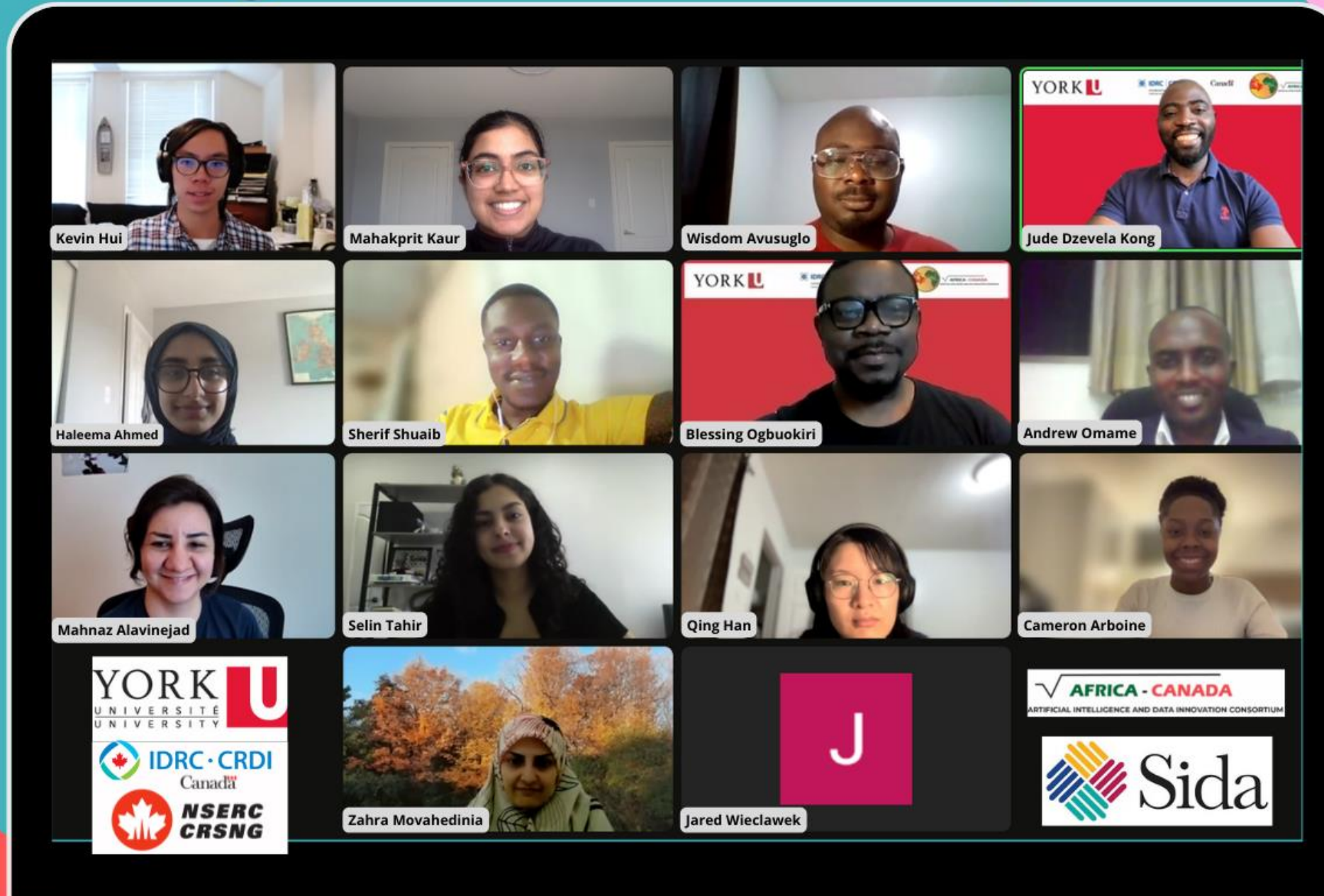
Notre cadre



- **Responsable** : Responsable, vérifiable, conforme, éthique, respectueux, sûr et sécurisé
- **Explicable** : Équitable, juste, interprétable, fiable, reproductible, transparent, digne de confiance, impartial
- **Local** : Autonome, bienveillant, connecté, décolonisé, centré sur l'humain et la communauté, inclusif, intentionnel, intersectionnel, juste, pratique, protecteur, basé sur des processus, durable

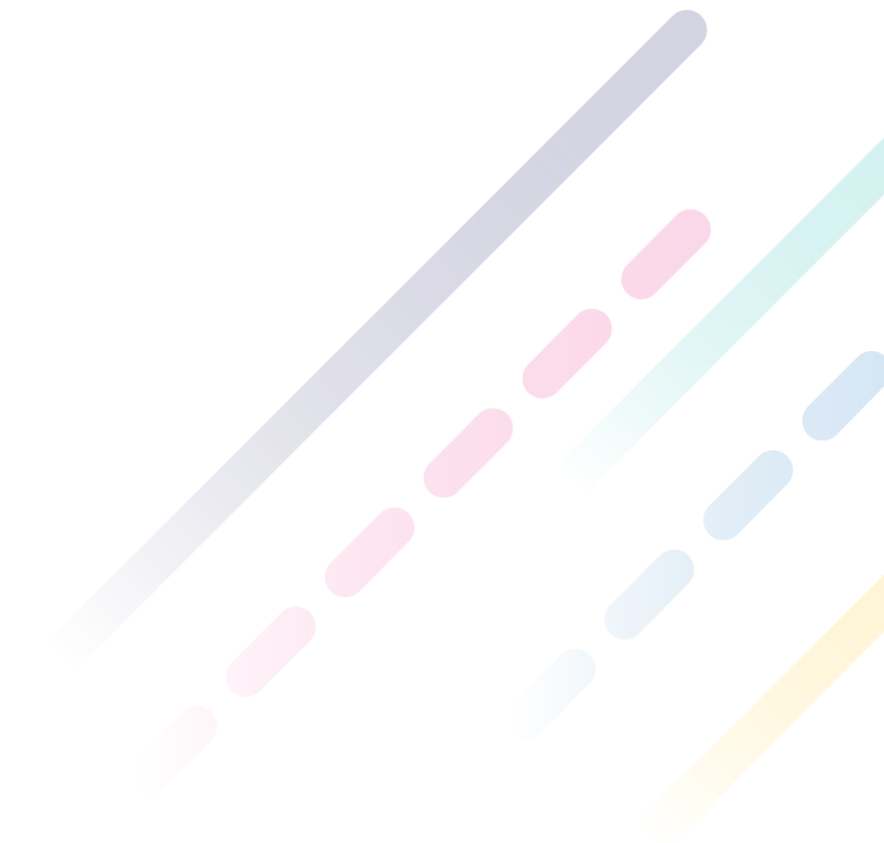


Kong Research Group 2022



Dr. Zahra Nia

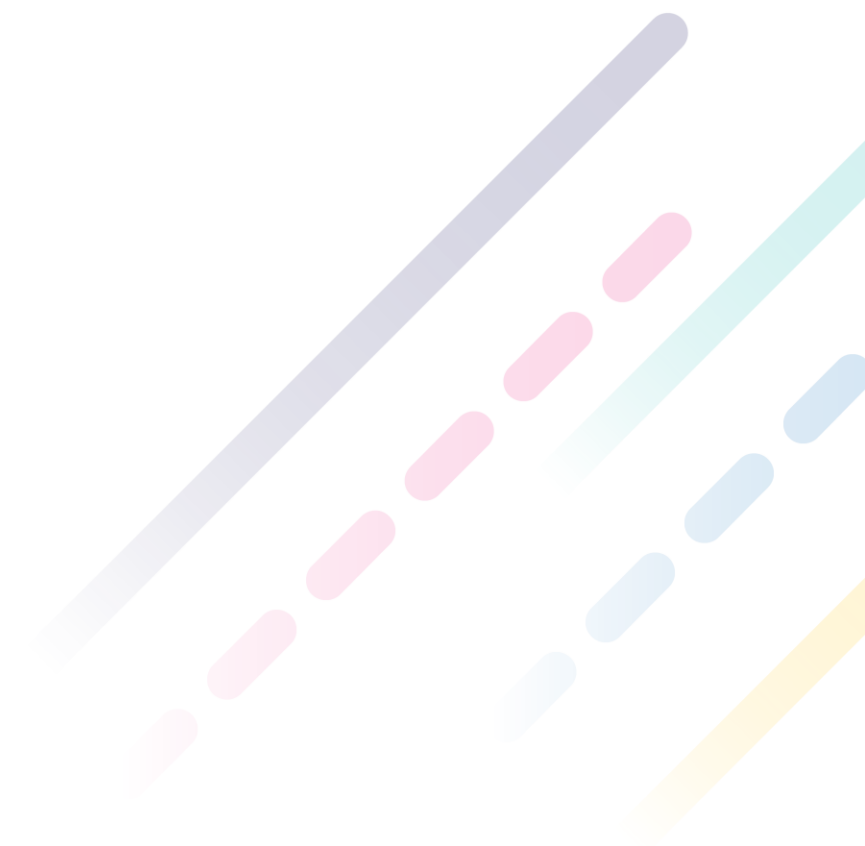
Remerciements : Organismes de financement



Conclusion



- Un système agile d'alerte précoce et de réponse (comme AI-Epidemix) est essentiel pour contrôler et contenir les maladies infectieuses.
- De multiples sources de données (p. ex. Google Trends, les médias sociaux, les données satellites et les images de l'outil Street View) sont collectées au niveau régional et communautaire.
- Les méthodes d'apprentissage automatique, en particulier les réseaux neuronaux récurrents et graphiques, sont utilisées pour prévoir les maladies émergentes et récurrentes.
- Ce cadre aidera les responsables des politiques, les responsables de la santé et les médecins à prévoir avec précision l'apparition de diverses maladies.



**Fourniture de services à partir de modèles :
épidémiologie multipathogène en temps
réel, suivi de la demande de soins de
courte durée et prévisions immédiates via
des modèles de transmission pilotés par les
particules de la méthode de Monte Carlo
par chaîne de Markov (PMCMC)**



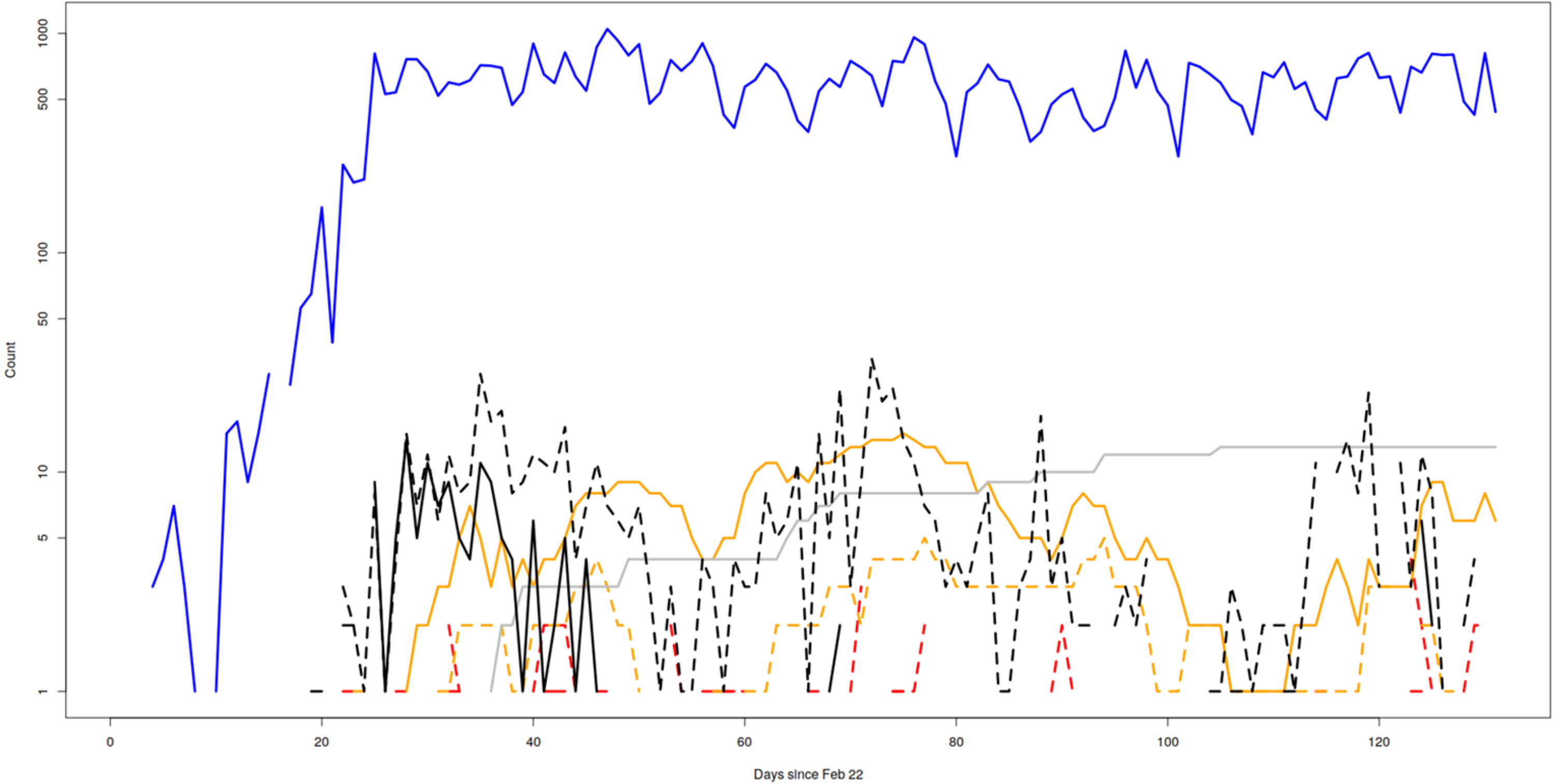
Professeur Nathaniel Osgood, Ph. D.

Laboratoire d'épidémiologie informatique et d'informatique de santé publique

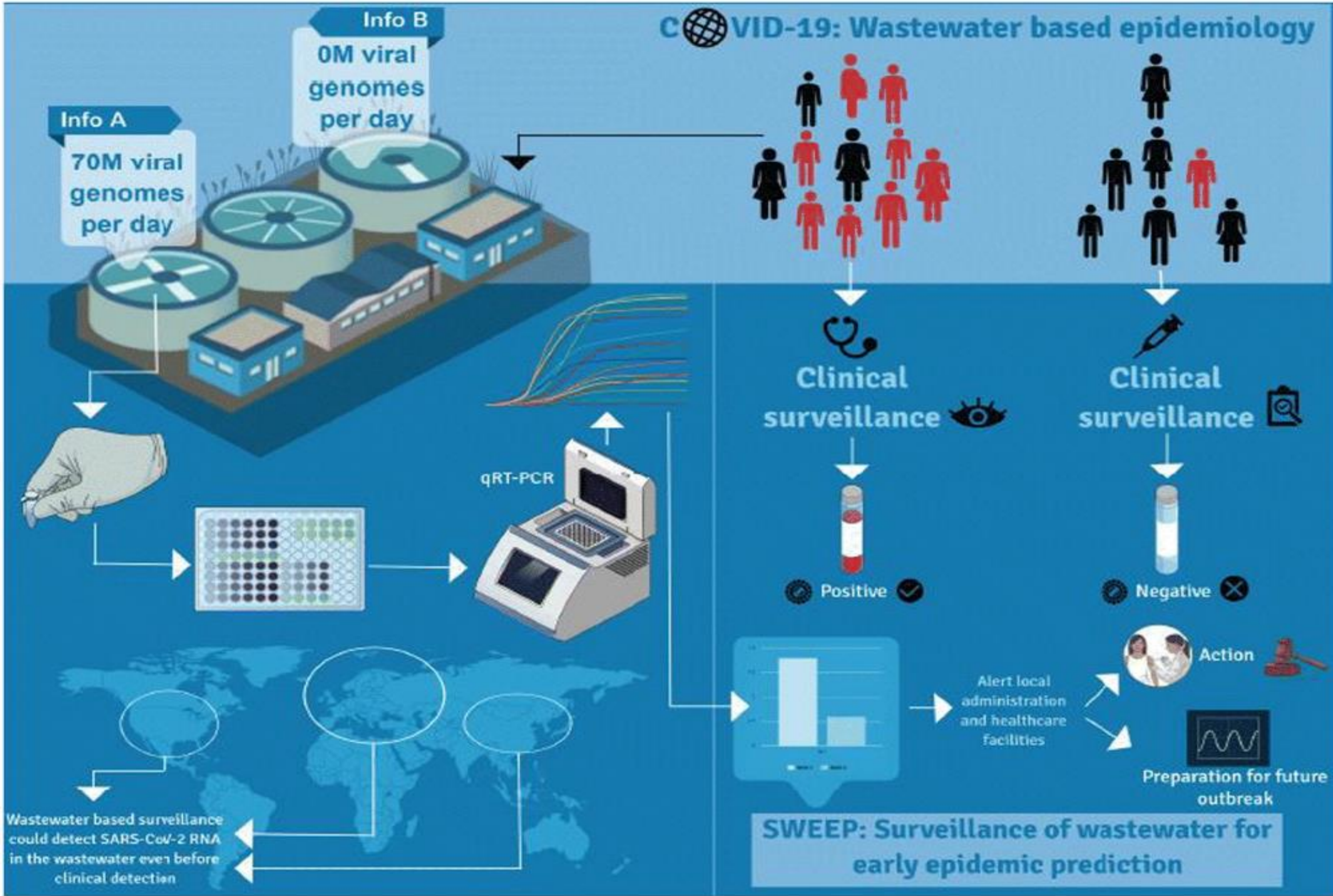
Université de la Saskatchewan



Diverses sources de données déclarées



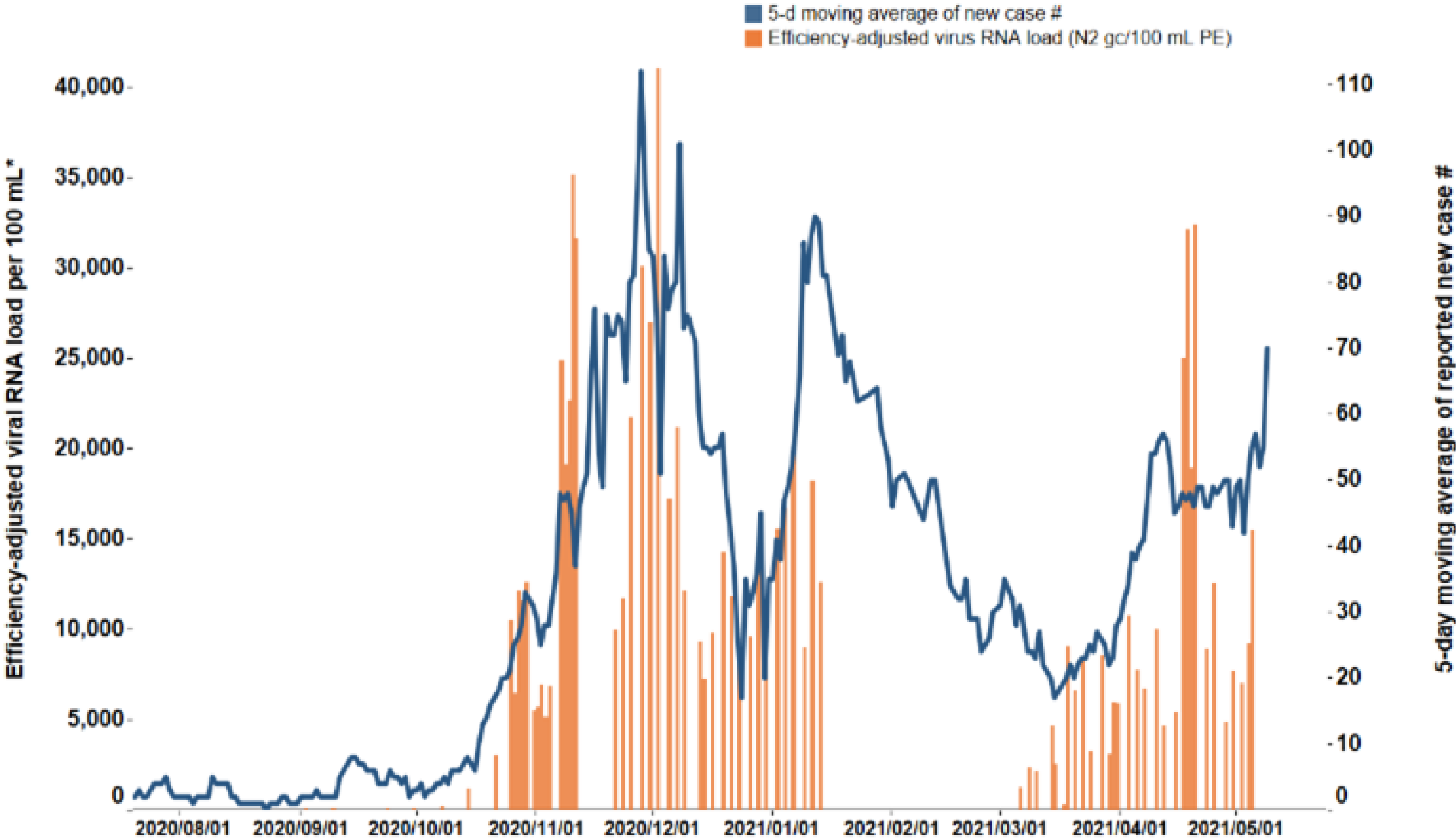
Épidémiologie des eaux usées



Source de l'image :
 Kumar, M., Mohapatra, S.,
 Mazumder, P. et al. Making Waves
 Perspectives of Modelling and
 Monitoring of SARS-CoV-2 in
 Aquatic Environment for COVID-19
 Pandemic. Curr Pollution Rep 6,
 468–479 (2020).
<https://doi.org/10.1007/s40726-020-00161-5>

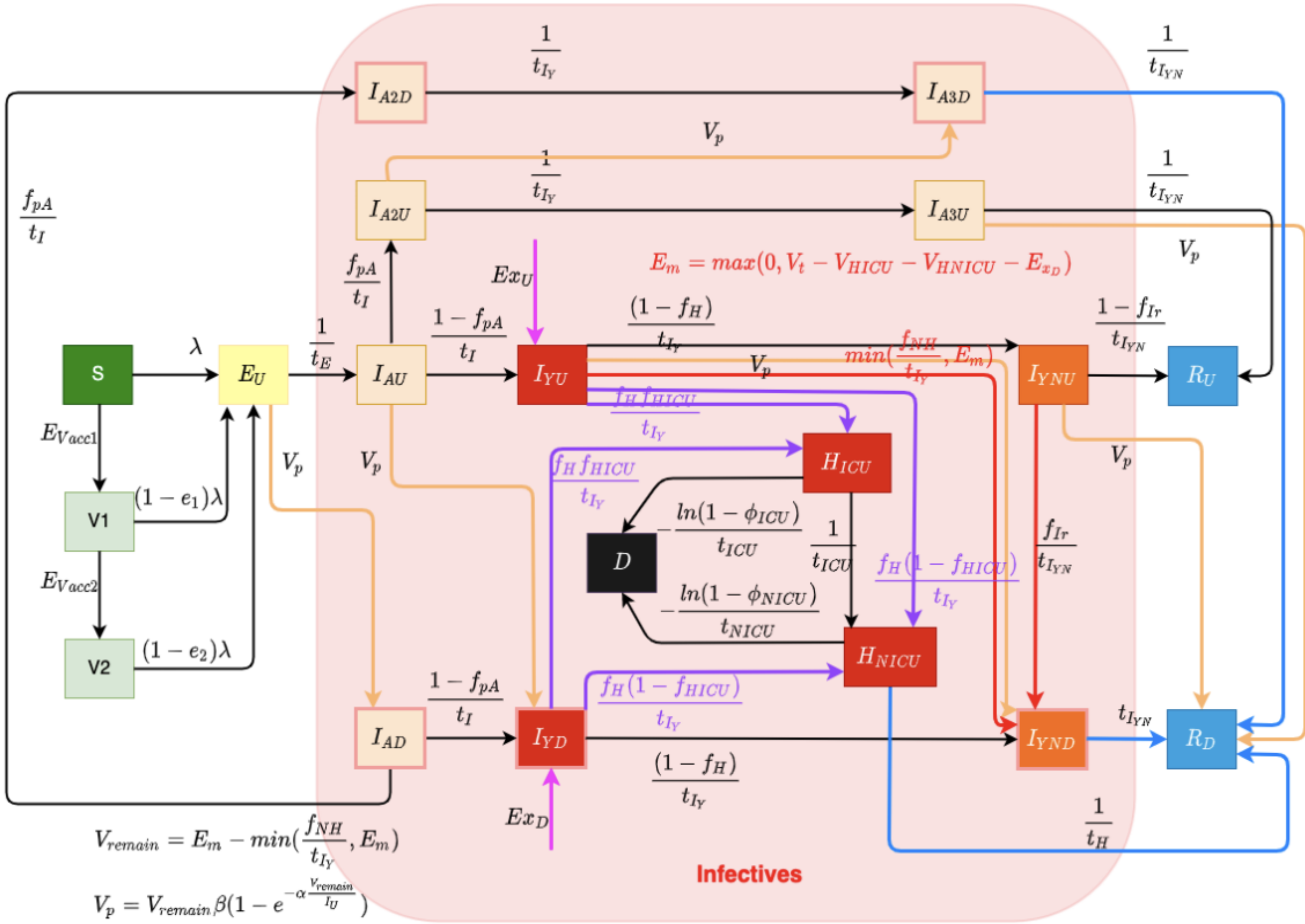
Données sur les eaux usées

Viral RNA load of SARS-CoV-2 in wastewater, Saskatoon



Source :
University of Saskatchewan
Toxicology Centre via COVID-19
Early Indicators,
<https://water.usask.ca/covid-19/#MeasuringVirusIndicatorsinWastewaterasanEarlyWarningofCOVID19Outbreaks>

Modèles de transmission



Réflexions



- Sans aide, tous les modèles s'écartent de la situation empirique au fil du temps.
 - Divergence entre l'état du modèle et l'état empirique
 - Quelques défis pertinents : stochastique, changements exogènes, approximations, omissions, hétérogénéité...
- Les divergences peuvent fortement limiter l'efficacité de l'évaluation des modèles d'intervention.
- Se contenter d'ajuster les paramètres ne permet pas d'assurer un alignement permanent entre le modèle et les données empiriques.



Donner du sens aux données probantes : des modèles dynamiques en tant que services toujours actualisés



- Rendre compte de la compréhension actuelle de l'histoire naturelle de l'infection et des diverses sources de données entrantes dans une image intégrée évolutive de l'épidémiologie sous-jacente actuelle :
 - Changements de comportement
 - Nombre de personnes infectées non diagnostiquées
 - Force de l'infection
 - Répartition régionale des cas
 - Taux de reproduction effectif
- Projection dans les jours et les semaines à venir à partir de la situation actuelle
- Capacité à évaluer des scénarios de politiques pour les besoins en matière de santé publique et de soins de courte durée, en se projetant dans l'avenir à partir de la situation actuelle





Comme pour les prévisions météorologiques, nous mettons à jour nos modèles pour refléter les données les plus récentes et les utiliser pour anticiper l'état futur (scénarios d'intervention, projections).

ROGERS -1° 83% 9:58 AM

Regina, Canada

-1°
SNOW
4°/-7°

SNOW ENDING 12:15 PM →

10 AM	11 AM	12 PM	1 PM	2 PM
-1°	0°	1°	2°	3°
95%	70%	45%		

ACCUMULATION 3-7 CM →

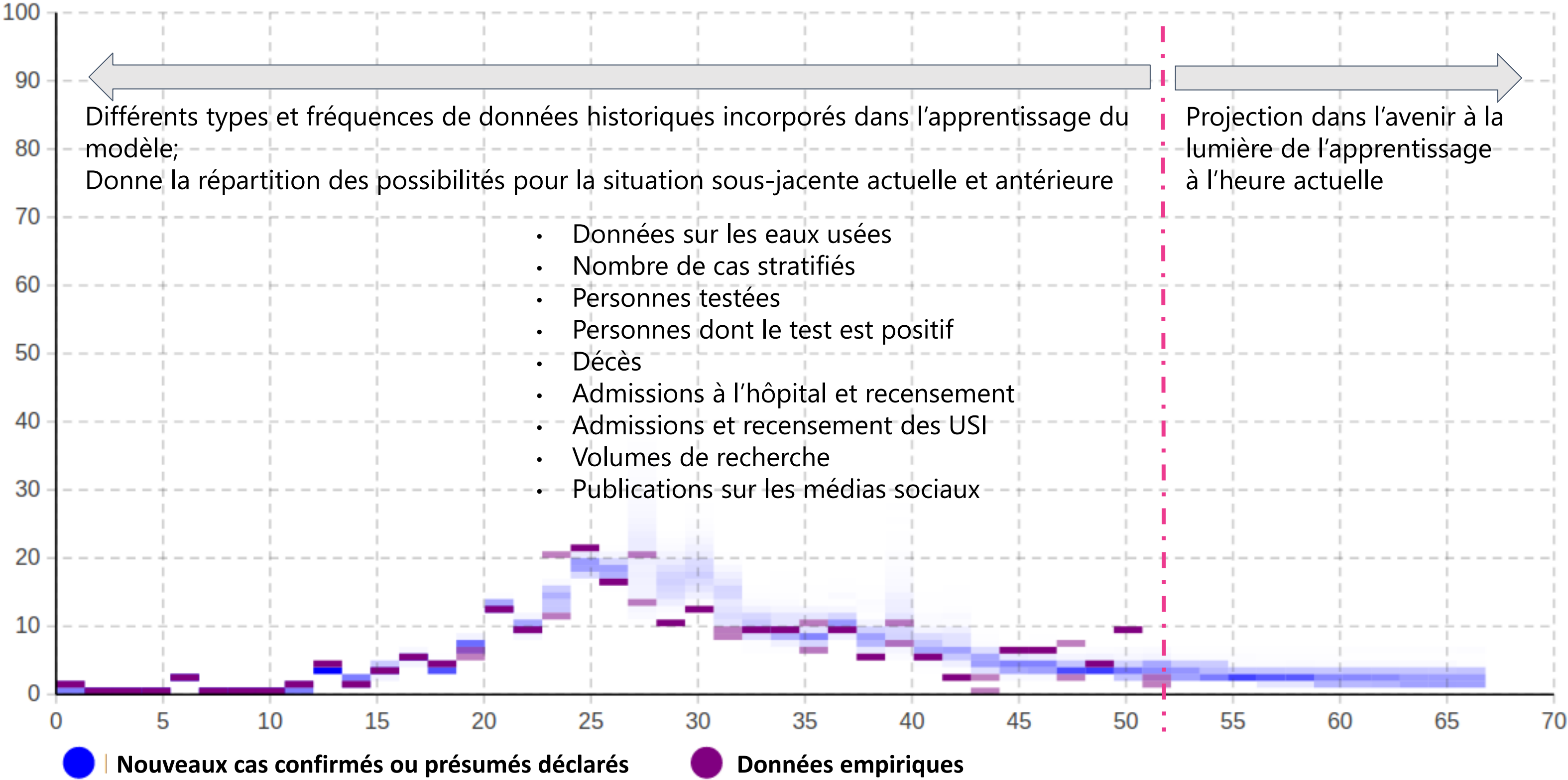
Points du niveau le plus élevé



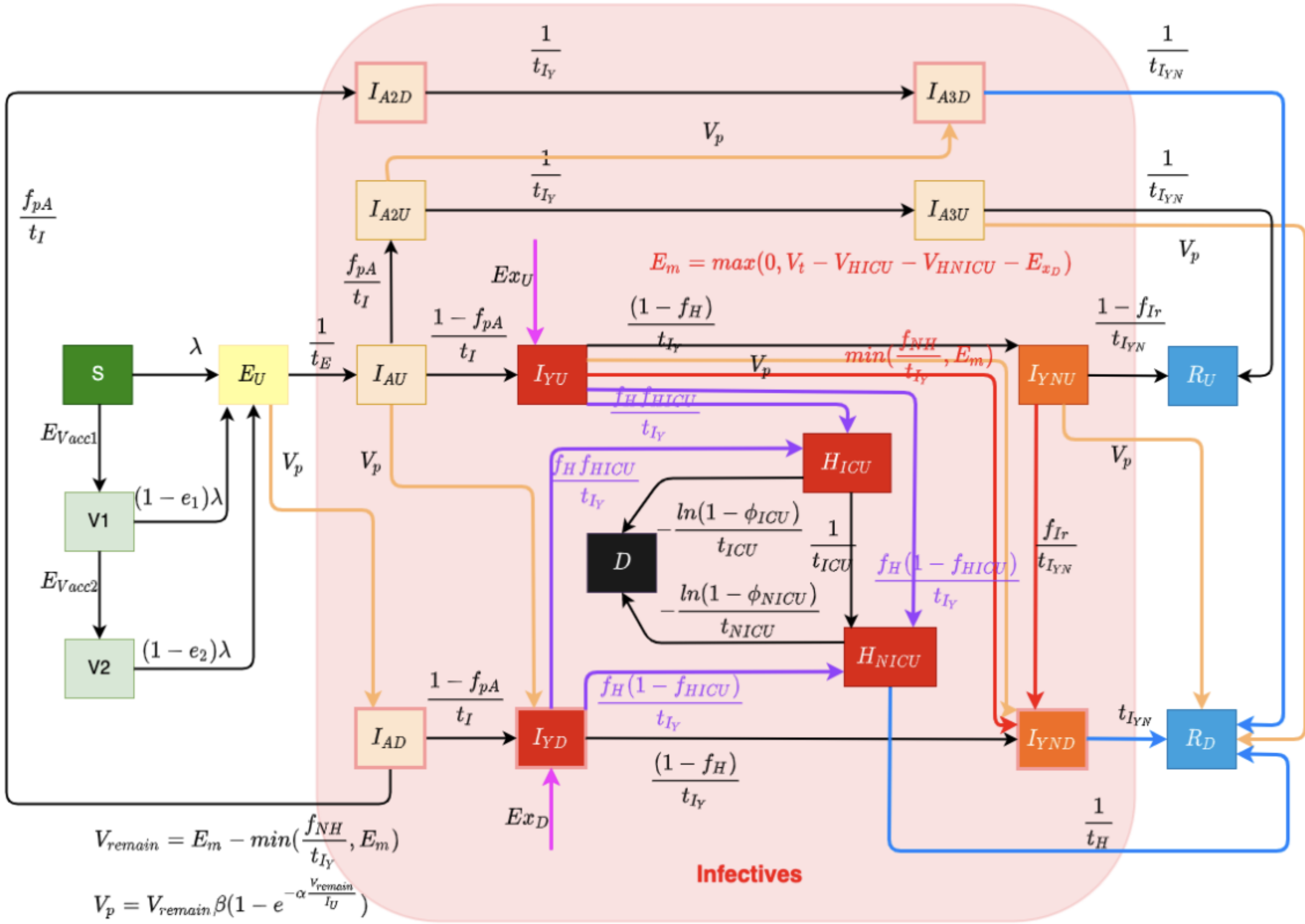
- Il ne s'agit pas de **modèles d'ajustement de courbes**. Ils tentent plutôt de déterminer ce qui se passe en termes de la situation sous-jacente (à la lumière de la théorie et des observations actuelles).
- Le modèle utilise des probabilités bayésiennes et des modèles dynamiques pour déterminer une compréhension cohérente – conforme à la compréhension clinique et épidémiologique de la COVID-19 – de la situation actuelle qui explique le mieux ce qui est observé à travers de nombreux types de données (p. ex. les cas, les volumes de tests, les admissions à l'hôpital, le recensement des hôpitaux, etc.).
 - Chaque type de données reflète une facette différente de cette situation sous-jacente.
- Ce processus implique une inférence de l'IA à partir des données observées d'une manière qui corresponde à la compréhension de l'histoire naturelle et de l'épidémiologie de la COVID-19 (telle qu'elle est définie dans le modèle).
- Puisque les modèles déduisent la situation sous-jacente conformément à la théorie, ils peuvent se projeter dans l'avenir avec ou sans interventions ou mesures supplémentaires.
- Ces projections ne sont pas des courbes prévisionnelles, mais plutôt des projections de la « dynamique » de la situation.



Planification adaptative : observer l'évolution des données probantes



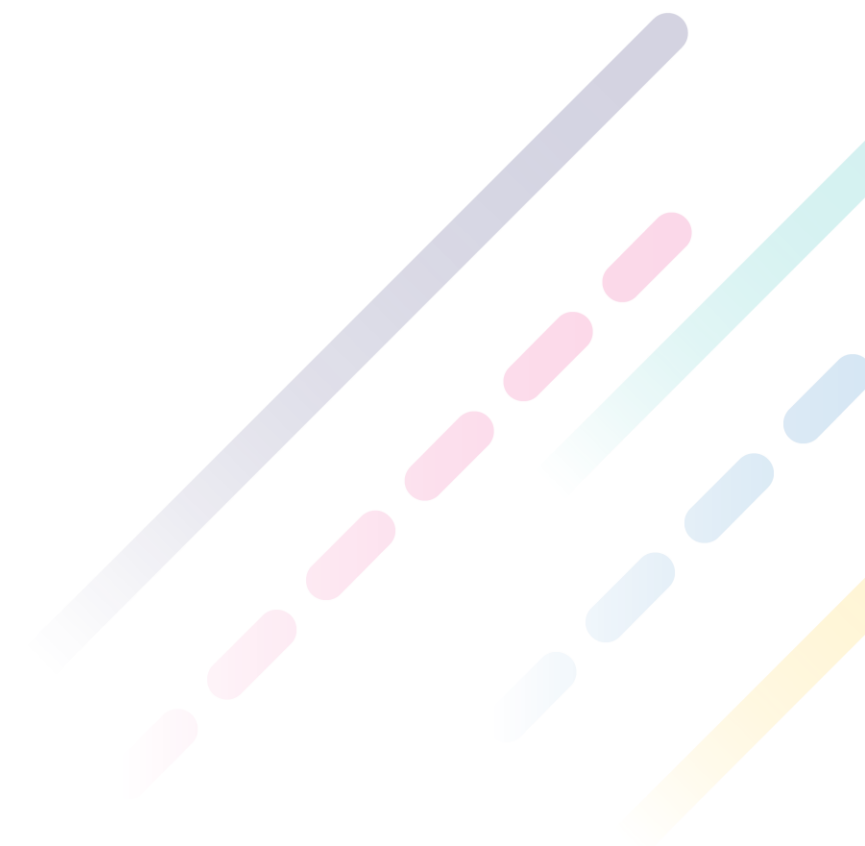
Modèles de transmission



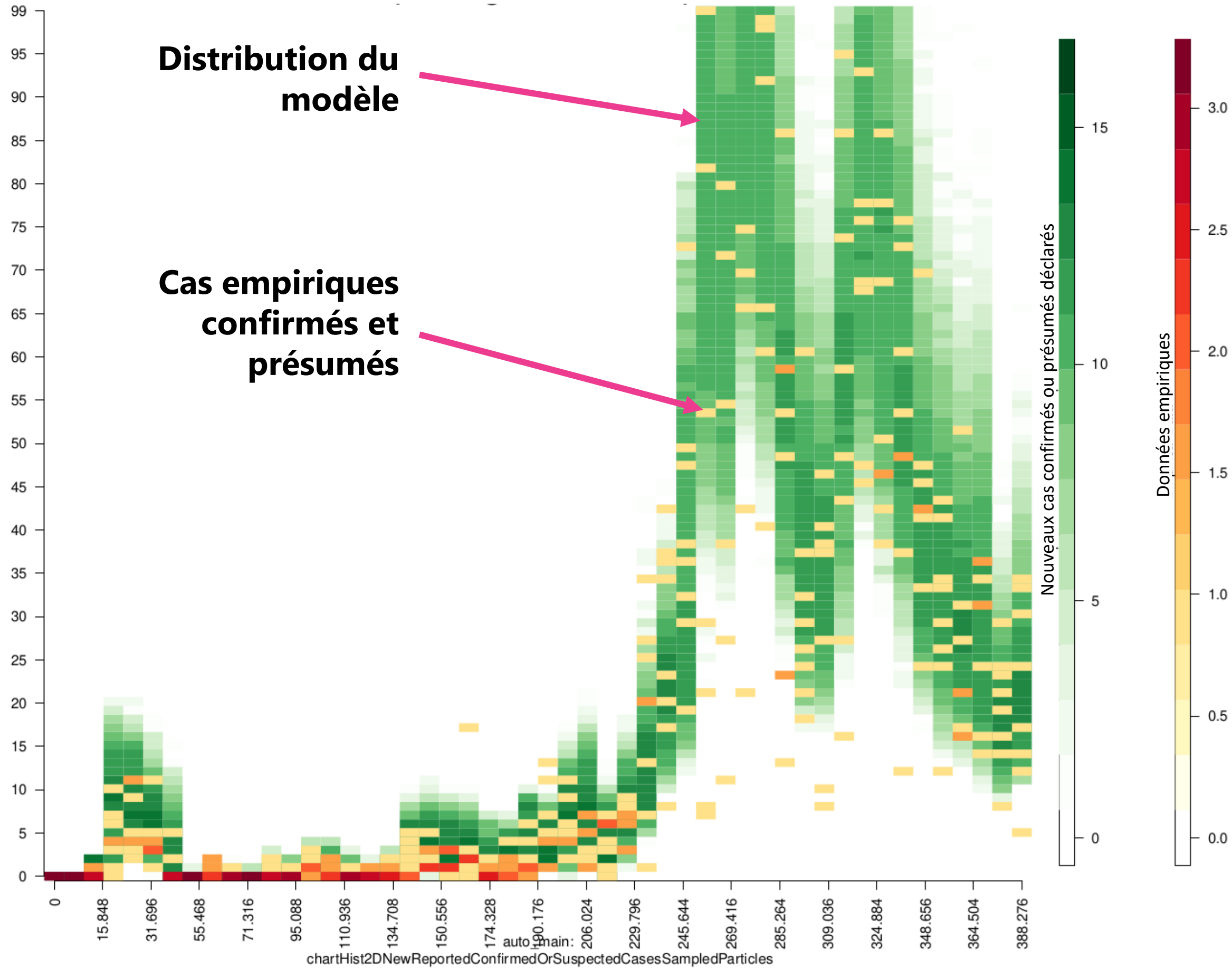
Utilisations du modèle du filtre à particules et des modèles de PMCMC



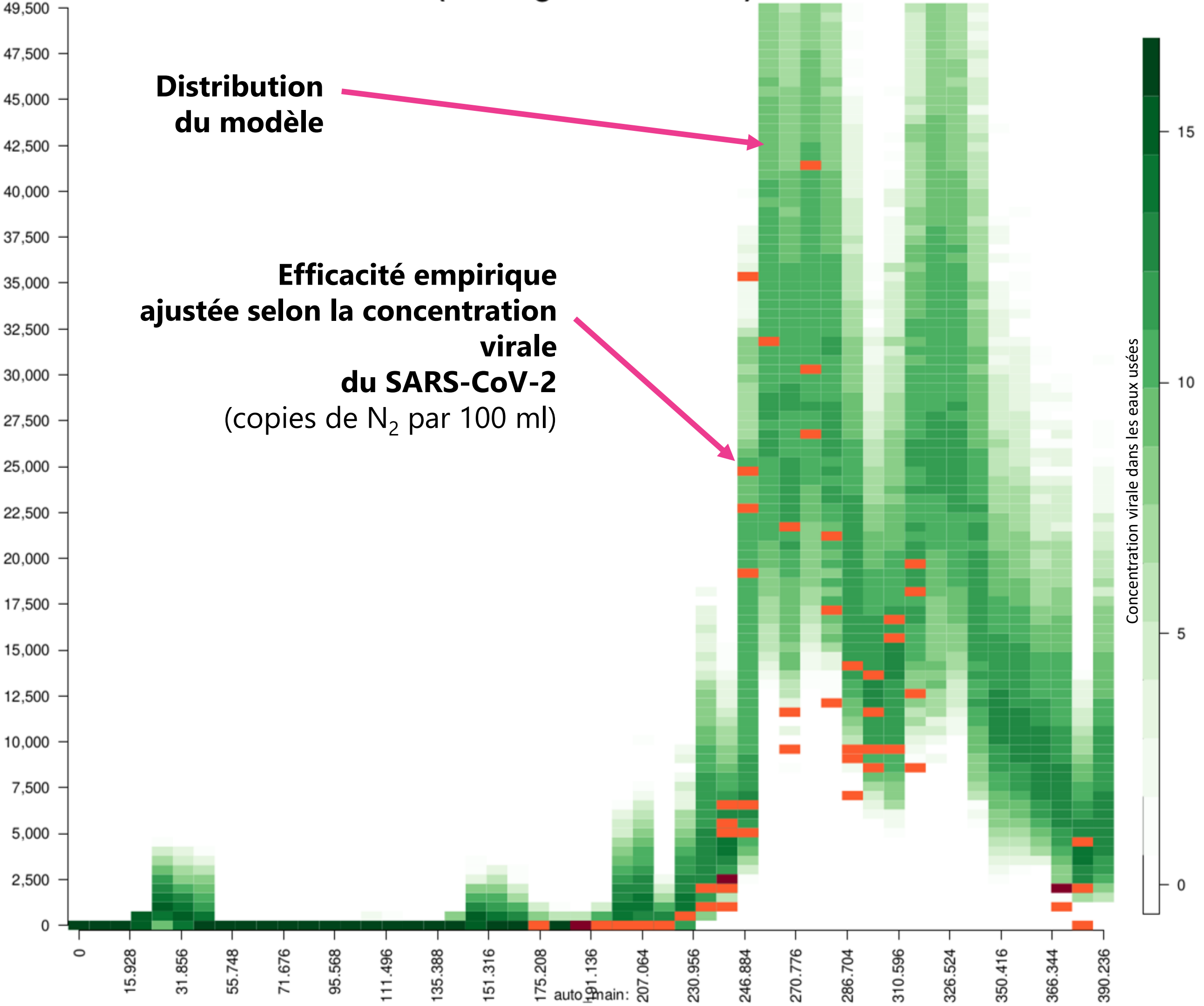
- **Tomographie des populations** : Fournir une interprétation probabiliste consensuelle (par le biais d'une distribution conjointe) de la situation actuelle et passée
- **Projection et prévision** : Projection à partir de maintenant avec la dynamique du modèle et les hypothèses de « statu quo » ou de diffusion concernant les tests actifs, les tendances de contact, etc.
- **Extrapolation rétrospective** : Reconstruction historique basée sur des données antérieures et postérieures
- **Évaluation des politiques** : Évaluation de portefeuilles d'interventions, de scénarios exogènes ou d'autres possibilités de simulation à l'aide d'une image cohérente fondée sur les données les plus récentes



Nouveaux cas confirmés ou présumés declares (sur l'échelle de couleur Log2)



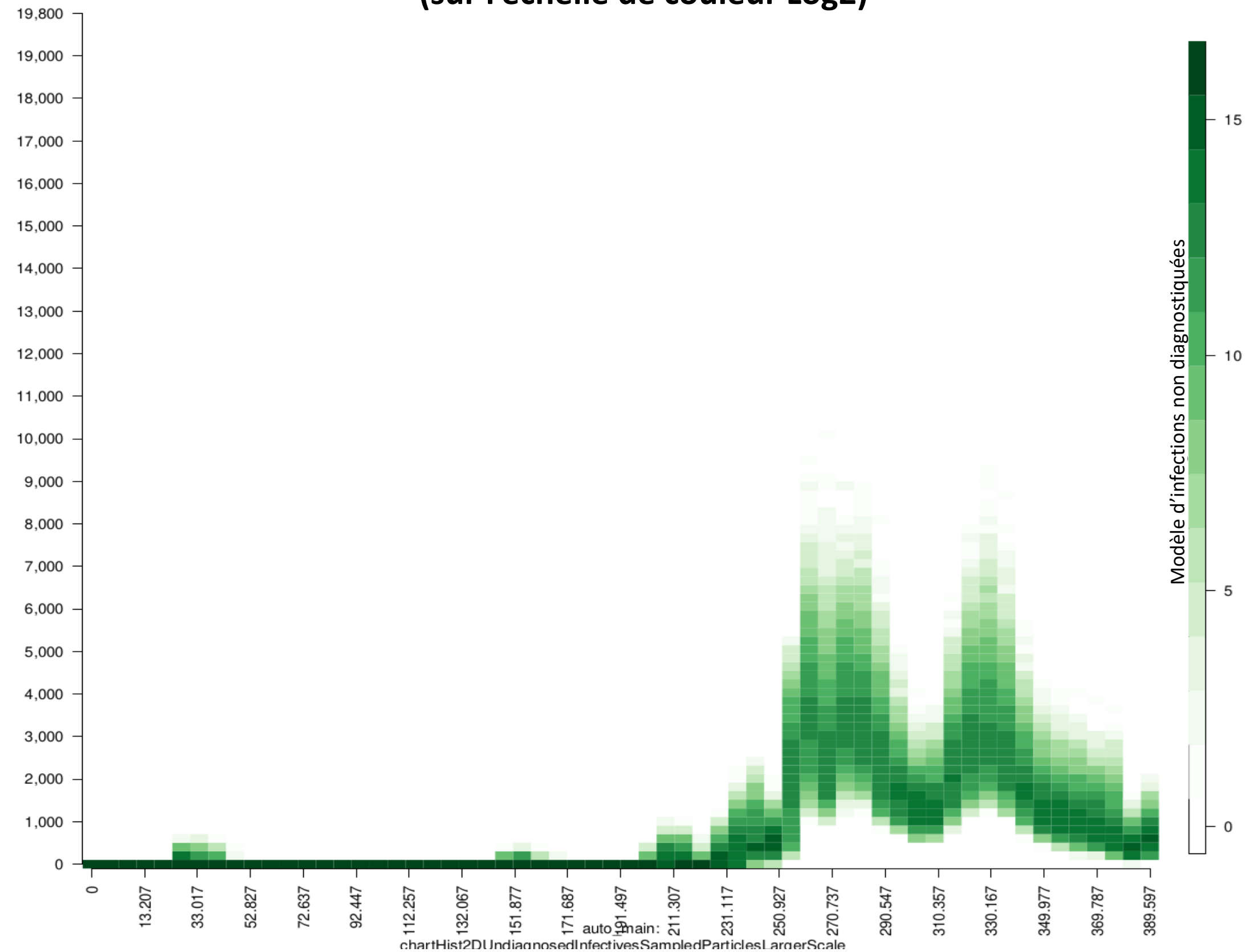
Série chronologique de données sur la concentration des eaux usées



Aperçu génératif des quantités non mesurées

État latent :
Infections non
diagnostiquées

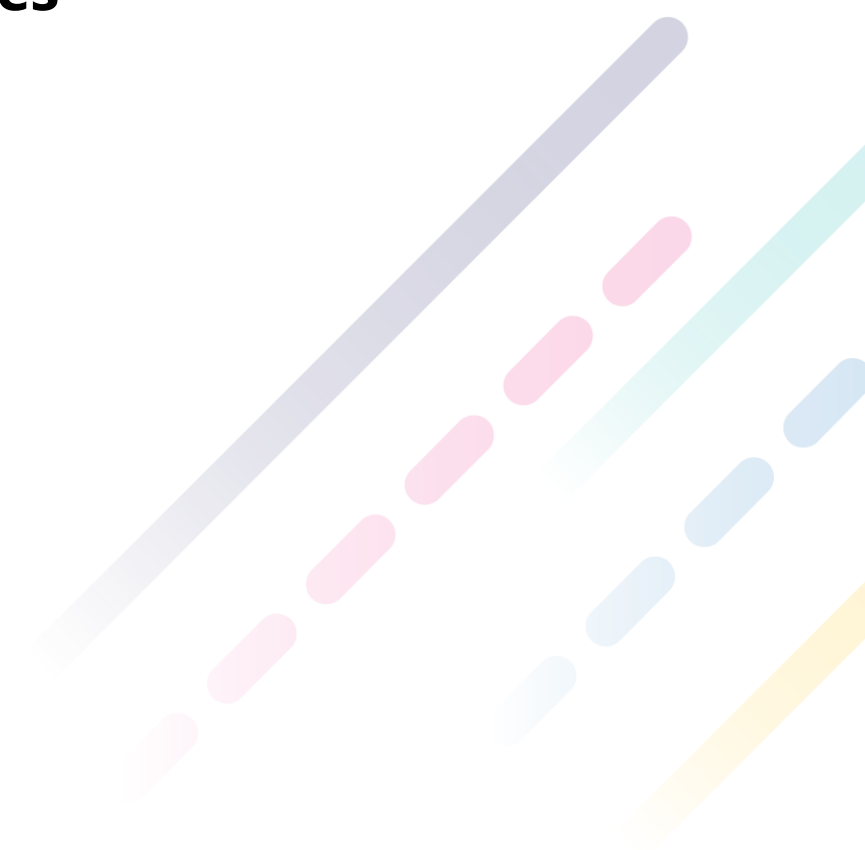
Modèle d'infections non diagnostiquées
(sur l'échelle de couleur Log2)



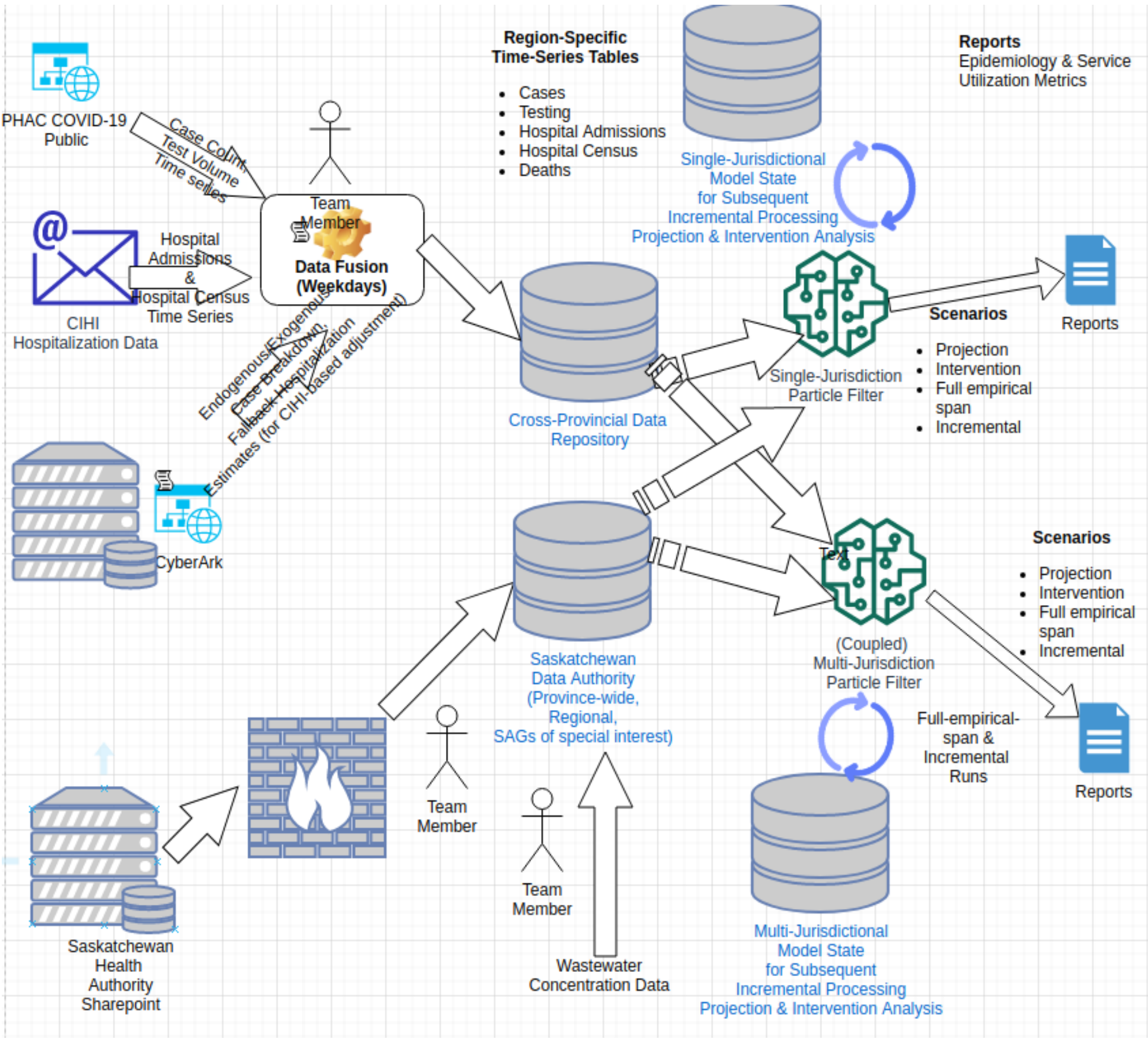
Avantages



- Obtenir une image unifiée, à l'échelle du système, cohérente avec l'histoire naturelle de l'infection à partir de diverses sources de données
- Capacité à estimer à la fois l'état épidémiologique sous-jacent et les paramètres
- Capacité à prendre en charge de nouvelles sources de données
- Tolérance pour les données manquantes et différents niveaux de qualité des données
- Applicabilité à une grande variété de modèles épidémiologiques
- Un tout plus grand que la somme des parties : capacité à affiner les estimations pour une infection donnée en prenant en compte les données relatives à **plusieurs agents pathogènes ayant des facteurs communs**
- Viabilité des cadres de surveillance épidémiologique et comportementale en temps réel, évolutifs et appuyés par des modèles, qui consomment divers types de données, y compris pour des agents pathogènes multiples



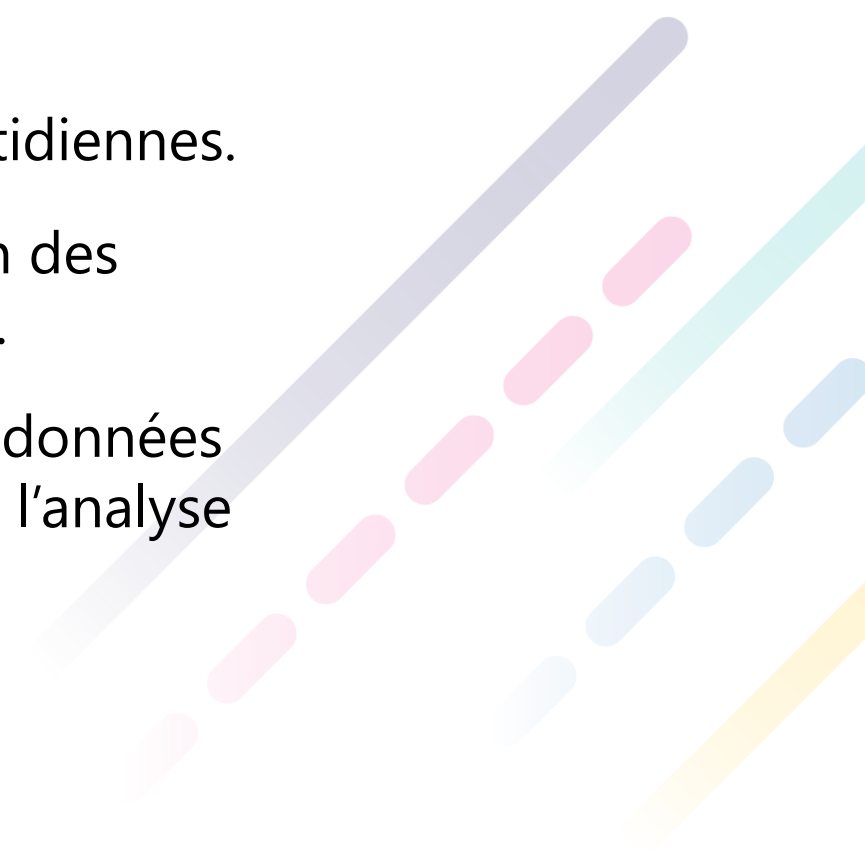
Cadre de force industrielle déployé durant la pandémie



Conclusions



- Les modèles acquièrent une valeur ajoutée considérable s'ils sont utilisés pour la fourniture de services avec des mises à jour régulières.
- La combinaison d'algorithmes bayésiens d'IA et d'apprentissage automatique avec des modèles théoriques permet d'obtenir des modèles « toujours actualisés » pour comprendre la situation actuelle et se projeter dans l'avenir.
- Ces méthodes sont synergiques avec la collecte de données à grande échelle utilisant des versions à grande vitesse des informations traditionnelles (p. ex. les essais) et nouvelles (p. ex. les eaux usées).
- En présence de modèles dynamiques agrégés, le filtrage particulière pour la COVID-19 peut donner de bons résultats aux niveaux national, régional et local.
- Ces méthodes permettent d'intégrer diverses séries chronologiques, notamment WW, SM et Sear.
- Grâce à l'informatique parallèle et distribuée actuelle, il est possible d'effectuer des mises à jour quotidiennes.
- Des pipelines de communication appropriés peuvent permettre une ingestion et une communication des données évolutives et efficaces, ainsi qu'une exploration interactive pour éclairer la prise de décision.
- Le tout est plus grand que la somme des parties : les premiers travaux suggèrent que l'utilisation de données relatives à plusieurs agents pathogènes présentant des facteurs de risque communs permet d'affiner l'analyse d'un facteur de risque donné.



Période de discussion

Des questions?

Veillez utiliser l'**onglet Q&R** pour soumettre vos questions à nos intervenants. Vous pouvez « **aimer** » les questions des autres pour les faire remonter dans l'ordre de priorité.



Mot de la fin



Merci!

Participez au prochain séminaire le mardi 30 janvier 2024 (de 13 h à 14 h, HE)!

Veillez répondre à notre **sondage** qui sera accessible peu de temps après le séminaire. Scannez le code QR.

L'enregistrement et les diapositives de la présentation du séminaire seront accessibles dans un délai de deux semaines sur le site Web du CCNMI : <https://ccnmi.ca/>.

Rendez-vous au <https://ccnmi.ca/serie-de-seminaires-sur-les-progres-realises-dans-le-domaine-de-la-surveillance/> pour obtenir plus de renseignements sur la série de séminaires sur les progrès dans le domaine de la surveillance.

