

*Présentation de DeepLearning intitulé :*

*GAN POUR LA PRÉDICTION DE SÉRIES  
CHRONOLOGIQUES, L'ASSIMILATION DE DONNÉES  
ET  
QUANTIFICATION DE L'INCERTITUDE*

1

**Réaliser Par:**

**ERRAFIA Chaimaa**

01-02-2022

# *Planning*



Introduction



Méthodes



Description du cas de test



Résultats



Conclusion

# Introduction

- Deep Learning a gagné en popularité dans le calcul scientifique
- Utilisés par les industries qui résolvent des problèmes complexe
- utilise différents types de réseaux de neurones pour effectuer des tâches spécifiques.
- fonctionne sur la base de la structure et de la fonction du cerveau humain.
- les soins de santé, le commerce électronique, le divertissement et la publicité

# Objectif

- proposer une nouvelle méthode dans laquelle un GAN est utilisé pour quantifier l'incertitude des simulations directes en présence de mesures.
- utiliser comme modèle de substitution pour prédire l'évolution de la distribution spatiale des états de simulation et les données observées
- quantifier l'incertitude d'un modèle compartimenté en épidémiologie, qui représente la propagation du COVID-19 dans une ville idéalisée

# *MÉTHODES*

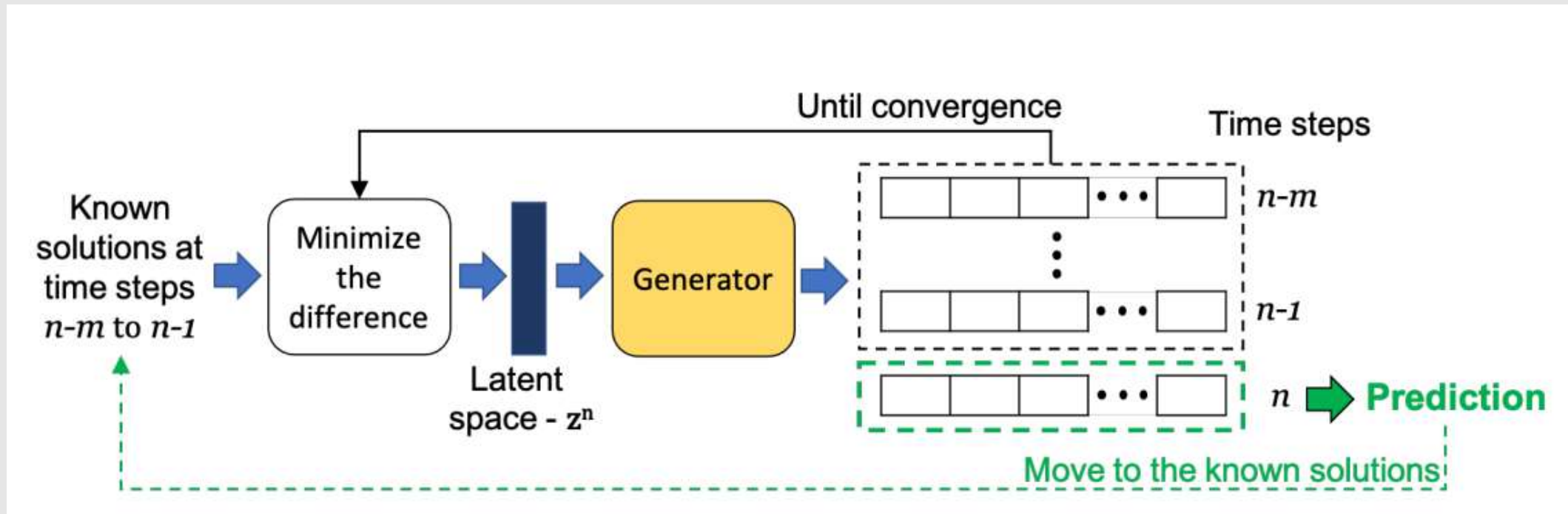
# *1 Réseaux antagonistes génératifs (GAN)*

- algorithme d'apprentissage profond génératif
- créer de nouvelles instances de données qui ressemblent aux données de formation.
- GAN a deux composants :
  - **un générateur**, qui apprend à générer de fausses données.
  - **un discriminateur**, qui apprend à partir de ces fausses informations.
- prédire des solutions spatio-temporelles pour l'écoulement de fluide à super-résolution

## 2 GAN pour la prédiction de séries temporelles(PredGAN)

- Nous entraînons le GAN à produire des données à une séquence de  $m + 1$  pas de temps
- étant donné un vecteur latent  $z$ , la sortie du générateur  $g(z)$  seront des données à des pas de temps  $n-m$  à  $n$
- effectuer une optimisation pour correspondre au premier  $m$  niveaux de temps en sortie du générateur avec les solutions connues
- Après convergence, le dernier pas de temps,  $m+ 1$ , dans la sortie du générateur est la prédiction.

# Aperçu du processus PredGAN.

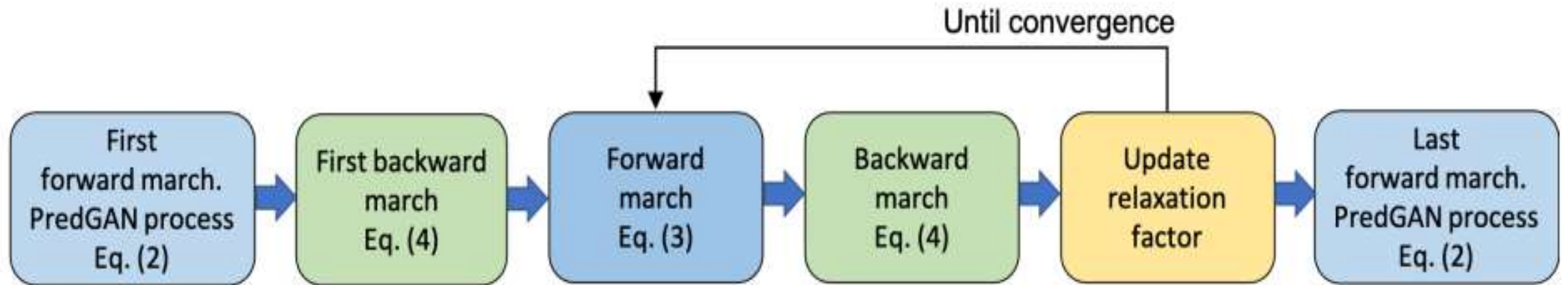




### 3 GAN pour l'assimilation de données (DA-PredGAN)

- L'assimilation de données est un type de problème inverse qui vise à incorporer des données observées dans des modèles mathématiques
- on a utiliser le GAN prédictif d'assimilation de données (DA-PredGAN) qui intègre trois changements principaux dans le PredGAN
  1. Un terme supplémentaire est inclus dans la fonction de perte dans l'équation. (2) pour tenir compte de l'inadéquation des données entre les données observées et les valeurs générées.
  2. Le but de l'assimilation des données est de faire correspondre les données observées et de déterminer les paramètres du modèle  $\mu_k$  (entrées du modèle numérique). Par conséquent, ils ne sont pas connus a priori, comme dans la prédiction.
  3. La marche en avant dans le temps est maintenant remplacée par la marche en avant et en arrière

*Vue d'ensemble du DA-PredGAN. Chaque marche représente le passage par tous les pas de temps.*



# *4 GAN pour la quantification de l'incertitude*

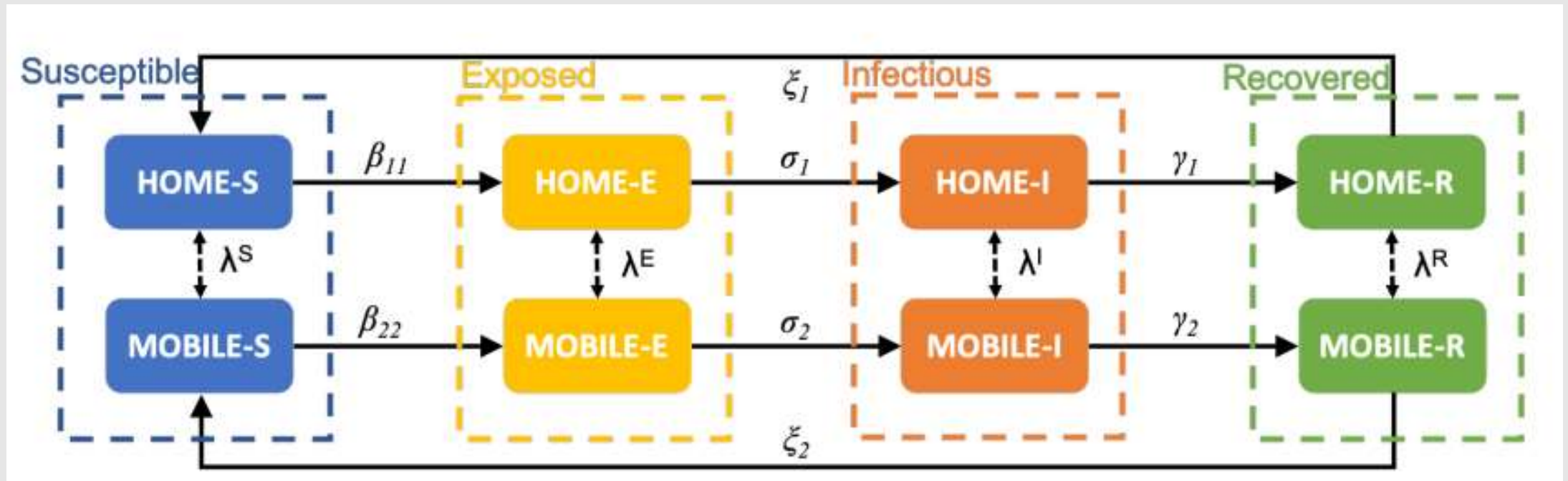
- la méthode proposée UQ-PredGAN permet de calculer des incertitudes en s'appuyant uniquement sur un ensemble de simulations numériques inconditionnées,
- Les prédictions, l'assimilation des données et la quantification de l'incertitude sont effectuées à l'aide de la capacité adjointe inhérente présente sur le GAN
- L'idée est de générer plusieurs modèles qui correspondent aux données observées et peuvent quantifier l'incertitude dans les états du modèle (sorties) et les paramètres des modèles (entrées).
- plusieurs assimilations de données à l'aide de l'algorithme DA-PredGAN

# *DESCRIPTION DU CAS DE TEST*

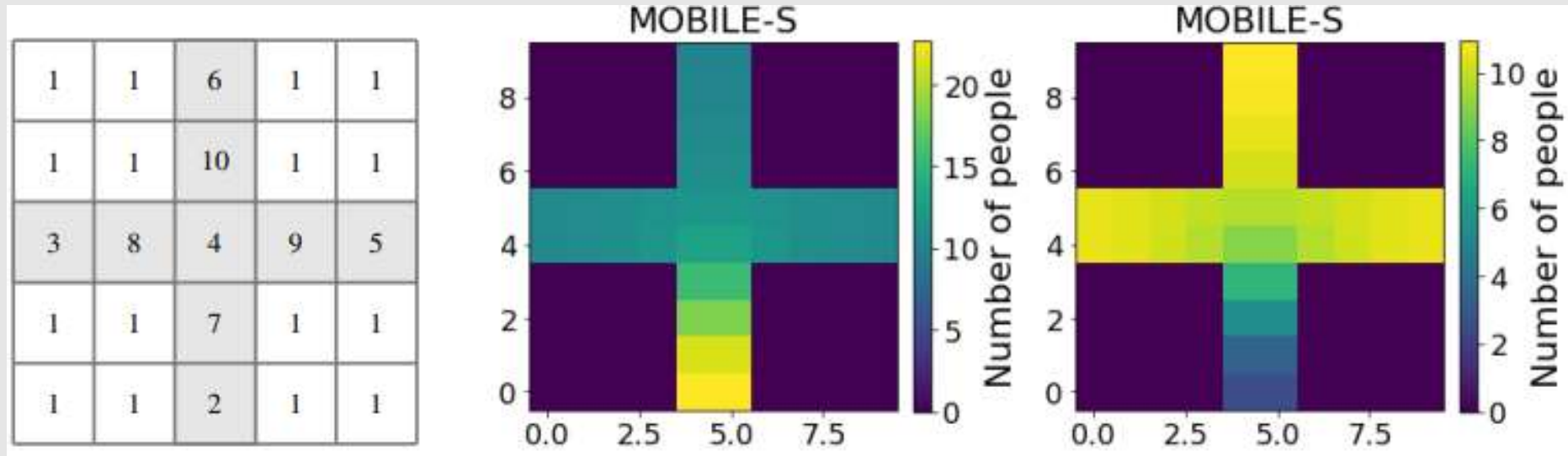
# 1 Modèle SEIRS étendu

- Le modèle SEIRS étendu est un modèle non linéaire avec quatre compartiments sensibles (S), exposés (E), infectieux (I) et récupérés (R) deux groupes de personnes (Maison – Mobile).
- reproduire le cycle quotidien de la nuit et du jour
- les personnes mobiles rentrent chez elles la nuit, et il y aura beaucoup de personnes qui quitteront leur domicile pendant la journée
- utilise un terme de diffusion et un terme d'interaction pour modéliser ce processus

# Schéma du modèle SEIRS étendu:



## 2 Configuration du problème



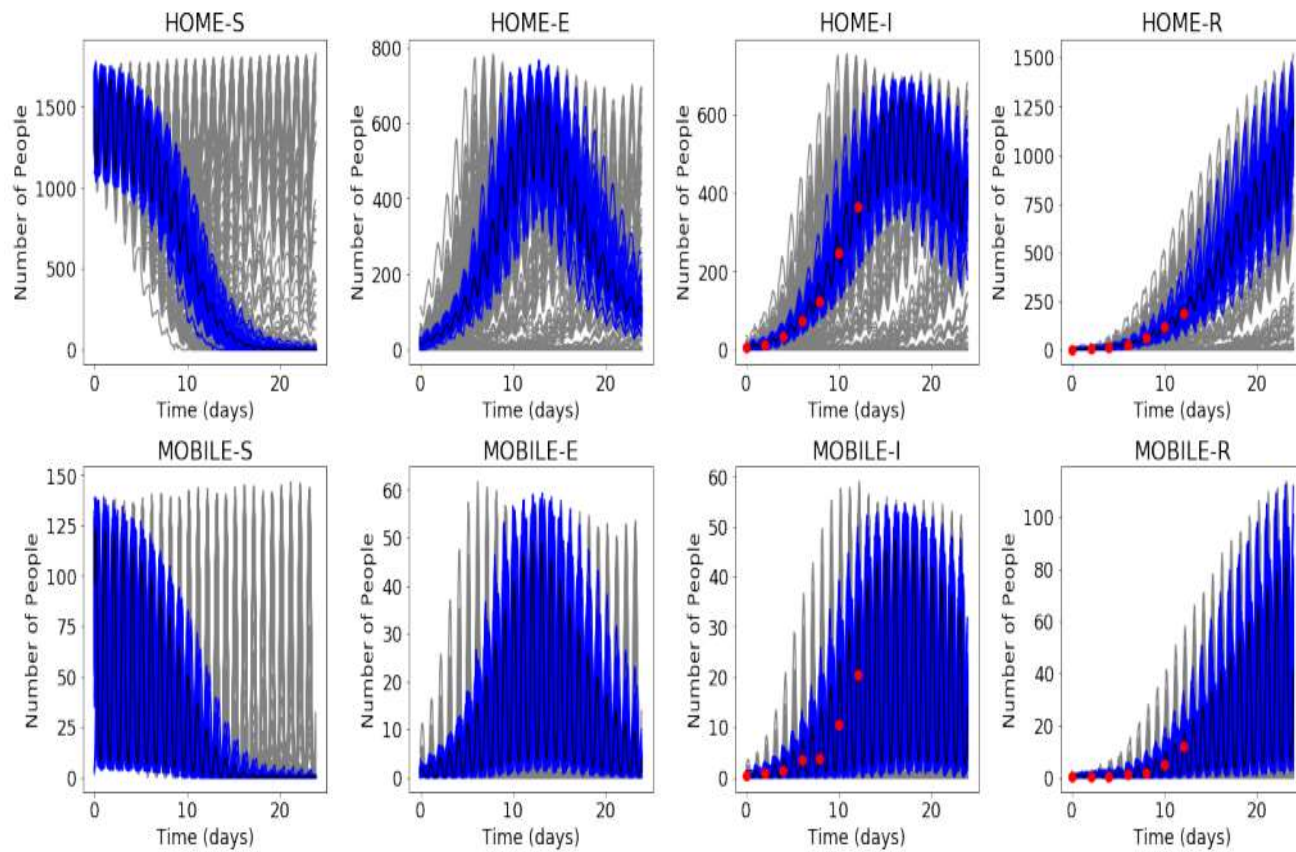
Ville idéalisée (100km × 100km) montrant les différentes régions. Les deux graphiques de droite montrent le nombre de personnes dans chaque cellule de la grille, pour une simulation et à deux moments de la journée.

### *3 Ensemble de données et processus de formation*

- 40 simulations numériques haute fidélité ont été réalisées afin de générer l'ensemble de données de formation
- nombre total est  $100 \times 8 = 800$  variables par pas de temps
- Après application de l'ACP, on aura 15 principaux composants capturent  $> 99.9999\%$  de la variance contenue dans les instantanés temporels.
- L'architecture GAN est basée sur DCGAN et implémentée à l'aide de 'TensorFlow'
- Nous entraînons le générateur et le discriminateur sur 5 000 epochs
- .Les réseaux reçoivent/génèrent les 10 niveaux de temps sous la forme d'un tableau bidimensionnel ("une image") avec 10 lignes et 17 colonnes.
- Chaque ligne représente un niveau de temps et chaque colonne comprend le 15 les coefficients PCA

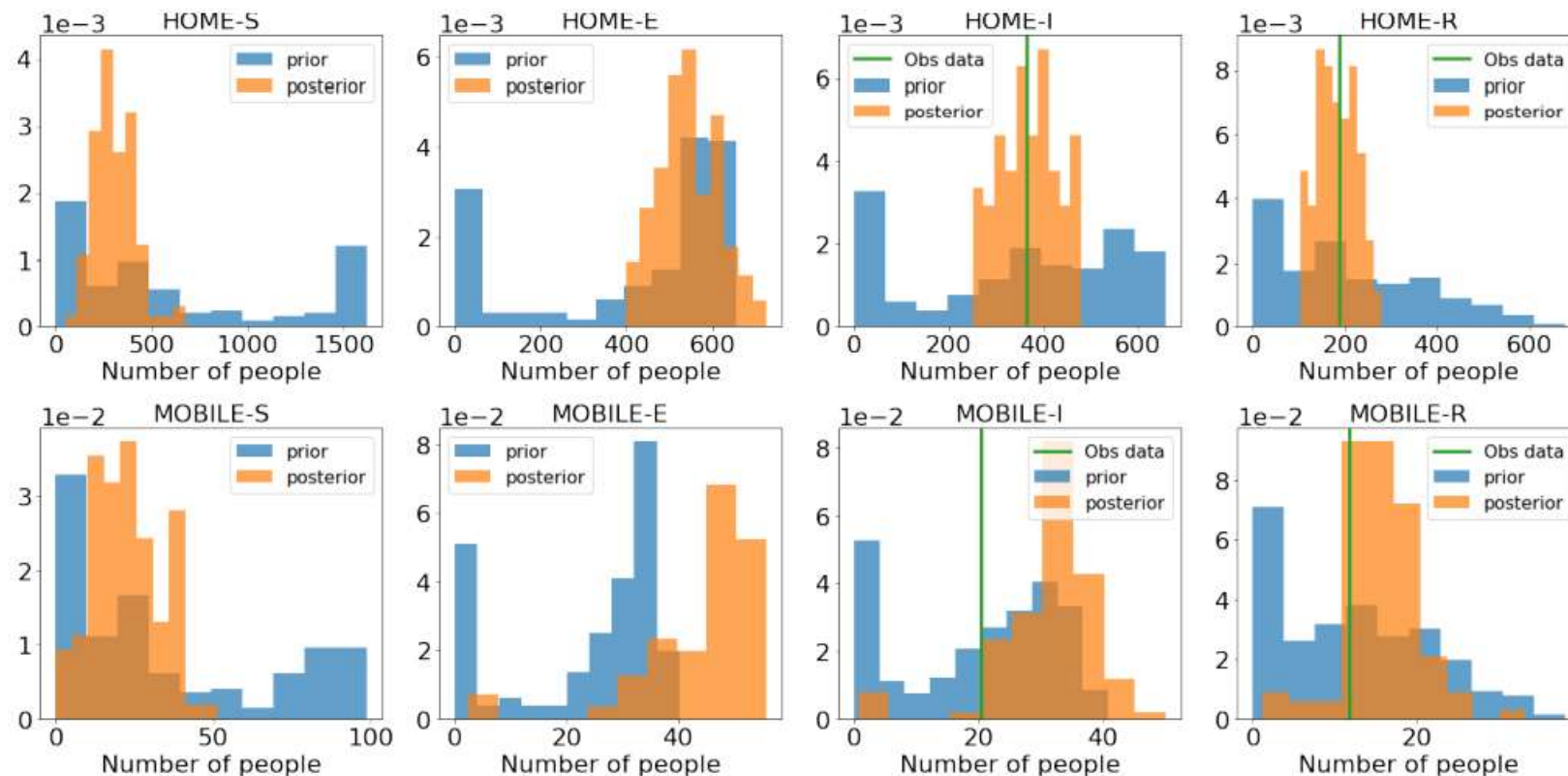


# *RÉSULTATS ET DISCUSSION*



## Résultats de l'UQ-PredGAN appliqués à la propagation du COVID-19 dans une ville idéalisée

- Dans chaque graphique les lignes grises les simulations inconditionnelles (avant assimilation des données), les lignes bleues les simulations conditionnelles (après assimilation des données) et la ligne noire la moyenne a posteriori.
- les points rouges représentent les données observées (mesures),



*Résultats de l'UQ-PredGAN appliqués à la propagation du COVID-19 dans une ville idéalisée*

- Fonction de densité de probabilité de chaque groupe et compartiment au jour 12.

# CONCLUSION

Dans ce travail, nous avons proposé une nouvelle utilisation d'un réseau contradictoire génératif (UQ-PredGAN) capable de quantifier l'incertitude dans les prédictions de séries chronologiques, compte tenu de la présence de mesures. L'objectif est de générer un modèle de substitution de la simulation numérique haute-fidélité, capable d'assimiler les données observées et de générer les incertitudes correspondantes. Nous avons appliqué la méthode proposée à un modèle SEIRS étendu qui représente la propagation du COVID-19 dans une ville idéalisée. Les résultats montrent que l'UQ-PredGAN correspond avec précision aux données observées et quantifie efficacement l'incertitude dans les états du modèle (groupes et compartiments) et les paramètres du modèle (nombres de reproduction de base).