

# Introduction

## Pouvons-nous prédire la courbe des anomalies de température globale en fonction de différents facteurs?

- données open data :



FAO ONU



kaggle

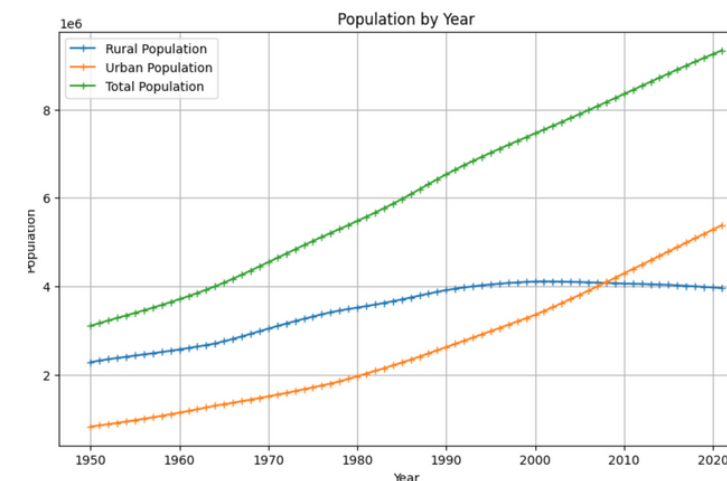


OMS



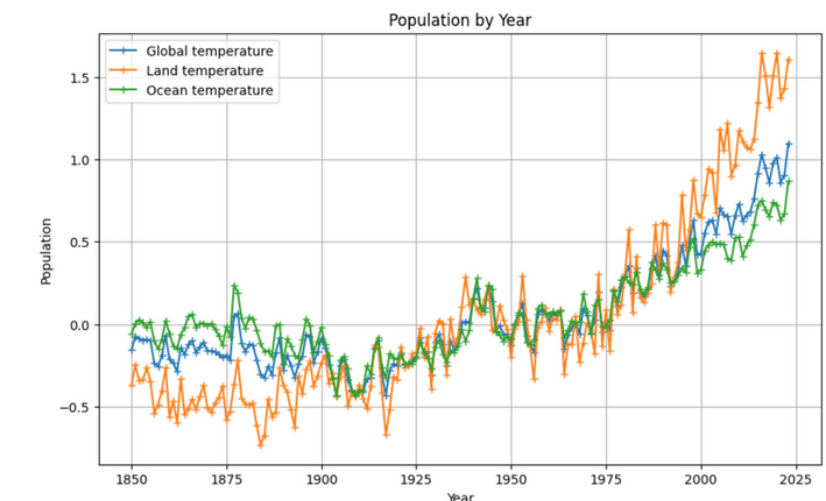
The World Bank

- facteurs démographiques : répartition rurale et urbaine.



- utilisation des terres et émissions de GHG dans l'agriculture
- émissions des GHG par les différents secteurs

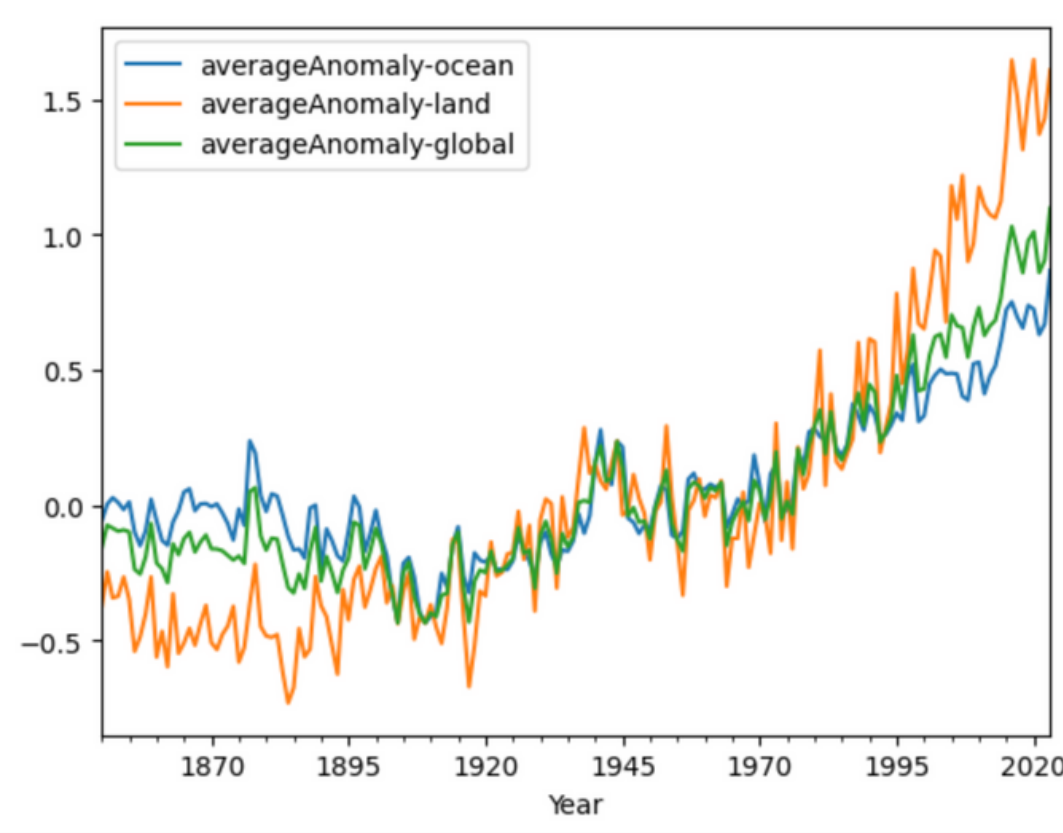
- évolution des températures : globale, terrestre, maritime



- évolution d'internet dans le monde
- production électrique par pays

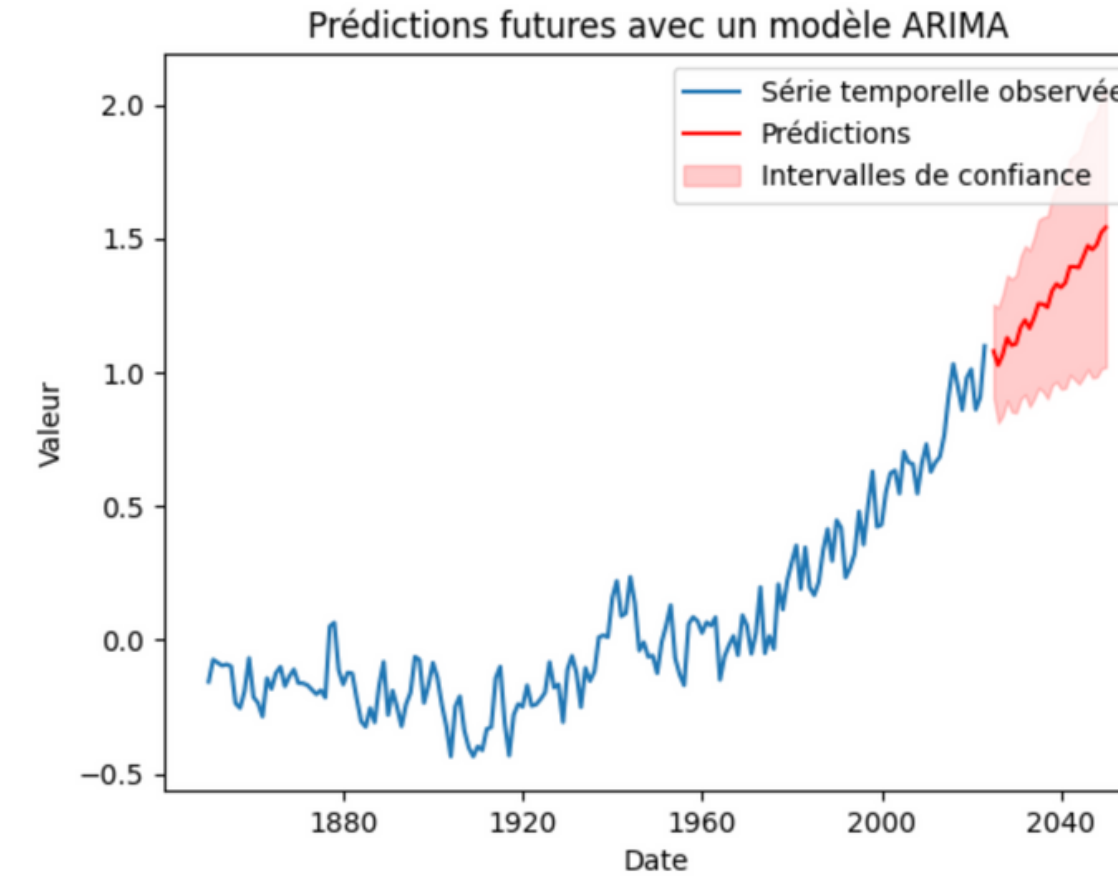
**But : extraire les données des différents facteurs pour créer notre dataframe et faire nos modèles dessus**

# Modèles préliminaires



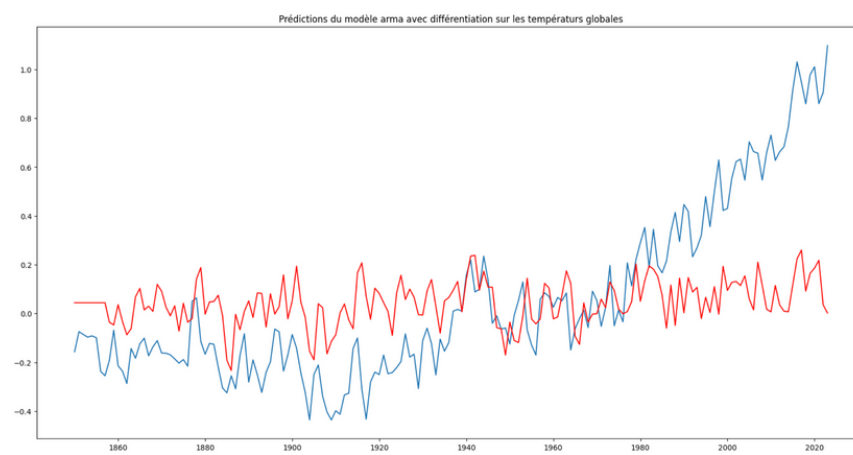
- test de Dickey Fuller

p-value adf : 0.997651554053703



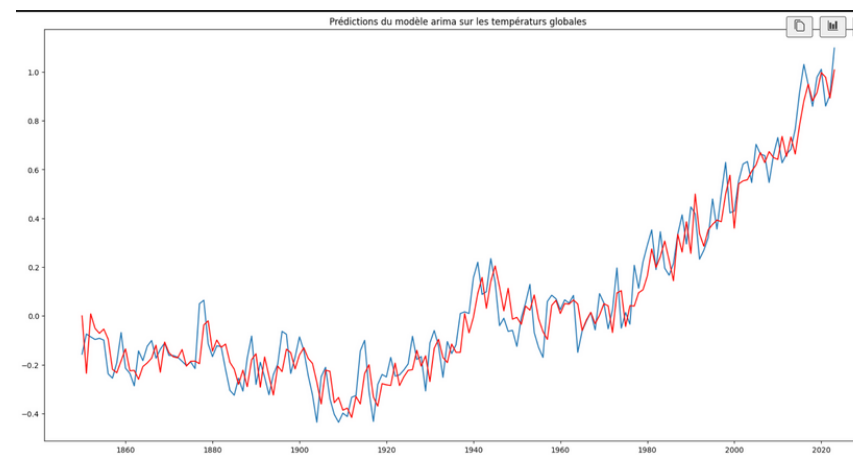
- forecast de 2023 à 2050

- ARMA

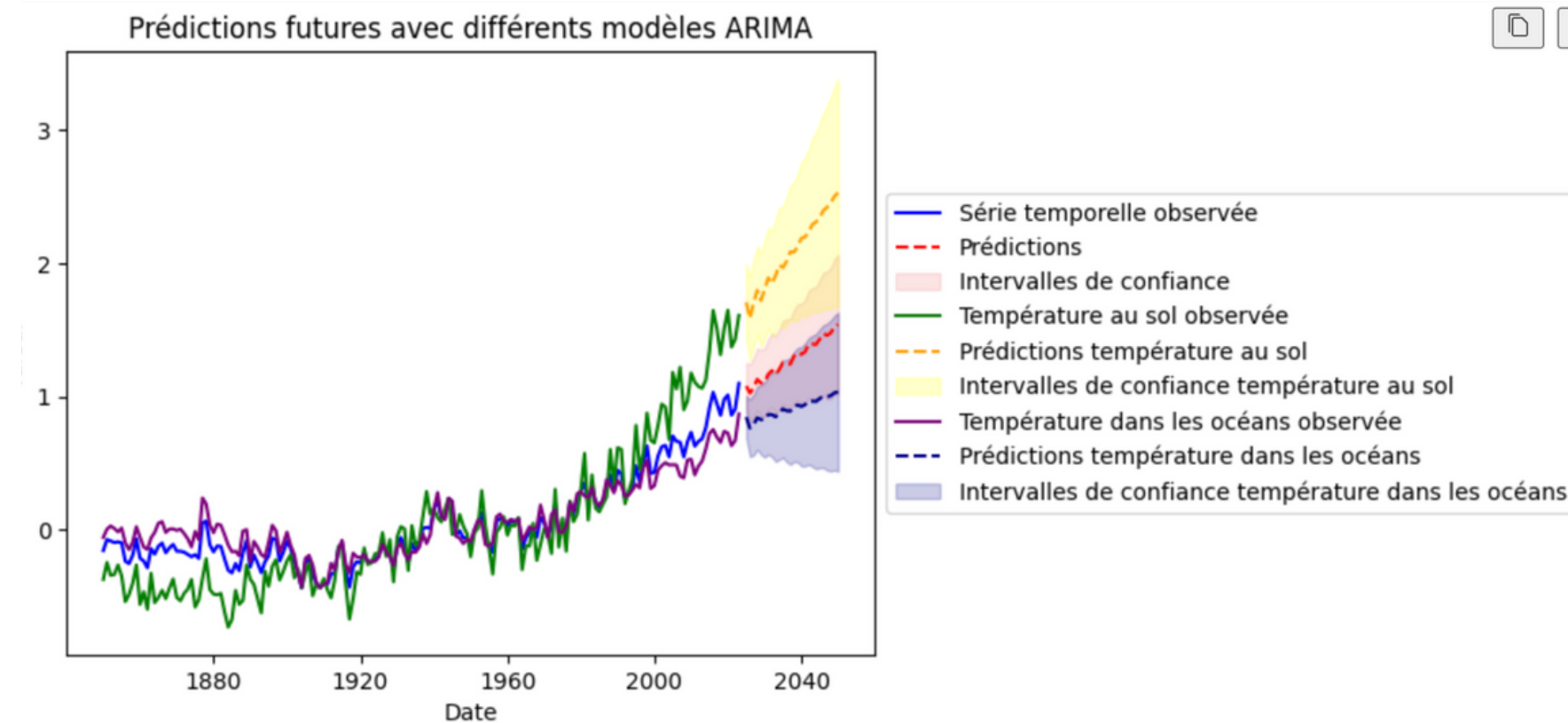


- choix (p,q)
- saisonnabilité
- différentiation

- ARIMA



- choix (p,d,q)



# Modèle ARIMAX

- modèle ARIMAX des températures globales

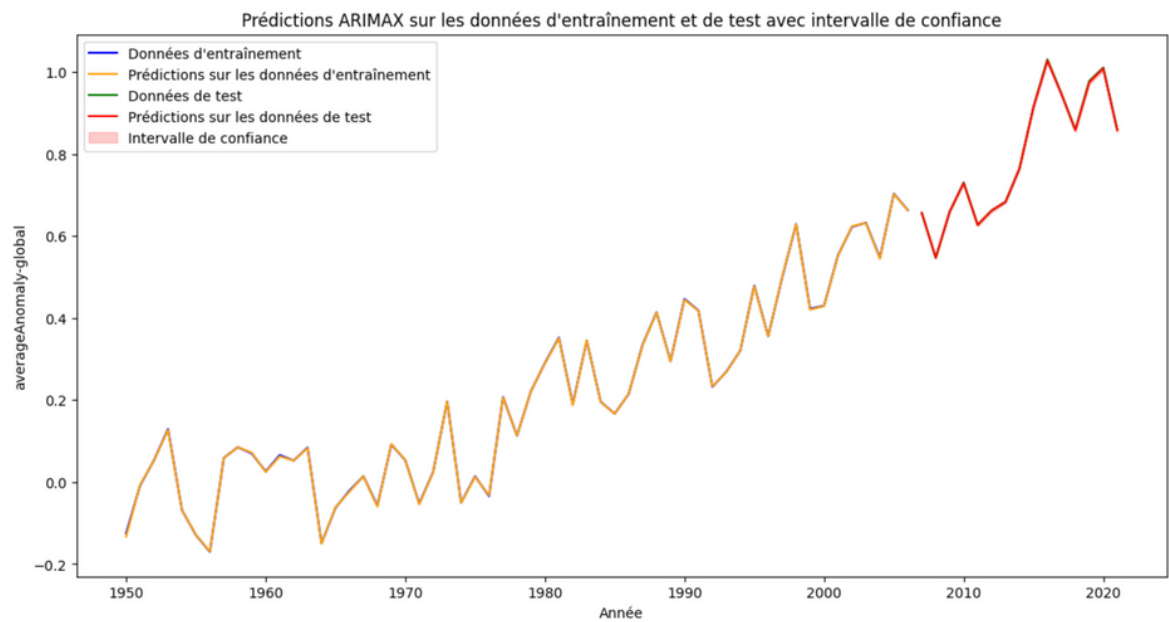
	averageAnomaly-ocean	averageAnomaly-land	averageAnomaly-global	Rural population	Urban population	Total population	log-rural-population	log-urban-population	log-total-population
Year									
1950-01-01	-0.090000	-0.202500	-0.125000	2280229.823	820258.014	3100487.837	2280229.823	820258.014	3100487.837
1951-01-01	0.000833	-0.023333	-0.008333	2314969.612	848615.115	3163584.727	2314969.612	848615.115	3163584.727
1952-01-01	0.072500	0.011667	0.053333	2346274.223	876864.771	3223138.994	2346274.223	876864.771	3223138.994
1953-01-01	0.054167	0.292500	0.129167	2374872.565	905829.315	3280701.880	2374872.565	905829.315	3280701.880
1954-01-01	-0.111667	0.027500	-0.068333	2401878.756	935690.068	3337568.824	2401878.756	935690.068	3337568.824
1955-01-01	-0.123333	-0.140000	-0.129167	2428203.147	966575.531	3394778.678	2428203.147	966575.531	3394778.678
1956-01-01	-0.095000	-0.333333	-0.170000	2454649.464	998455.413	3453104.877	2454649.464	998455.413	3453104.877
1957-01-01	0.095833	-0.020000	0.059167	2481796.367	1031267.971	3513064.338	2481796.367	1031267.971	3513064.338
1958-01-01	0.117500	0.016667	0.085000	2509646.174	1065298.708	3574944.882	2509646.174	1065298.708	3574944.882
1959-01-01	0.062500	0.096667	0.070000	2538426.784	1100422.902	3638849.686	2538426.784	1100422.902	3638849.686

feature endog : température global

- features exogènes :
  - températures maritimes et terrestres
  - population modiale, répartitions rurales et urbaines

anomalies détectées

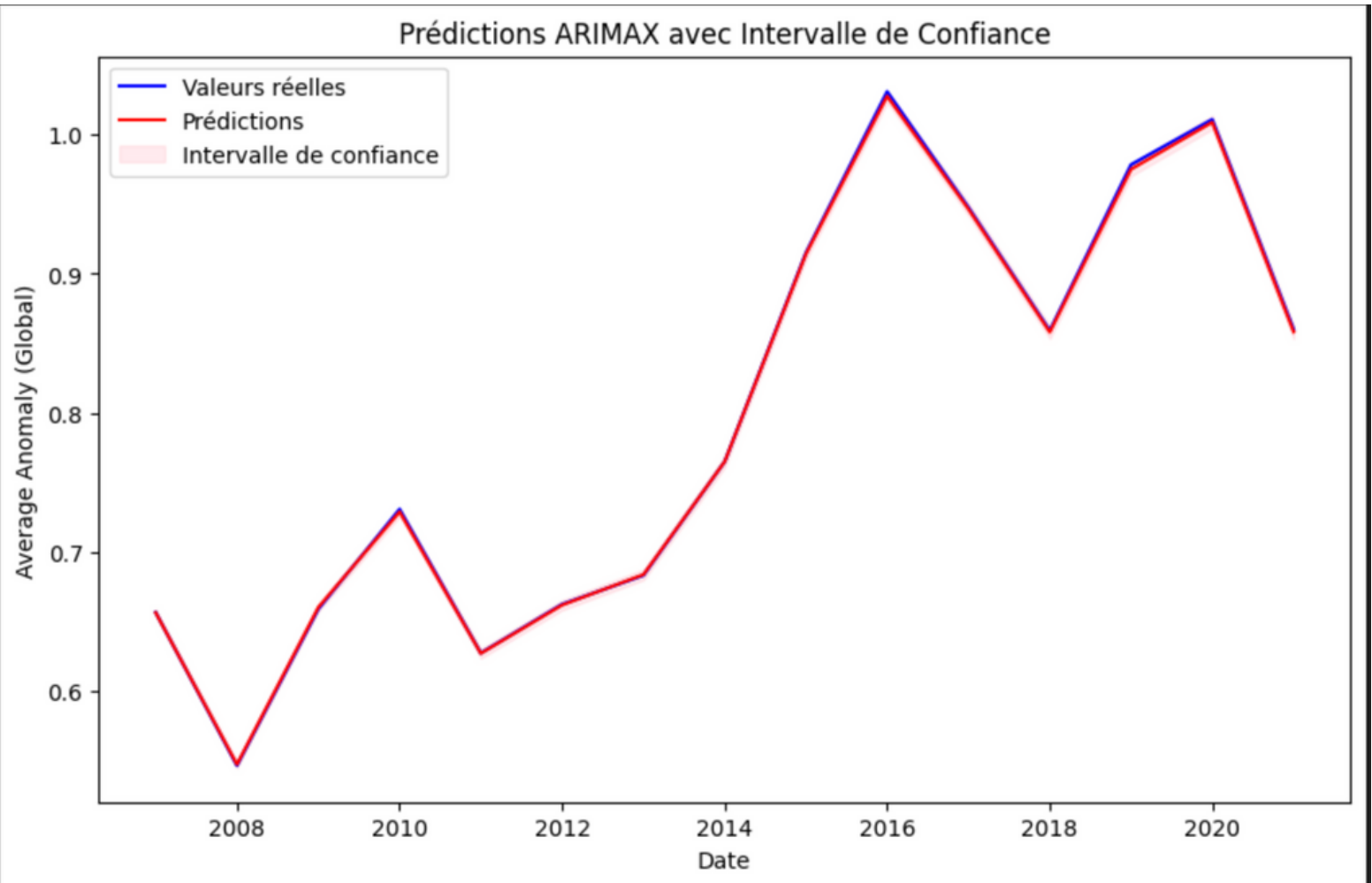
Apprentissage sur la base de train et predict



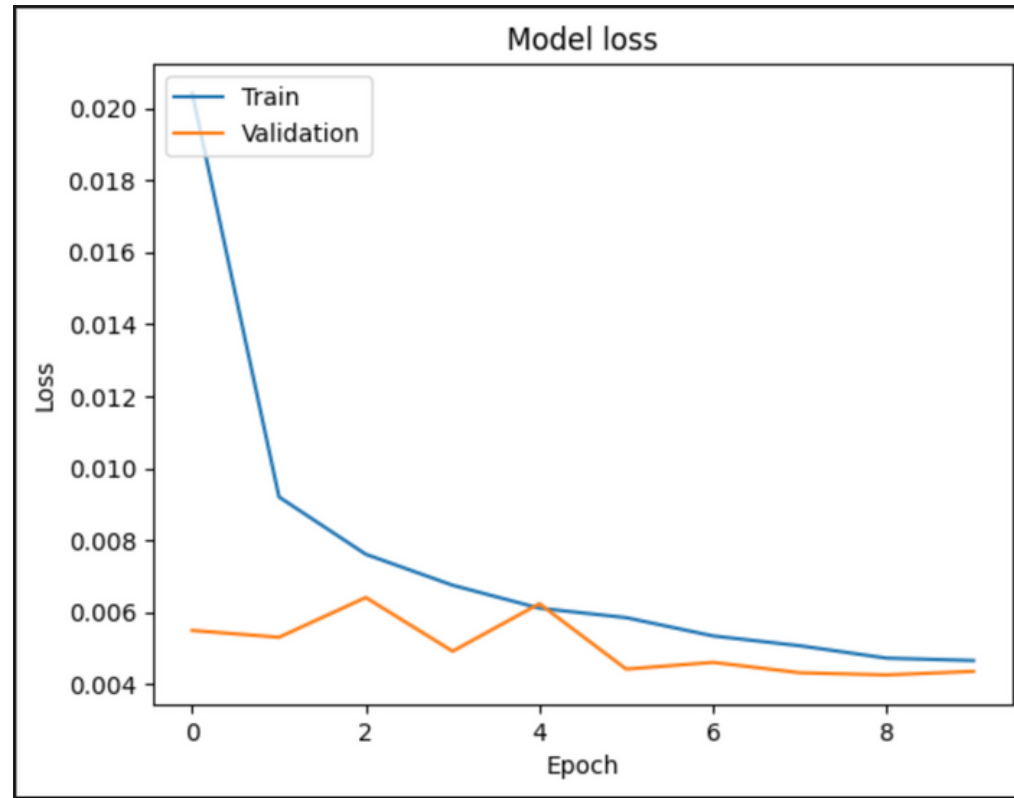
Validation croisée avec 10 folds

```
MSE pour la validation croisée 1: 2.4701313764352248e-06
MSE pour la validation croisée 2: 3.0308798790034885e-06
MSE pour la validation croisée 3: 8.074694579963697e-06
MSE pour la validation croisée 4: 5.908049263741429e-07
MSE pour la validation croisée 5: 8.974095967170234e-07
MSE pour la validation croisée 6: 1.3162633692646019e-06
MSE pour la validation croisée 7: 1.3363098502151137e-06
MSE pour la validation croisée 8: 9.658294297167288e-07
MSE pour la validation croisée 9: 2.3479028473180683e-06
MSE pour la validation croisée 10: 6.755843596345773e-06
Moyenne des MSE : 2.7786069451353864e-06
```

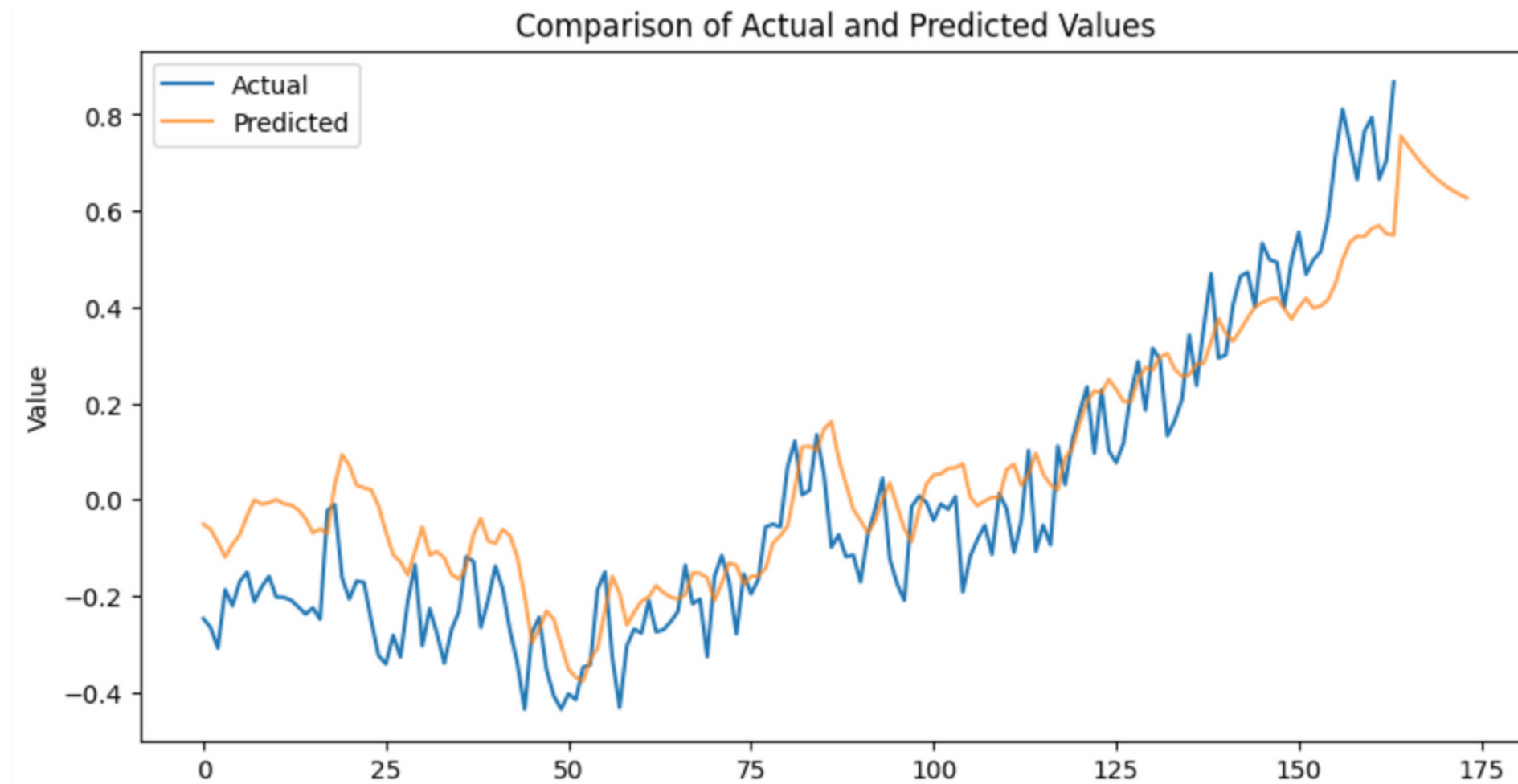
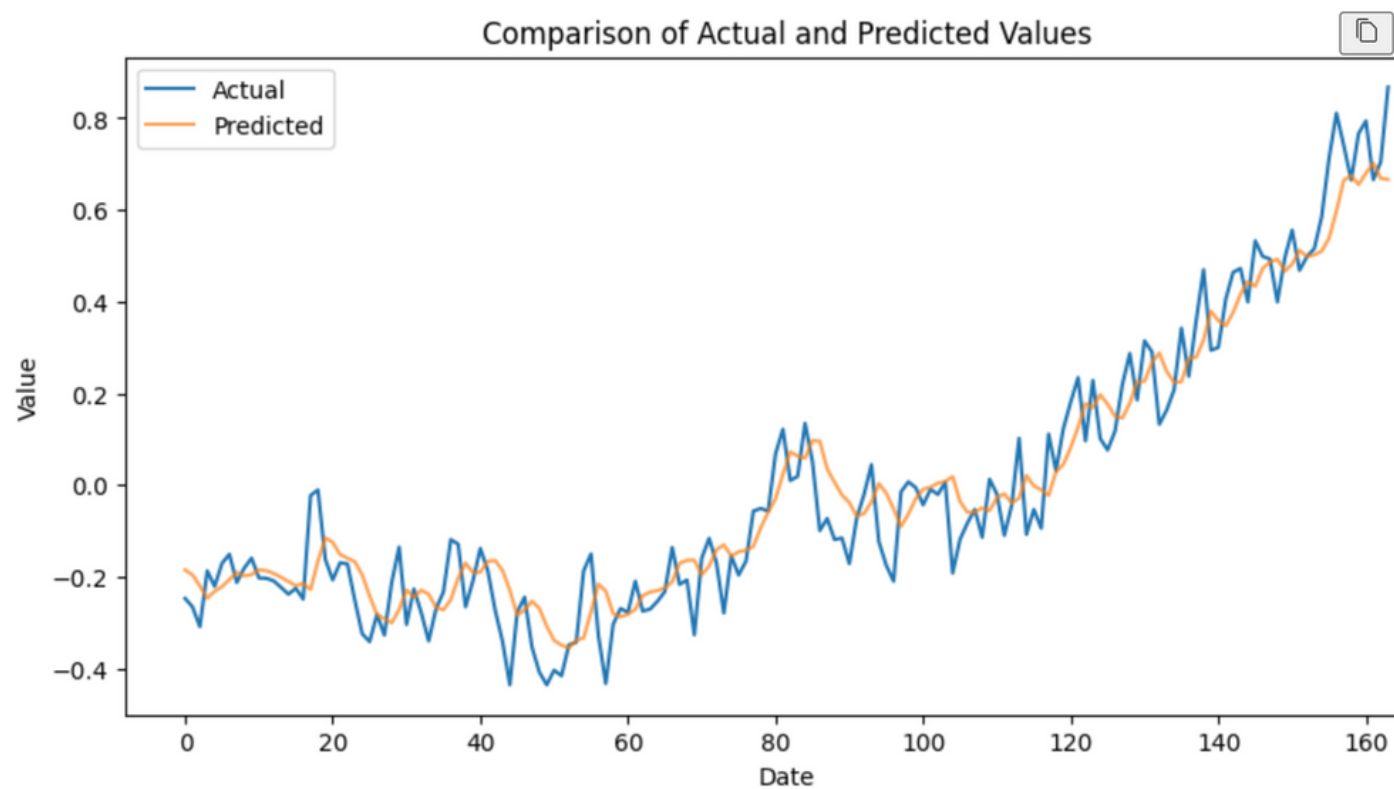
Predict sur les tests



# Modèle RNN



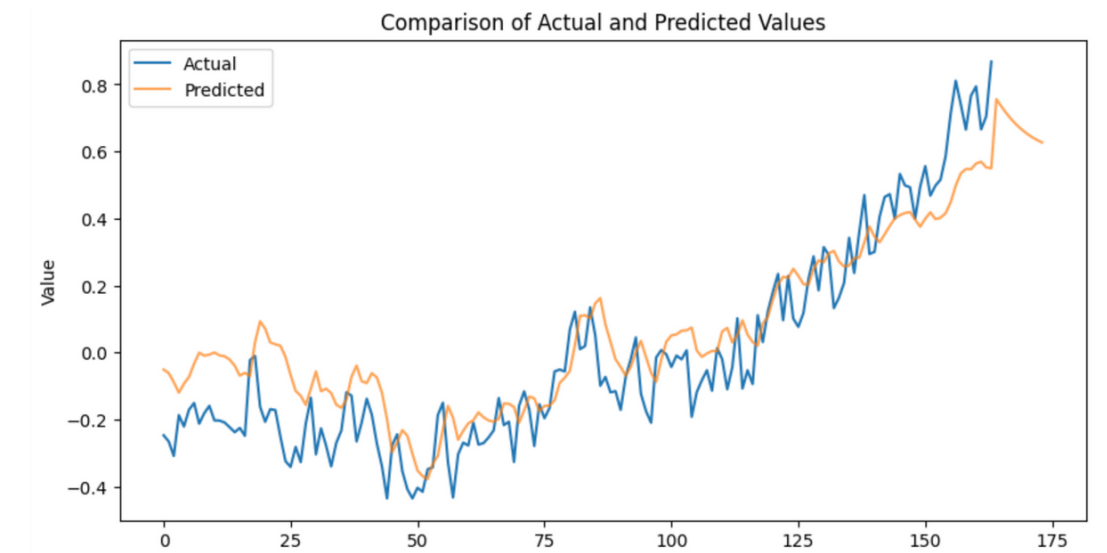
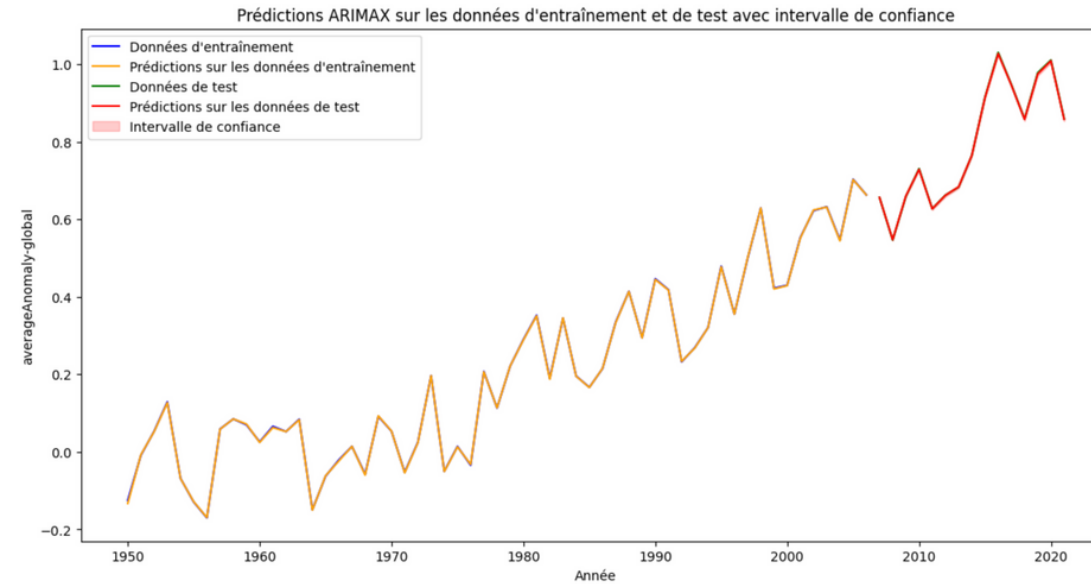
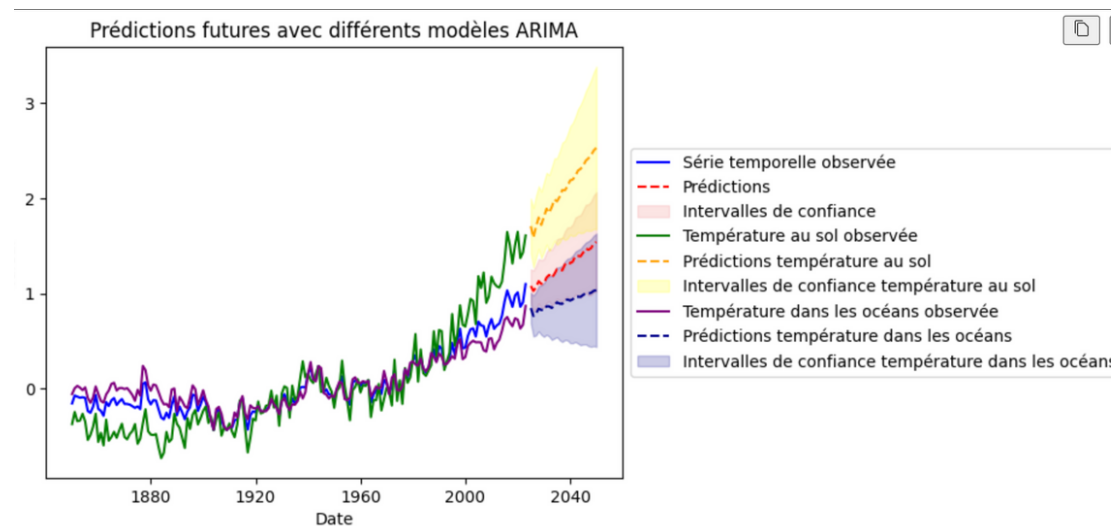
**MSE: 0.007583997455746853, MAE: 0.06960163827044598**





# Conclusion et perspectives

## Comparaison des modèles



## Prise de recul et perspectives

### Projet trop complexe ?

- Jeux de données pas toujours trouvables
- dépendance aux variables et aux jeux de données : différentes origines et plages de données

### Méthode de travail pas adaptée



**Peut-on répondre à la problématique ?  
Quelles conclusions pour le futur ?**