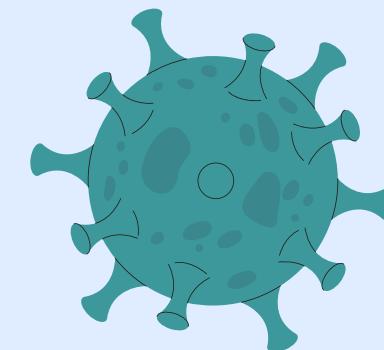


# Projet Covid-19

## Étude exploratoire d'estimation du taux de reproduction de la Covid-19

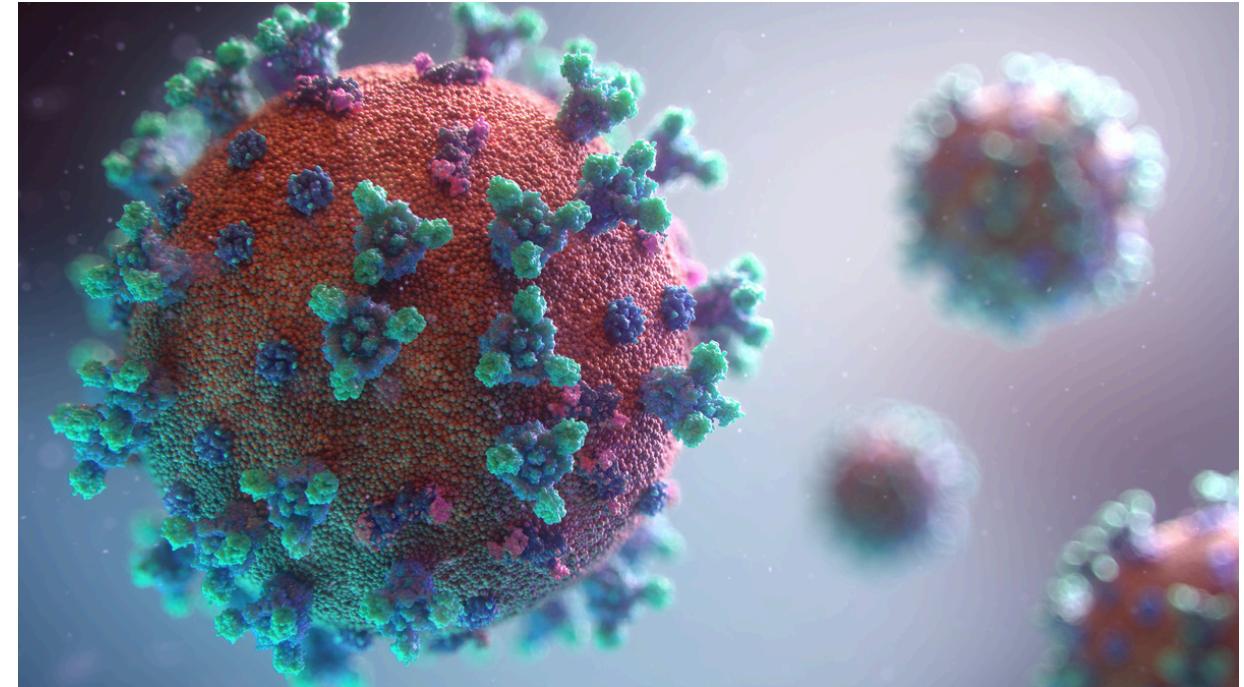


**Emma Greverie**  
**Lucas Tramonte**  
**Romain Dalcant**  
**Jia Yi Ang**



# Quid du R0 ?

Le R0, aussi appelé **taux de reproduction de base**, est une mesure utilisée pour évaluer la propagation d'une maladie infectieuse. Il représente le **nombre moyen de personnes** qu'une personne infectée peut contaminer.



Ce dernier aspect est utile aux autorités sanitaires car cela leur donne la possibilité de **planifier** et de mettre en place des **mesures de contrôle appropriées**, telles que des campagnes de vaccination, des confinements ou des mesures de distanciation sociale.

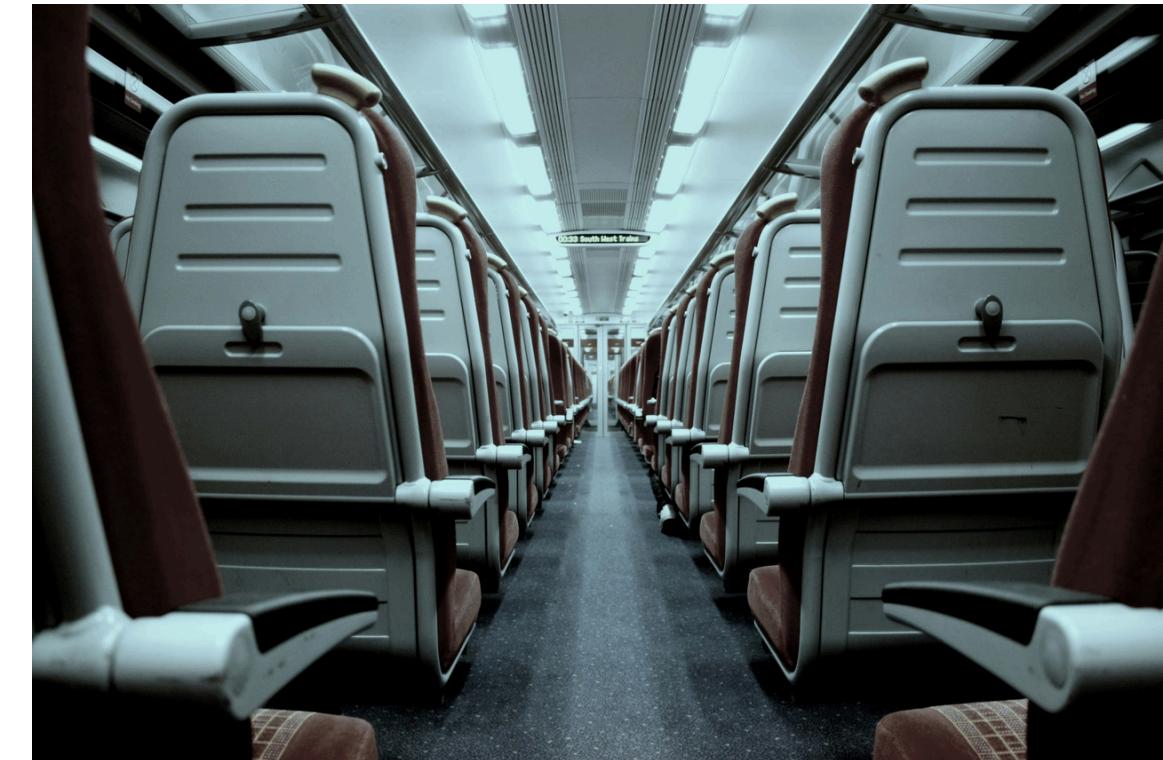
Le R0 permet à la fois d'évaluer la transmission du Covid-19 en donnant des informations sur sa **vitesse de propagation** dans une population donnée mais aussi de **prédir l'évolution de la maladie**.



# La suite du Projet démarré en S7

**Au semestre précédent, notre projet consistait à utiliser différents modèles afin d'obtenir le R0 de l'Ile de France.**

Dans cette première approche, déterminer le R0 de l'Ile de France se limitait alors à **considérer l'Ile de France comme une région isolée**, ce qui dans les faits, est loin d'être pertinent.



**Ce semestre, nous avons voulu élargir notre démarche aux départements à proximité de l'Ile de France.**

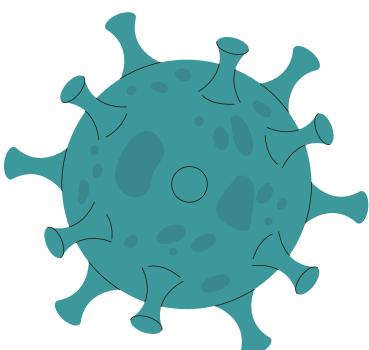
Ainsi, nous avons dû modifier nos modèles existants afin de tenir compte des interactions qui existent avec les différentes régions voisines.

# Sommaire

**I/ Base de données & Nouvelle Géographie Étudiée**

**II/ Modèle SIR**

**III/ Modèle Autorégressif & Lissage Spatial**



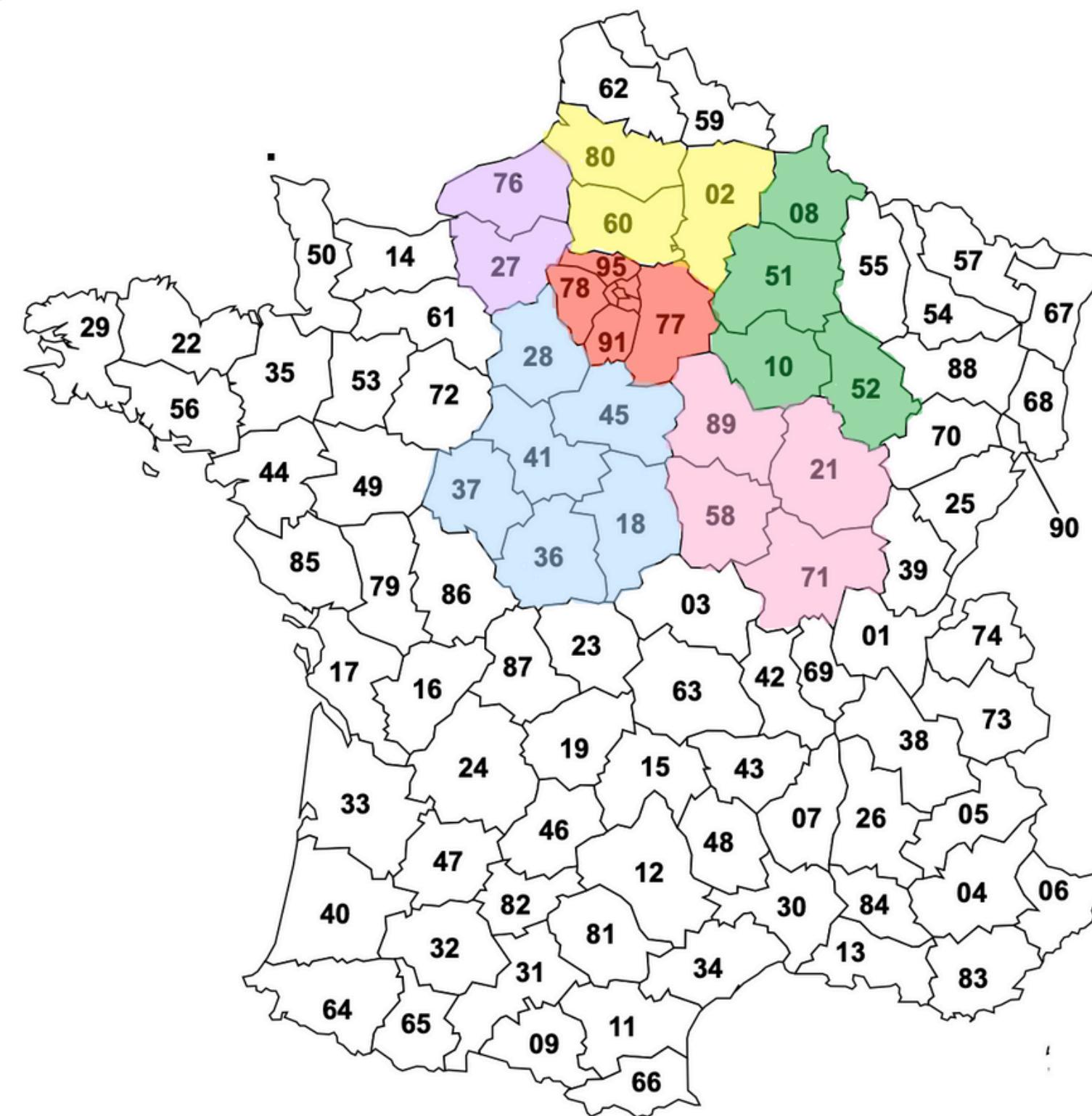
**IV/ Résultats Obtenus**

**V/ Conclusion & Perspectives**

# I/ Base de Données & Nouvelle Géographie Étudiée



# 1) Nouvelle géographie étudiée



Région Jaune : Aisne (02) + Oise (60) + Somme (80)

Région Rose : Côte-d'Or (21) + Nièvre (58) + Saône-et-Loire (71) + Yonne (89)

Région Bleue : Cher (18) + Eure-et-Loir (28) + Indre (36) + Indre-et-Loire (37) + Loir-et-Cher (41) + Loiret (45)

Région Verte : Ardennes (08) + Aube (10) + Marne (51) + Haute-Marne (52)

Région Violette : Eure (27) + Seine-Maritime (76)

Région Rouge : Paris (75) + Seine-et-Marne (77) + Yvelines (78) + Essonne (91) + Hauts-de-Seine (92) + Seine-Saint-Denis (93)

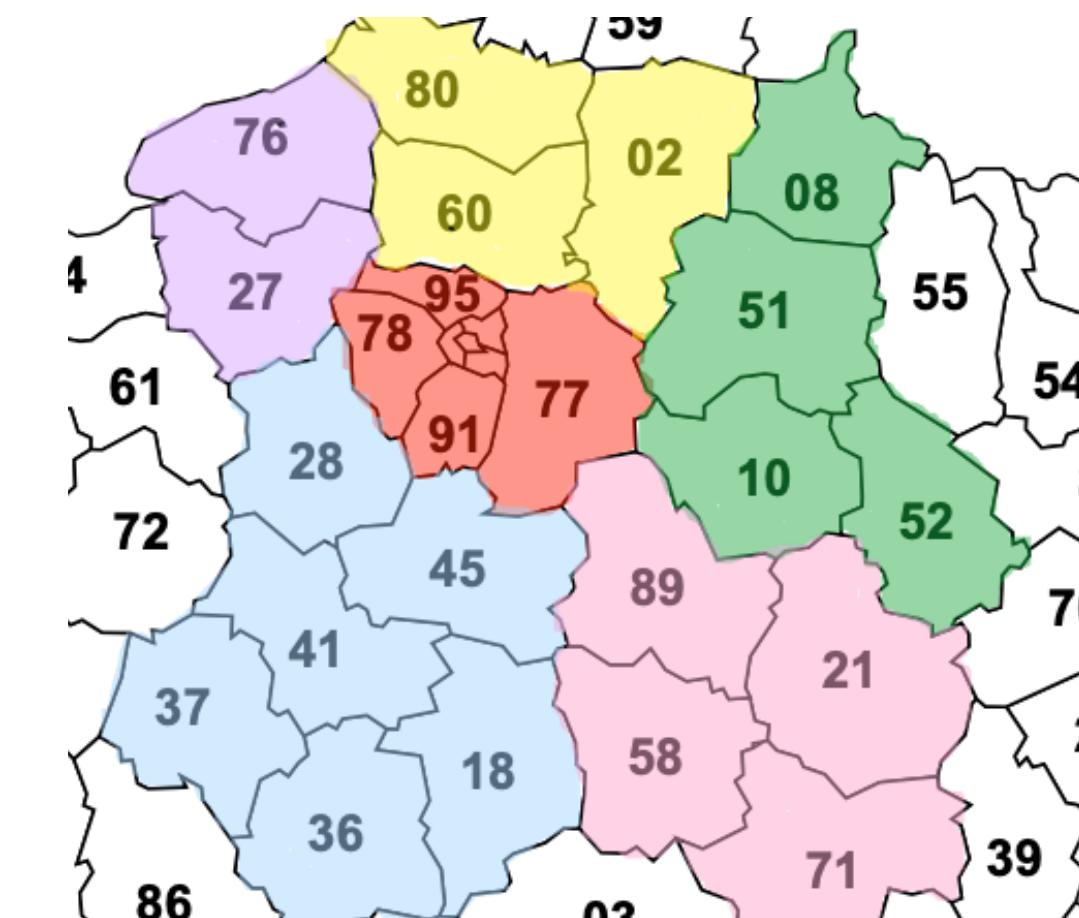
# 1) Nouvelle géographie étudiée

## Distance entre les différentes régions

196.7 km (Bourgogne - Centre)  
183.5 km (Bourgogne - Champagne)  
335.9 km (Bourgogne - Haute Normandie)  
227.2 km (Bourgogne - Île de France)  
294.8 km (Bourgogne - Picardie)  
258.0 km (Centre - Champagne)  
208.6 km (Centre - Haute Normandie)  
151.5 km (Centre - Île de France)  
250.0 km (Centre - Picardie)  
259.9 km (Champagne - Haute Normandie)  
160.0 km (Champagne - Île de France)  
159.2 km (Champagne - Picardie)  
112.3 km (Haute Normandie - Île de France)  
126.1 km (Haute Normandie - Picardie)  
98.6 km (Île de France - Picardie)

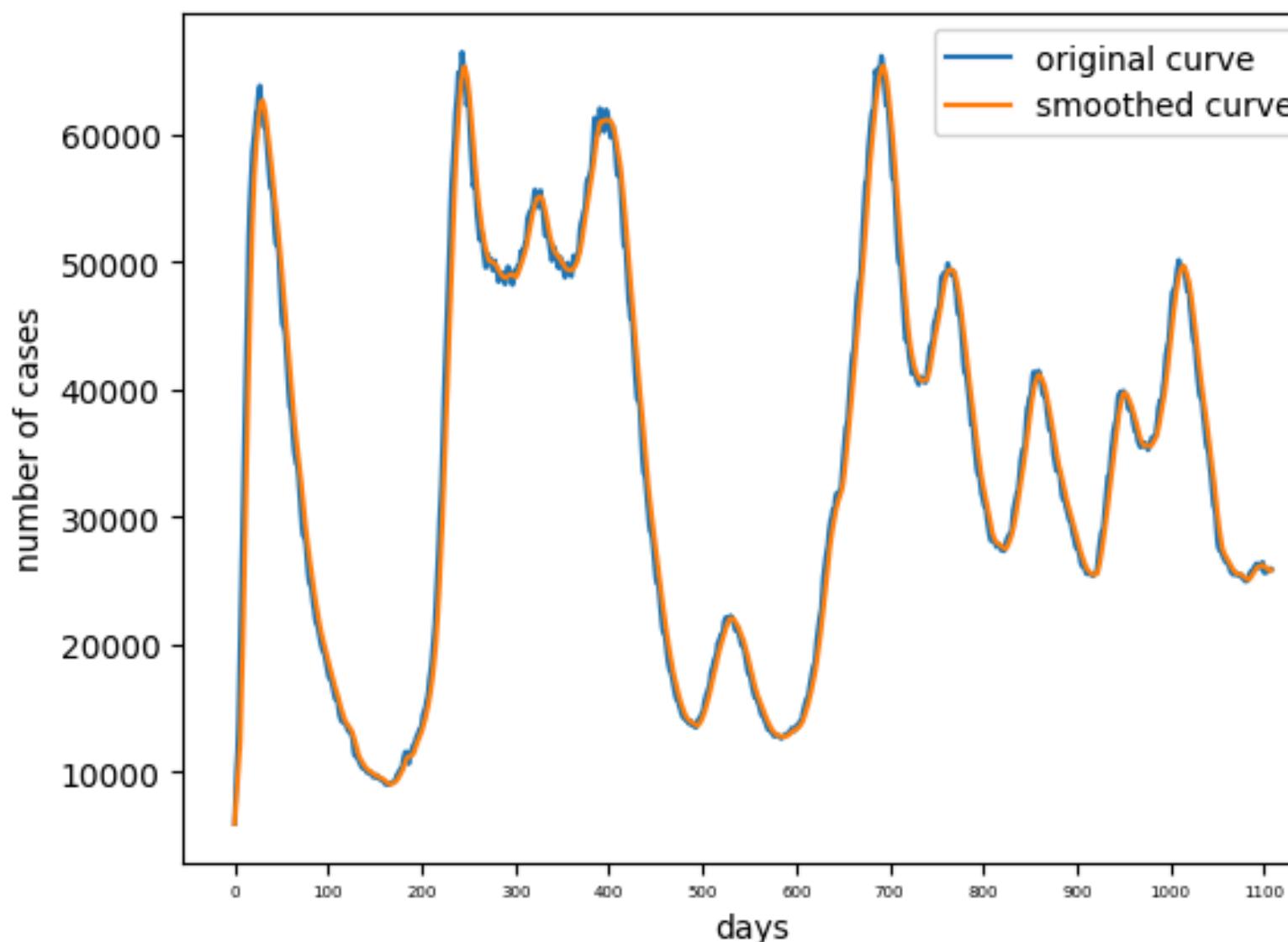
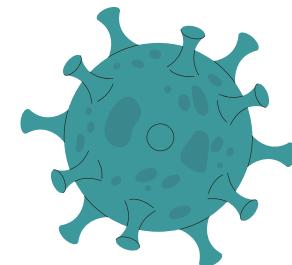
## Population par région (en millions d'habitants)

Région Jaune - « Picardie » : 1.575  
Région Rose - « Bourgogne » : 1.624  
Région Verte - « Champagne » : 1.316  
Région Bleue - « Centre » : 2.573  
Région Violette - « Haute Normandie » : 1.847  
Région Rouge - « Île de France » : 11.451



## 2) Base de données

Il s'agit d'une base de données contenant le nombre de cas par jour et par département issue de Data.gouv



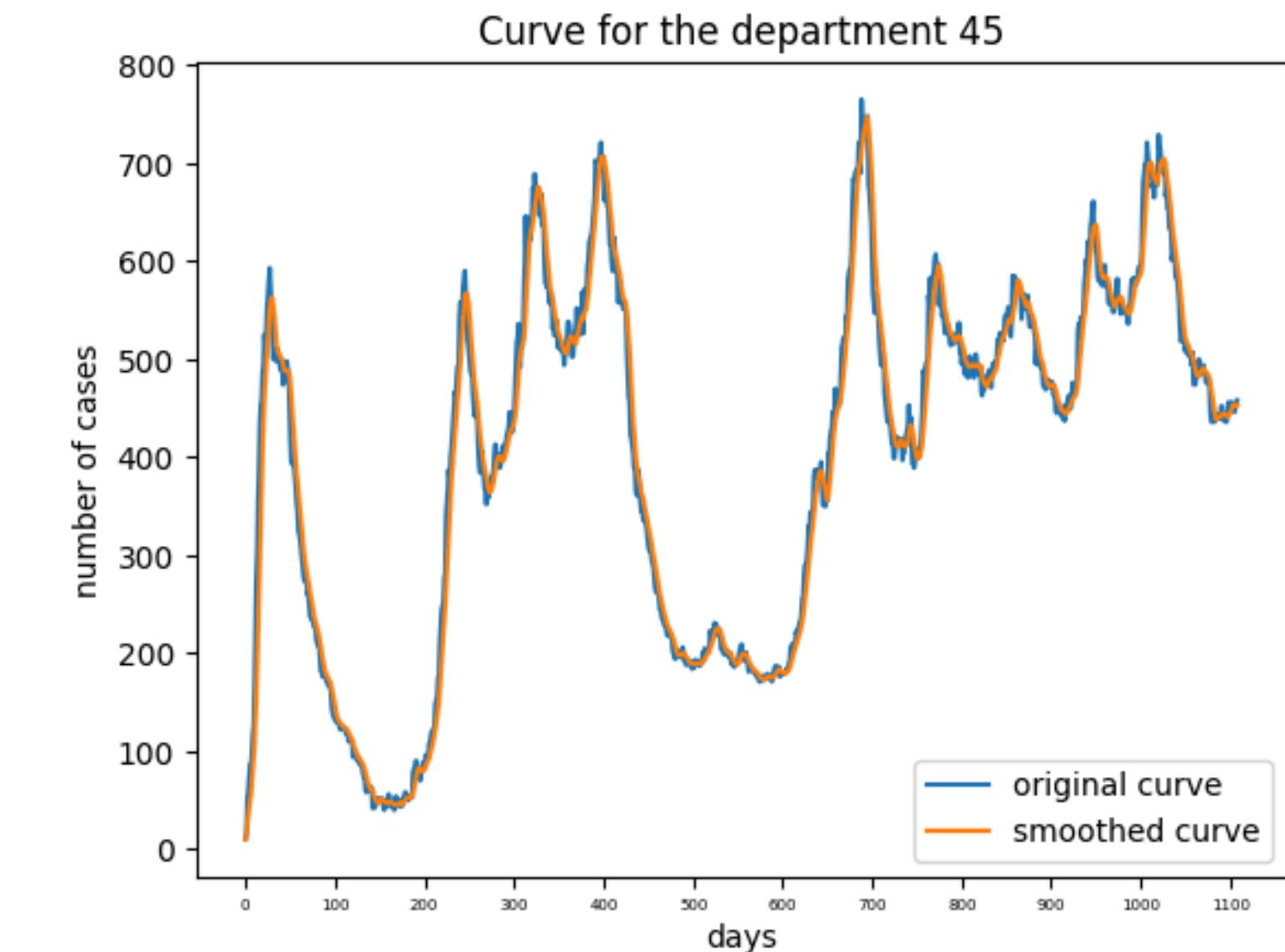
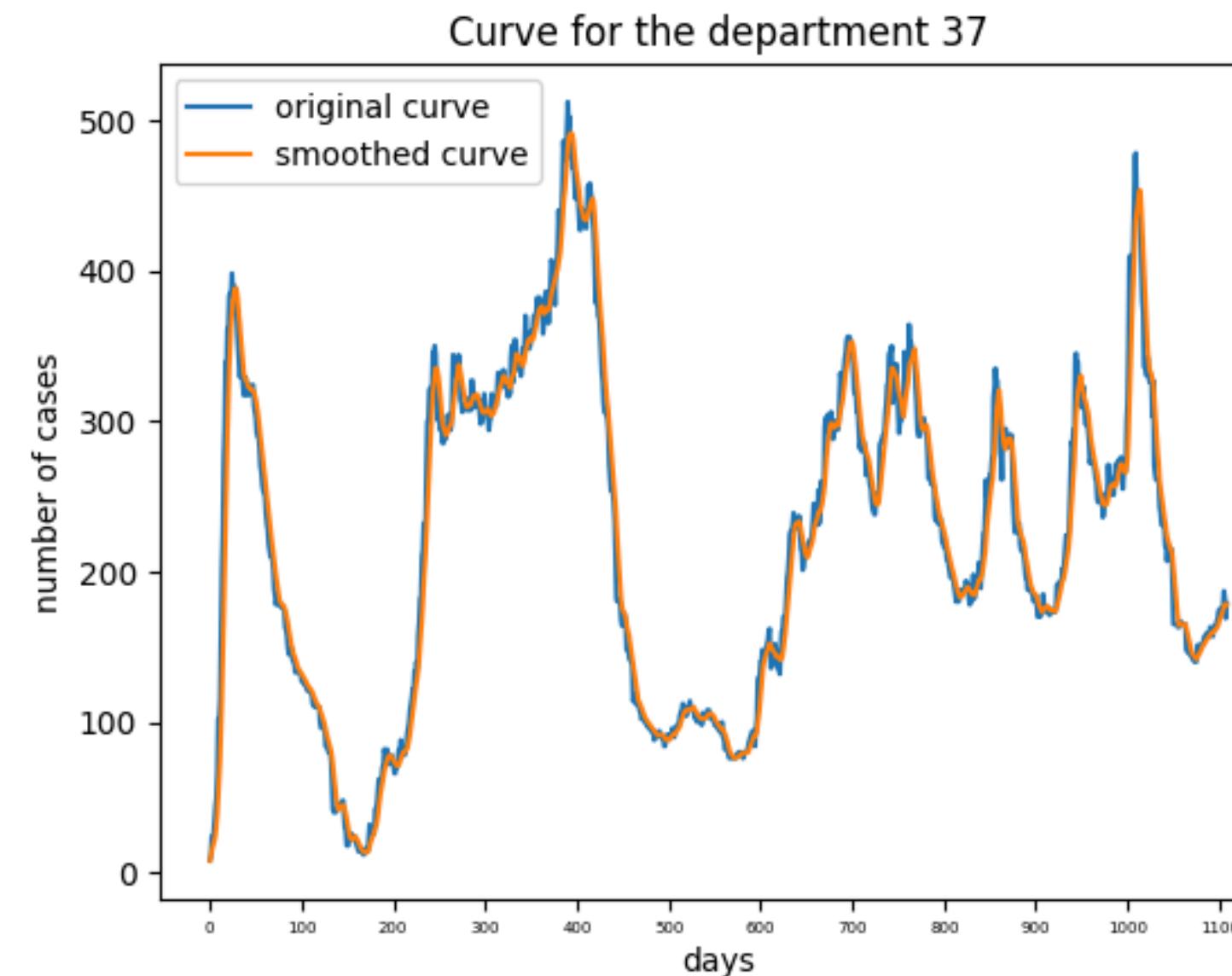
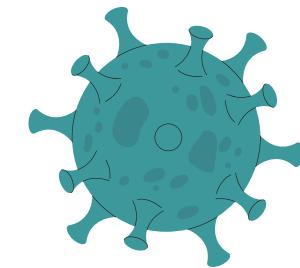
Nombre de cas par jour pour tous les départements de la nouvelle géographie étudiée.

Comparaison entre :

- La courbe réelle
- La courbe lissée obtenue à partir de la méthode de la moyenne mobile

## 2) Base de données - Pour chaque département

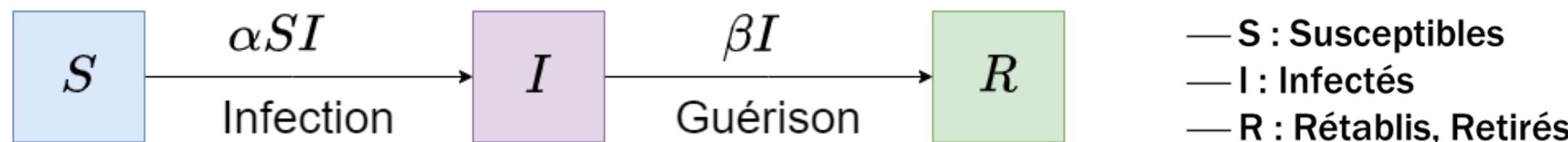
Il s'agit d'une base de données contenant le nombre de cas par jour et par département issue de Data.gouv



## III/ Modèle SIR



# Modèle SIR



α : Taux de contamination  
β : Taux de guérison

## Équations différentielles du modèle SIR

$$\frac{dS}{dt} = -\alpha \cdot S \cdot I$$

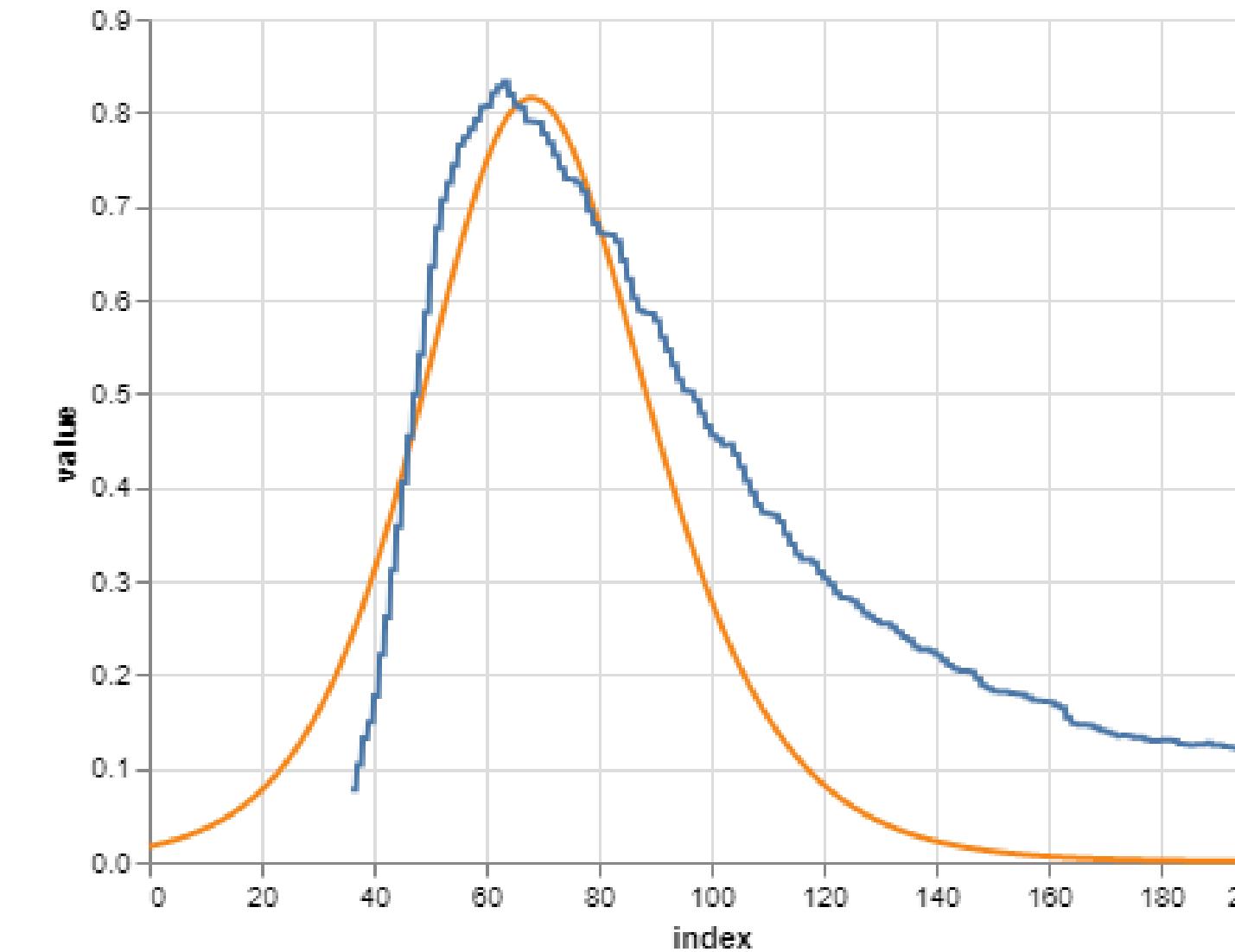
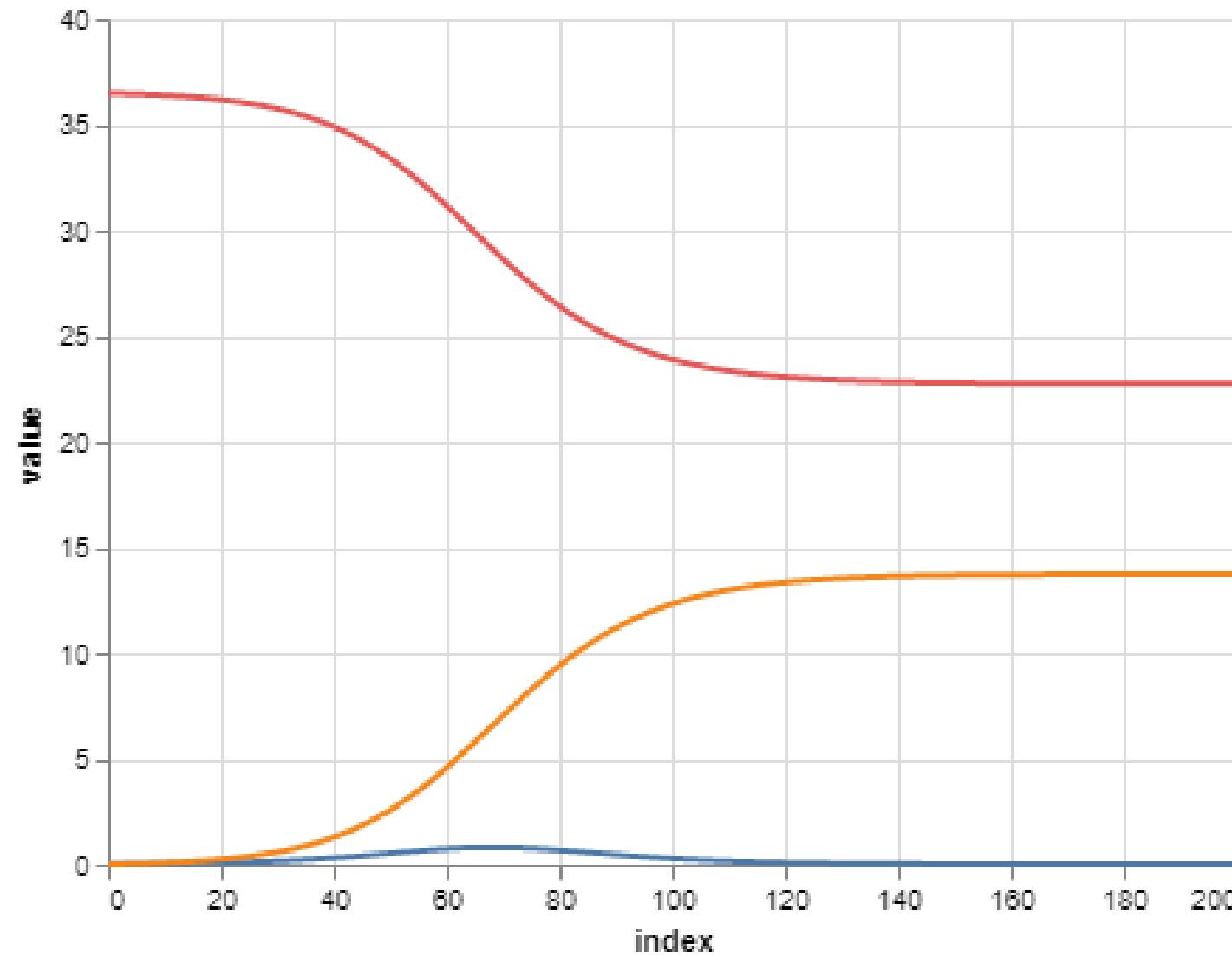
$$\frac{dI}{dt} = \alpha \cdot S \cdot I - \beta \cdot I$$

$$\frac{dR}{dt} = \beta \cdot I$$

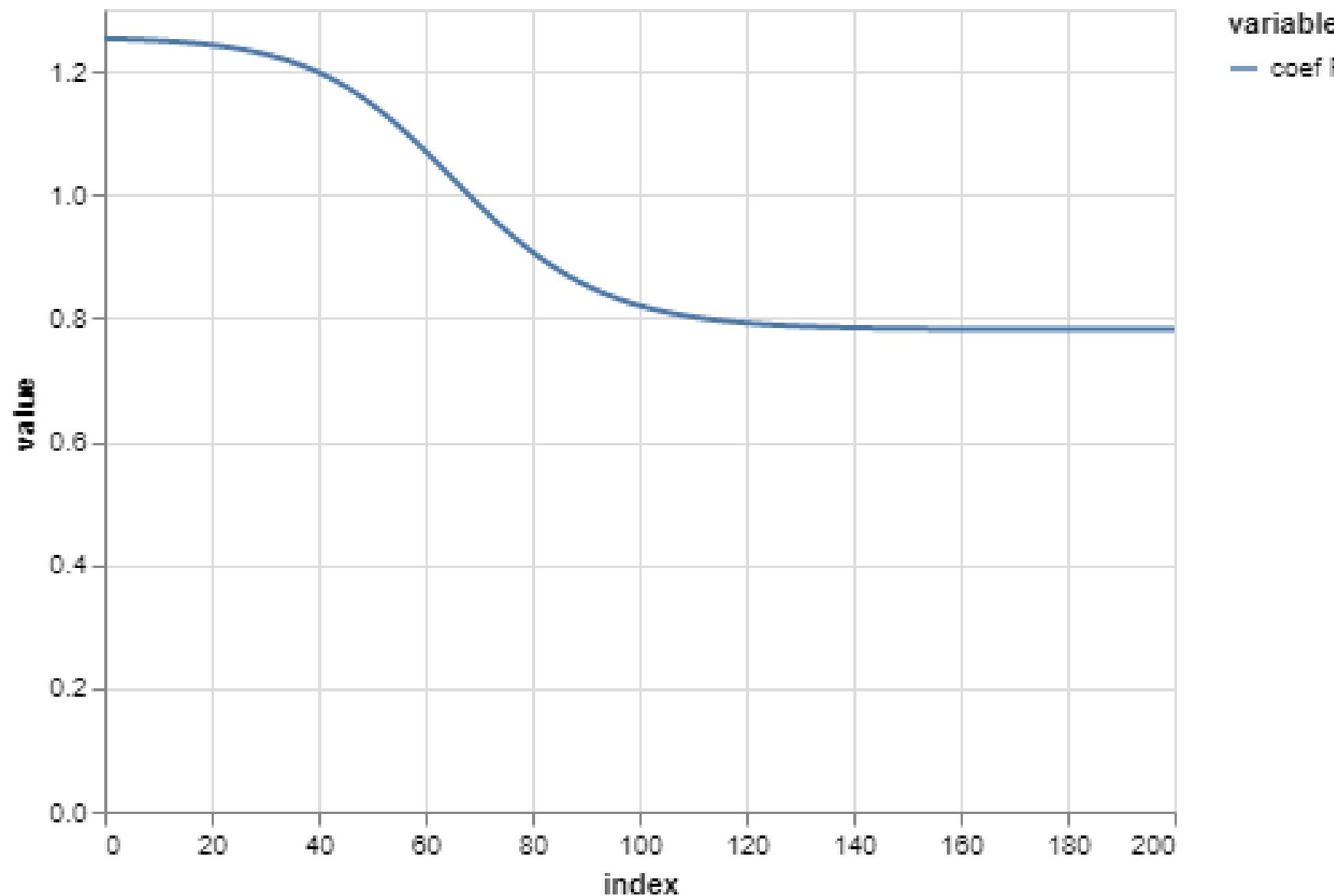
Estimation du coefficient de reproduction  $R_0$

$$R_0 = \frac{S\alpha}{\beta}$$

# Modèle SIR - Tous les départements

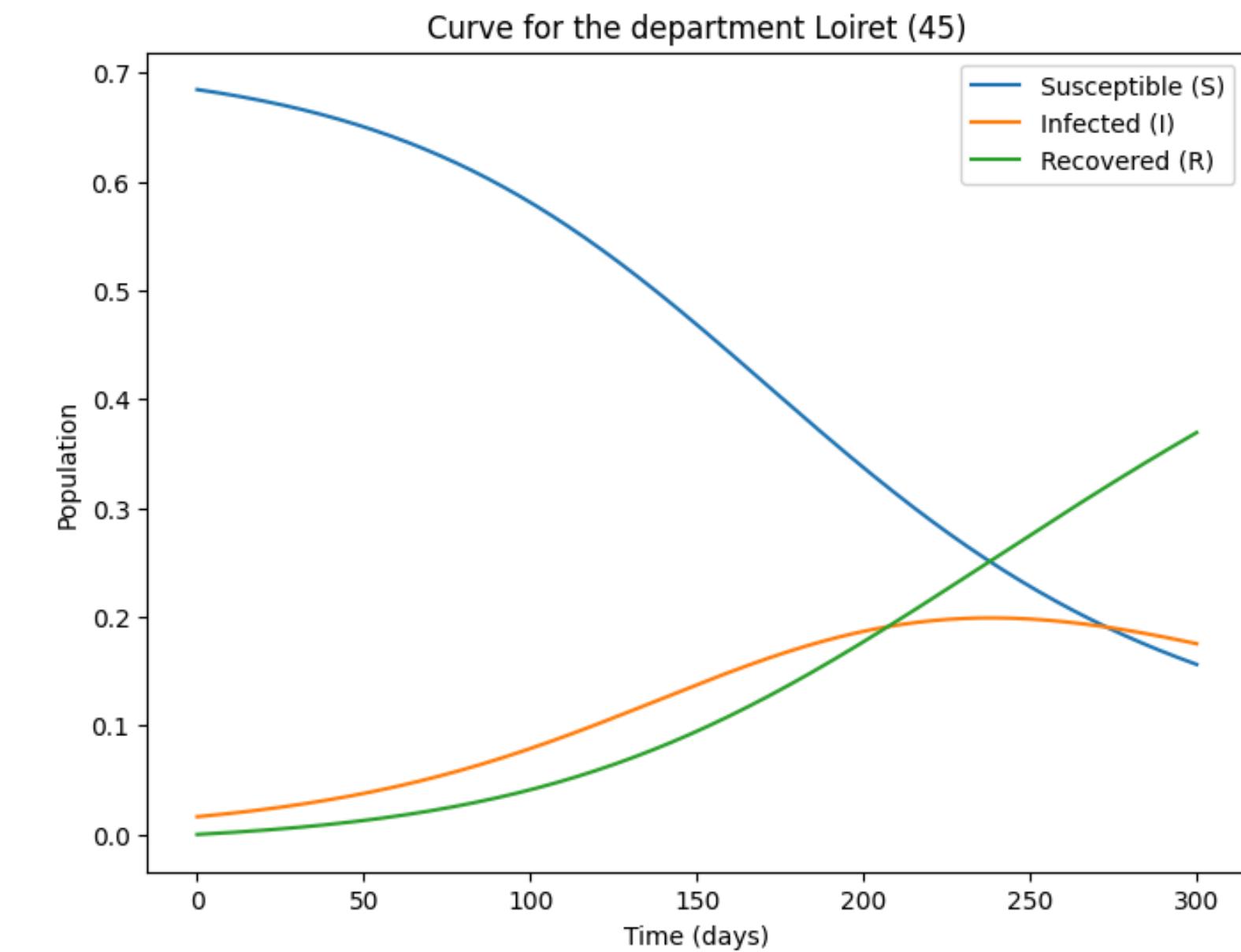
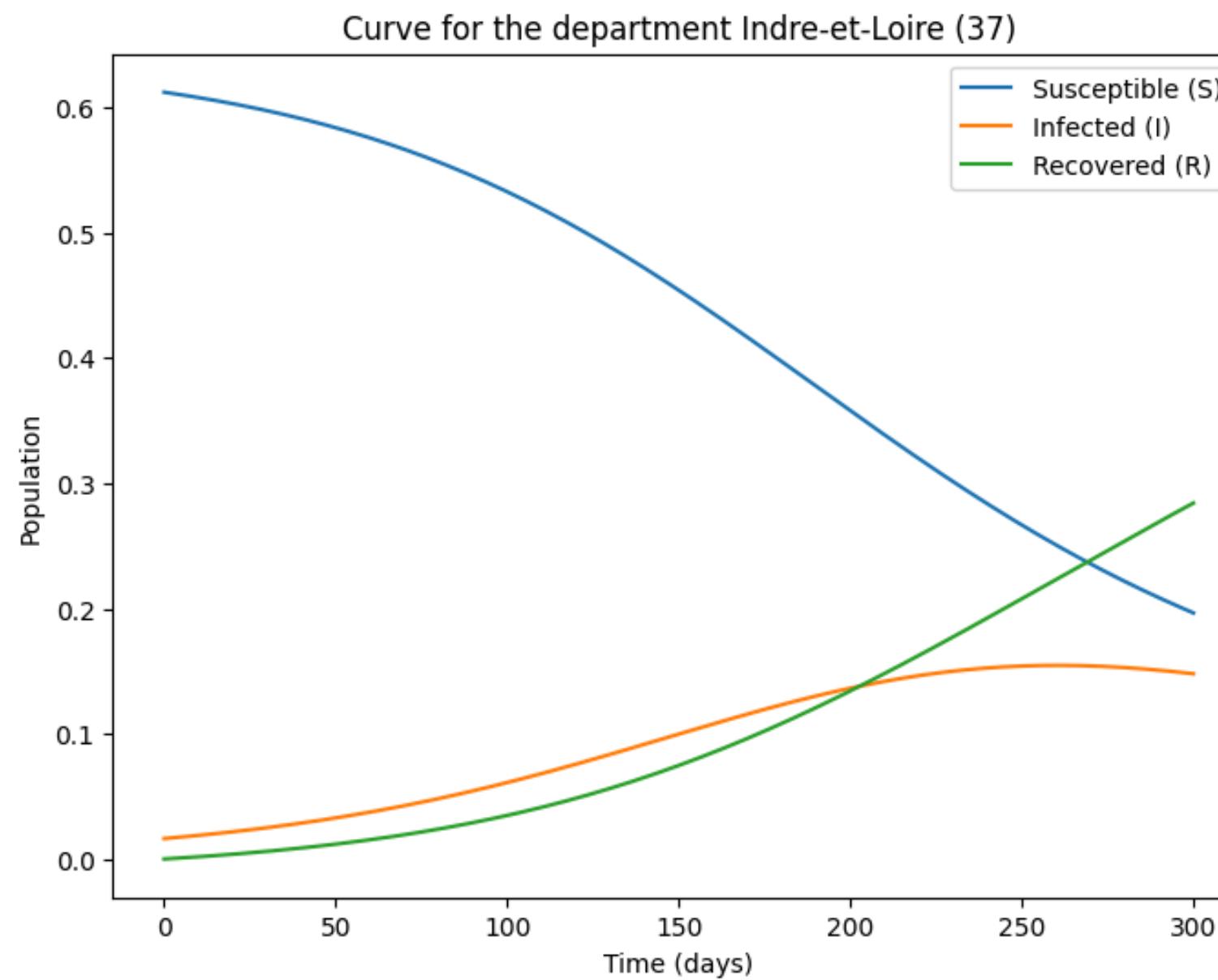


# Modèle SIR - R<sub>0</sub> - Tous les départements

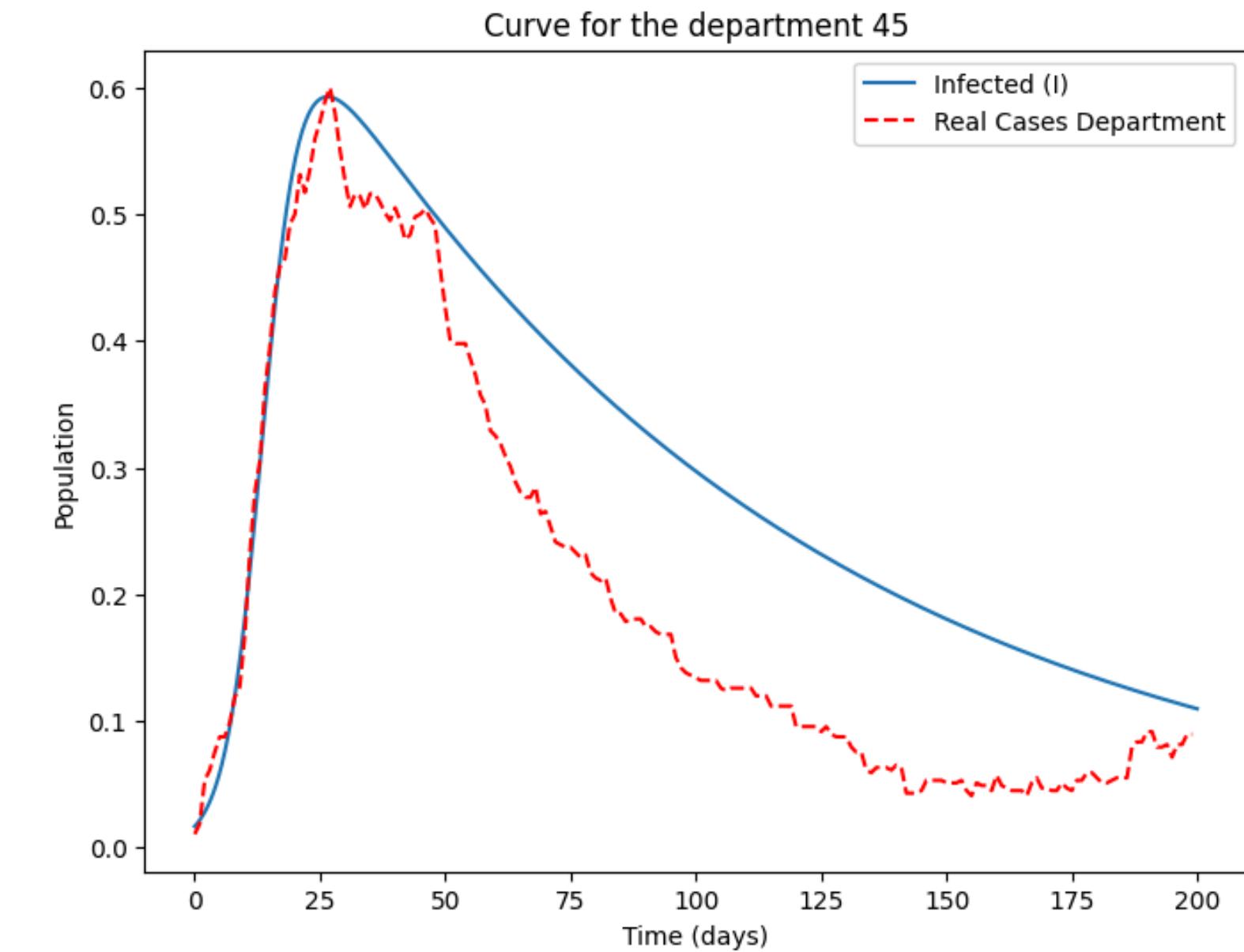
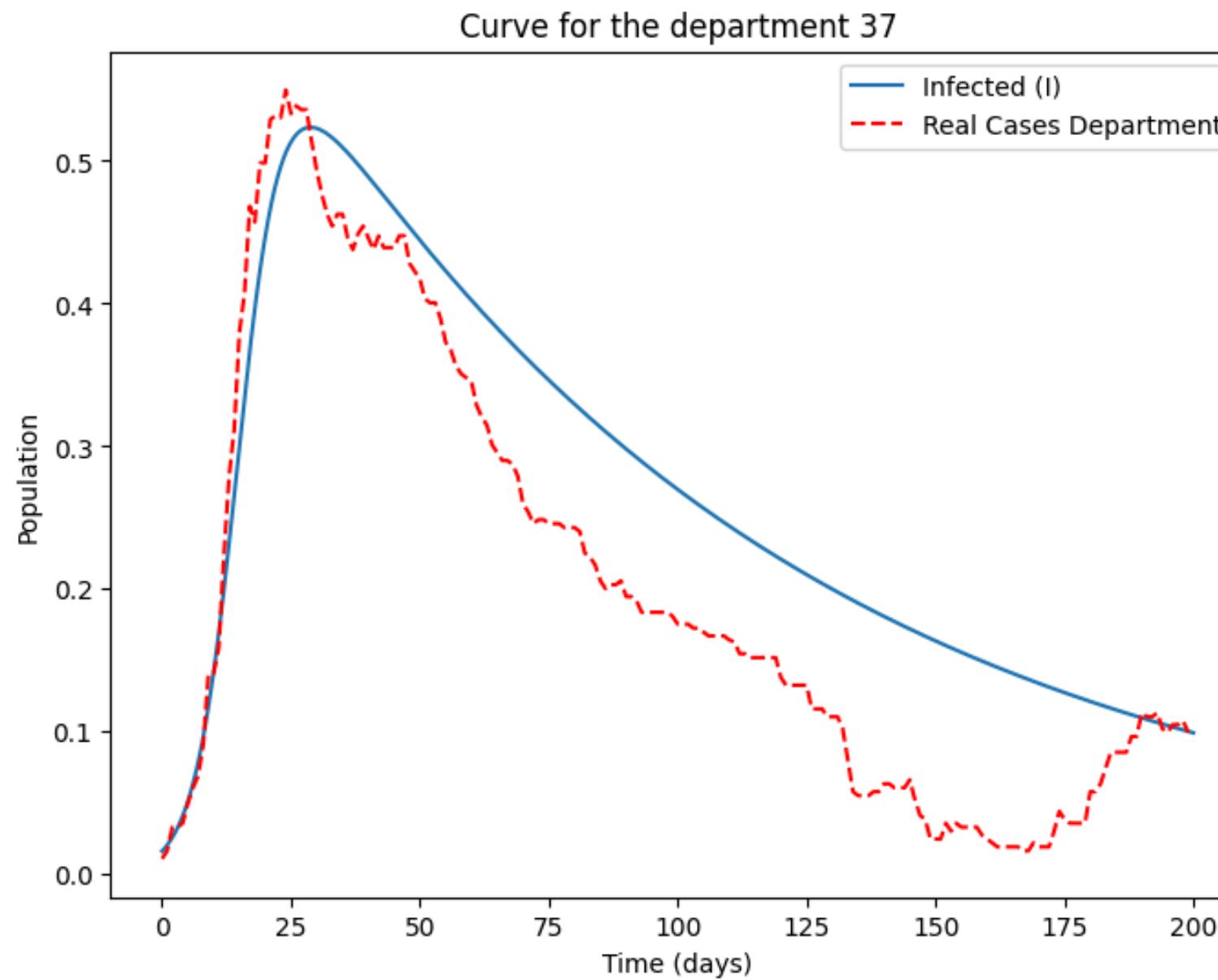


**R<sub>0</sub> above 1 means that transmission of the disease is increasing.**

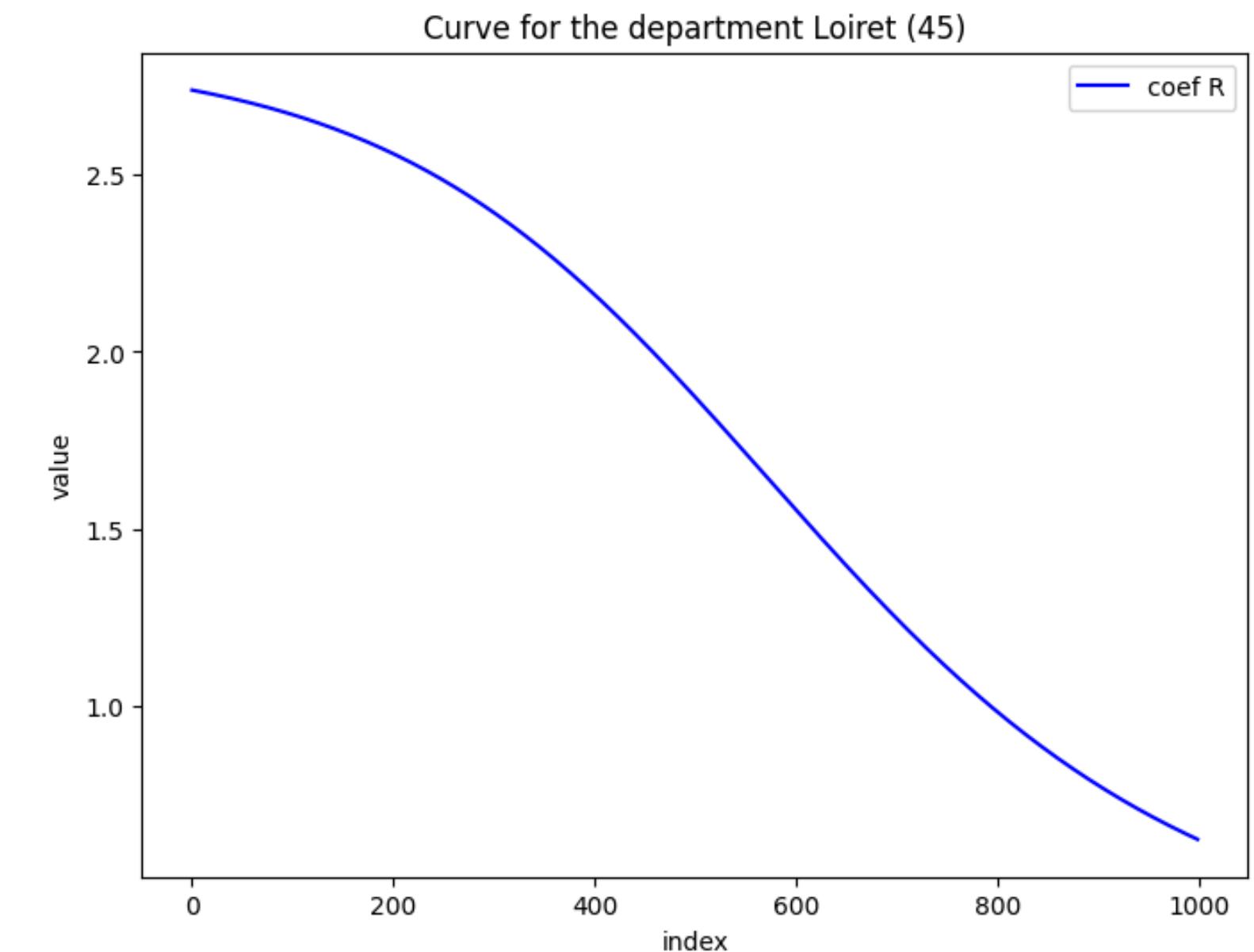
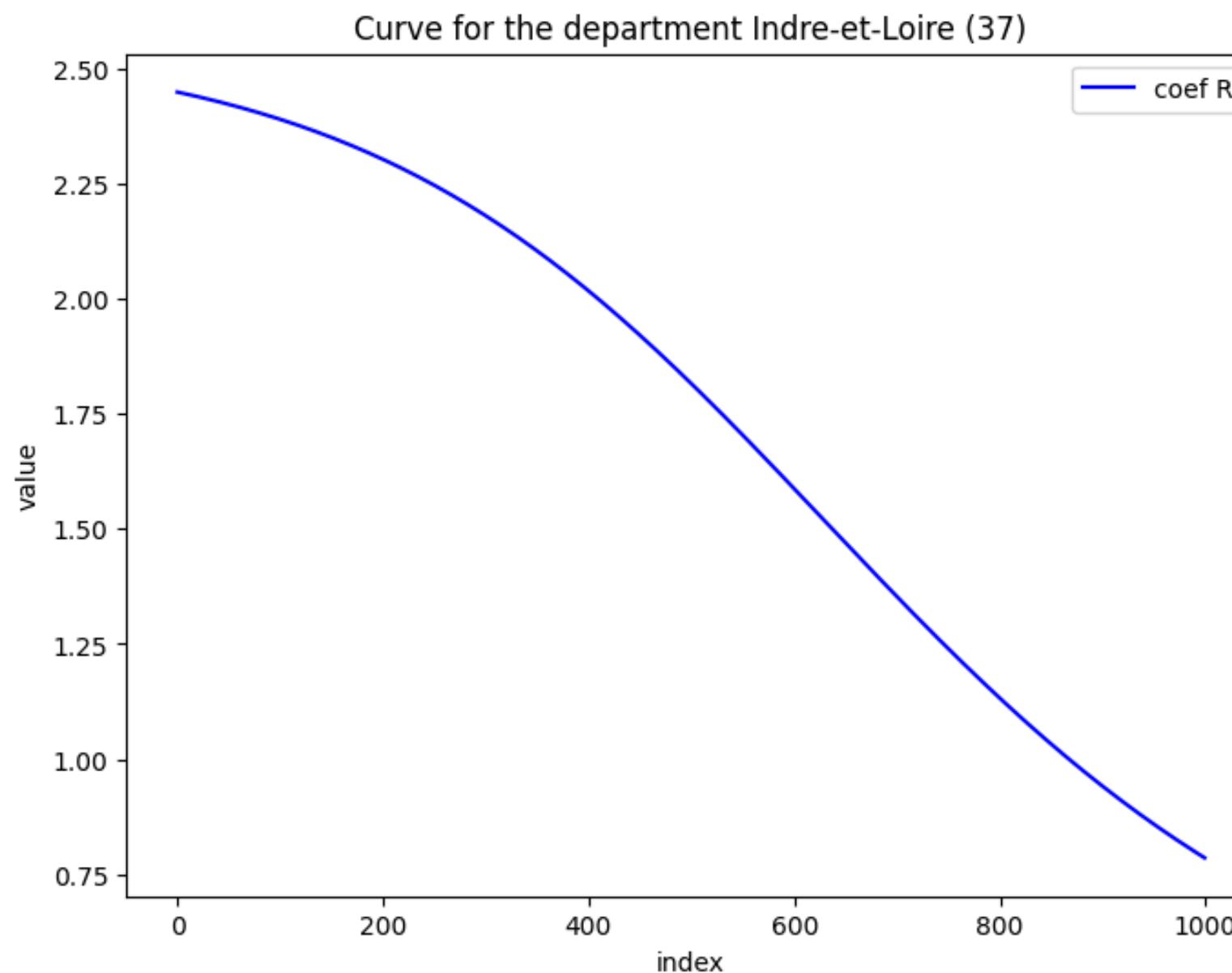
# Modèle SIR - Pour chaque département



# Modèle SIR - Pour chaque département

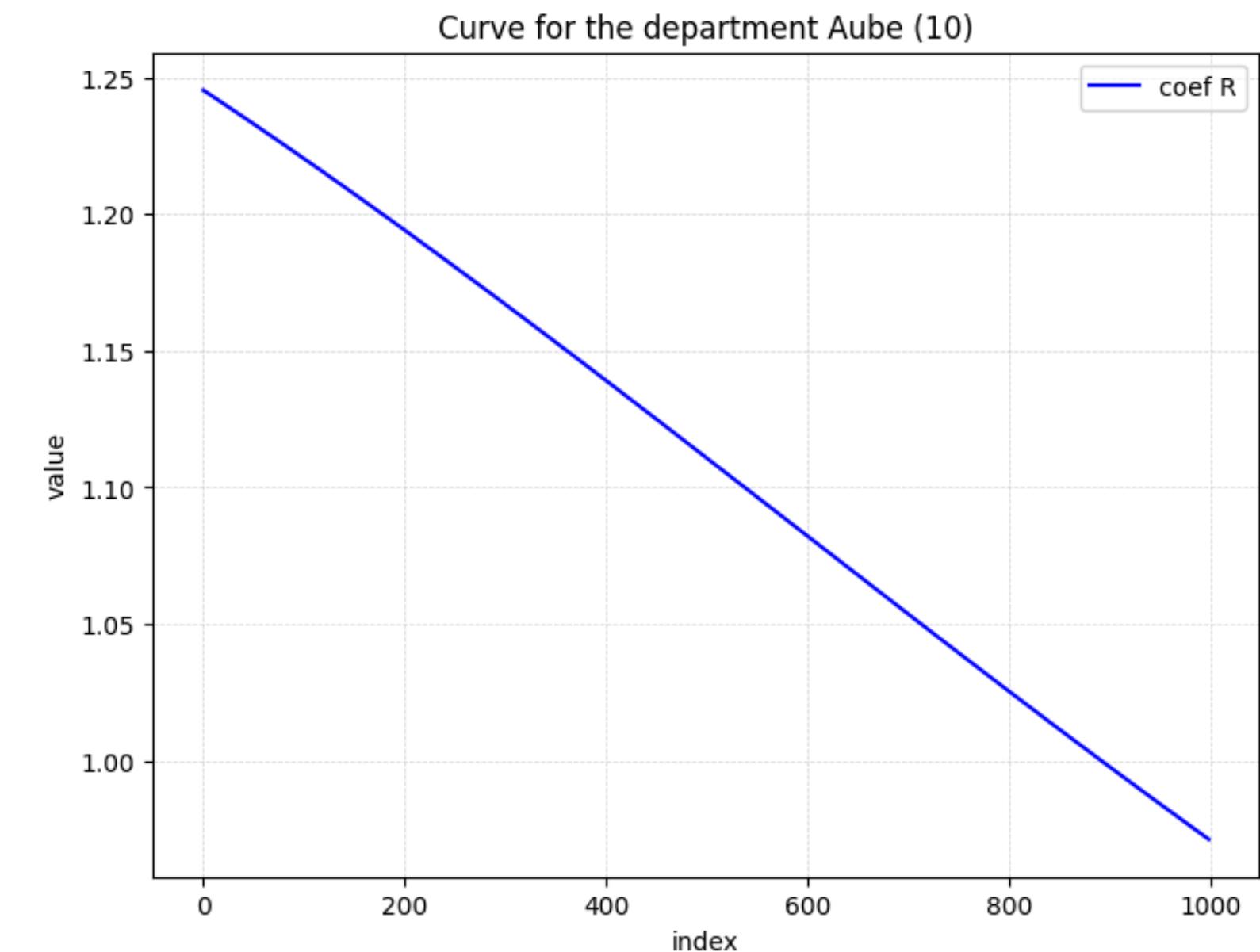
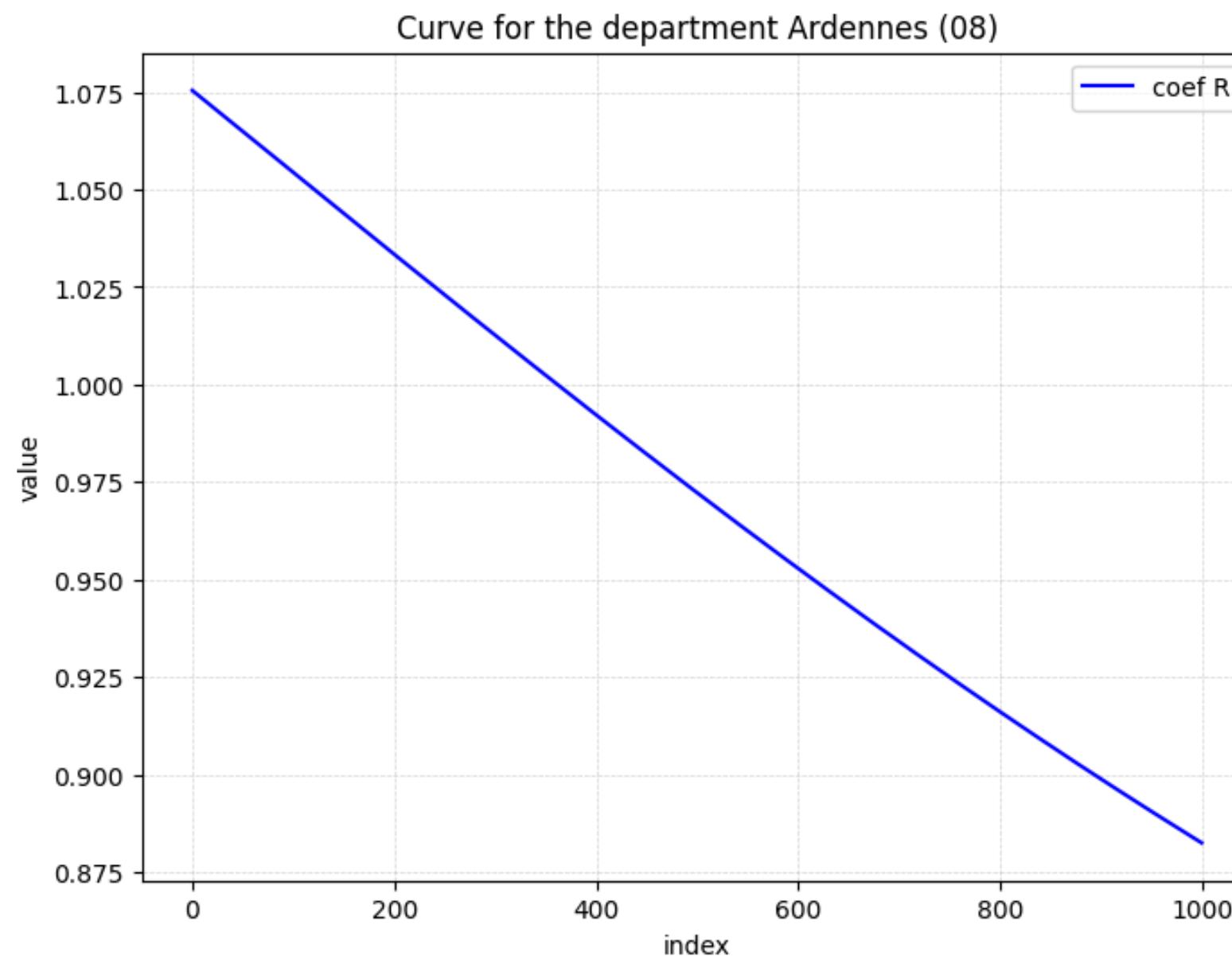


# Modèle SIR - R<sub>0</sub> - Pour chaque département



**R<sub>0</sub> above 1 means that transmission of the disease is increasing.**

# Modèle SIR - R<sub>0</sub> - Pour chaque département



**R<sub>0</sub> above 1 means that transmission of the disease is increasing.**

# Limites du modèle SIR

- Homogénéité de la population (chaque individu est susceptible d'attraper la maladie avec la même probabilité)
- Immunité permanente (les individus récupérés acquièrent une immunité permanente à la maladie)

## Perspectives d'amélioration

- Considérer les données de mobilité (déplacements entre les départements)
- Analyser les clusters (K-means : identifier des schémas de propagation similaires entre les départements)

### **III/ Modèle Autorégressif & Lissage Spatial**



# Définition de la fonction de minimisation

**Ajouter une dimension  
spaciale au modèle**

**Modéliser l'influence des  
différententes régions**

**AIC  $\rightarrow n = 11$**

$$Z(t) = \sum_{k=1}^L \alpha(k) Z(t-k)$$

$$\hat{\alpha} = \operatorname{argmin}_{\alpha} \|Z(t) - \sum_{k=1}^L \alpha(k) Z(t-k)\|_2^2$$

$$Z(t, c) = \sum_{k=1}^L \alpha(k, c) Z(t-k, c)$$

$$\hat{\alpha} = \operatorname{argmin}_{\alpha} \sum_{c=1}^6 \|Z(t, c) - \sum_{k=1}^L \alpha(k, c) Z(t-k, c)\|_2^2 + \lambda \sum_{c, c'} \omega(c, c') \|\alpha(c) - \alpha(c')\|_2^2$$

$$\omega(c, c') = \frac{\operatorname{pop}(c) * \operatorname{pop}(c')}{2 * d(c, c')}$$

## IV/ Résultats Obtenus



# Prédiction du nombre de cas et évaluation du modèle

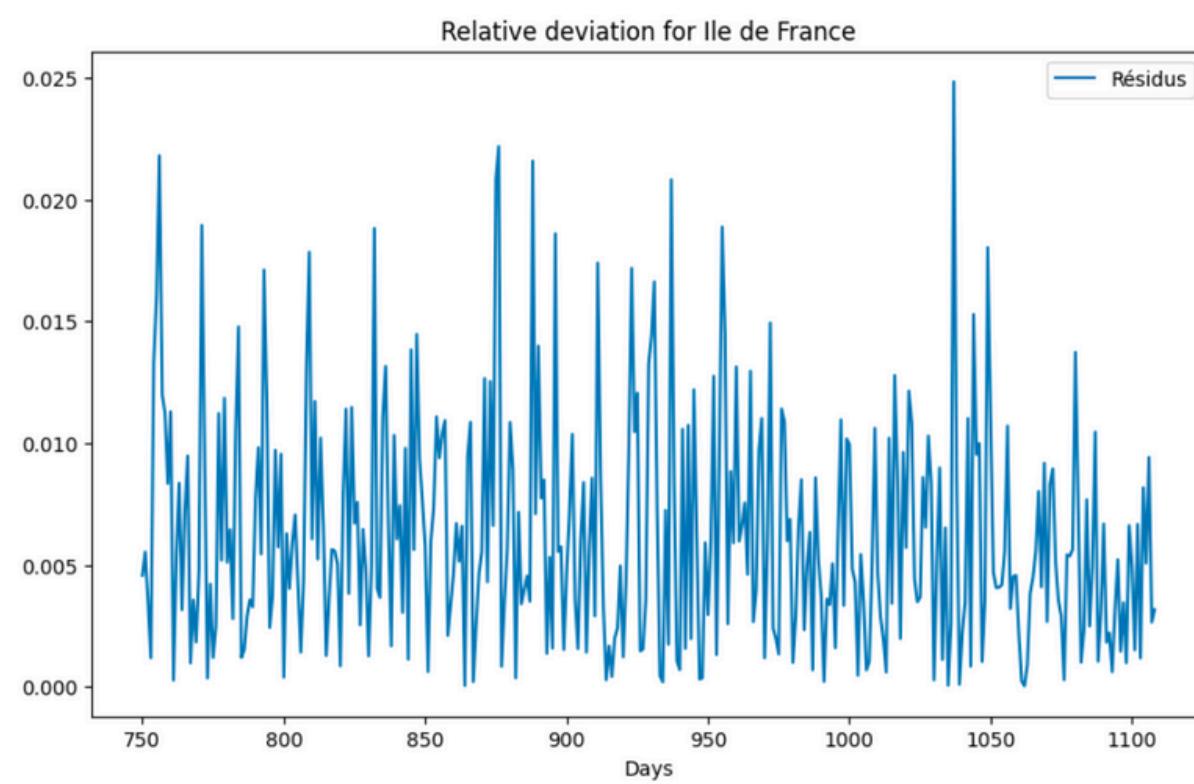


FIGURE 19 – IDF Relative deviations

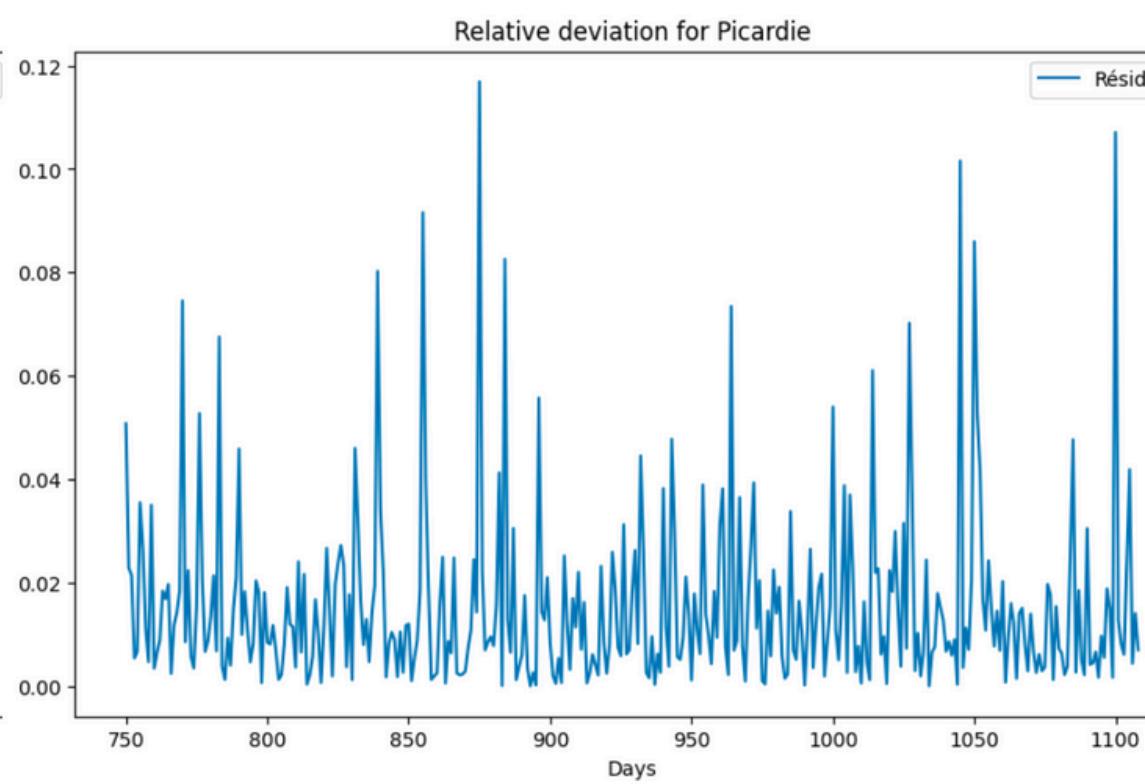


FIGURE 20 – Picardie Relative differences

$$MAE = \left( \frac{1}{n} \right) \sum_{i=1}^n |Y_i - \hat{Y}_i|$$

$$MSE = \left( \frac{1}{n} \right) \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

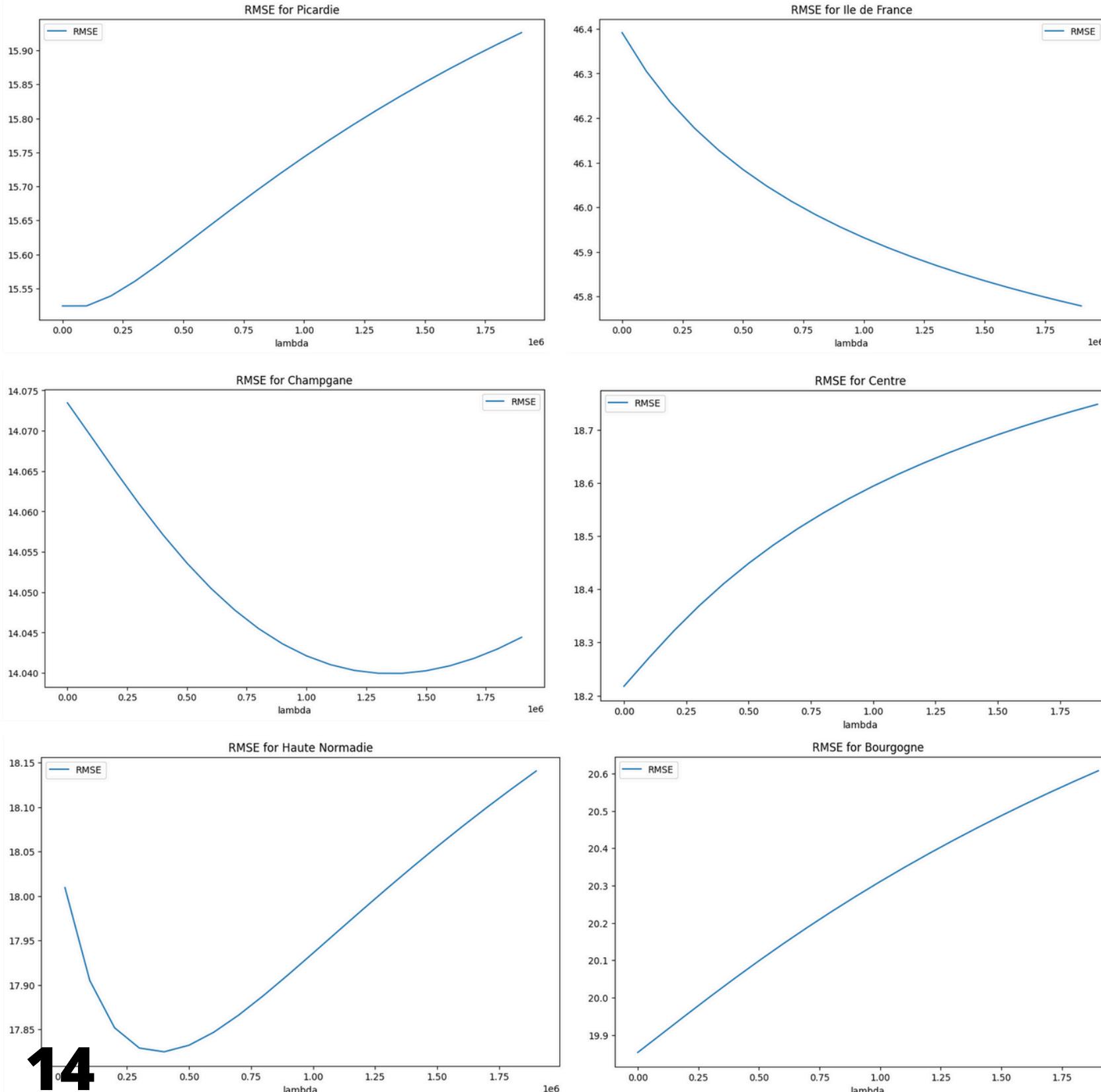
$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}$$

Region	RMSE Value
Bourgogne	18.95
Centre	17.56
Champagne	13.70
Haute Normandie	17.35
Ile de France	42.25
Picardie	15.45

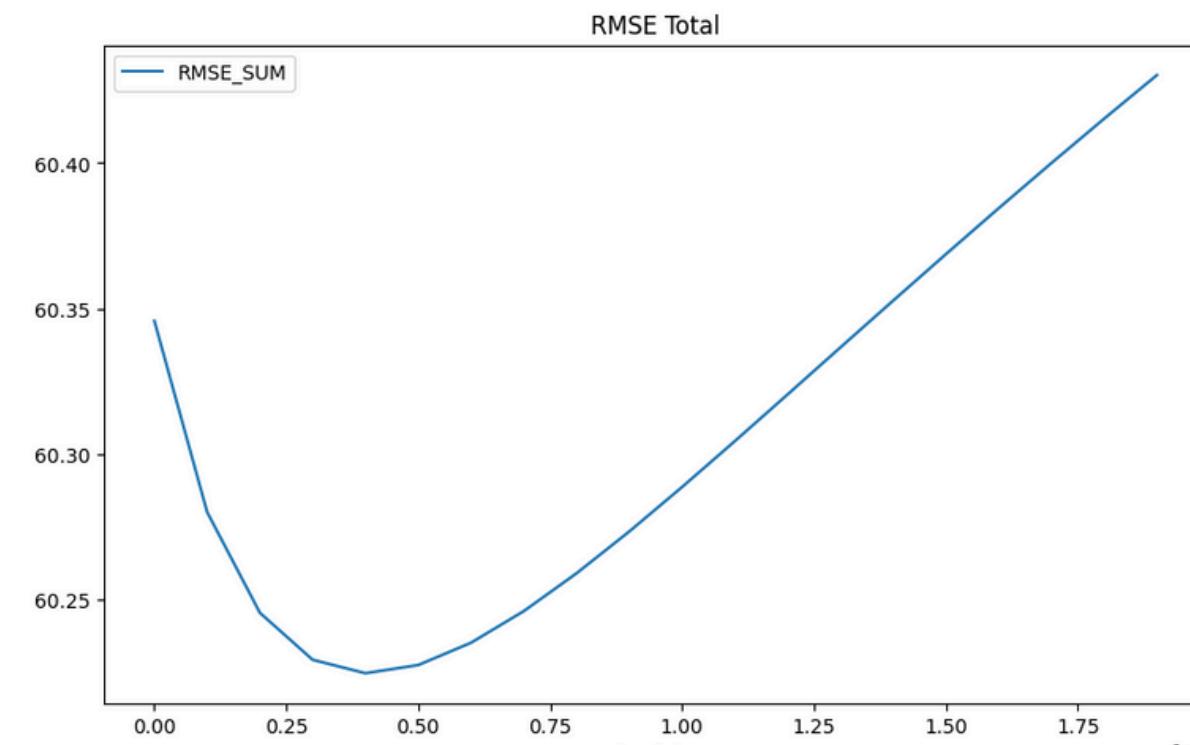
TABLE 3 – RMSE Values for  $\lambda = 500000$  and  $n = 11$

# Pénalité Lambda

$$\hat{\alpha} = \operatorname{argmin}_{\alpha} \sum_{c=1}^6 \|Z(t, c) - \sum_{k=1}^L \alpha(k, c) Z(t - k, c)\|_2^2 + \lambda \sum_{c, c'} \omega(c, c') \|\alpha(c) - \alpha(c')\|_2^2$$



$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}$$



## IV/ Résultats Obtenus

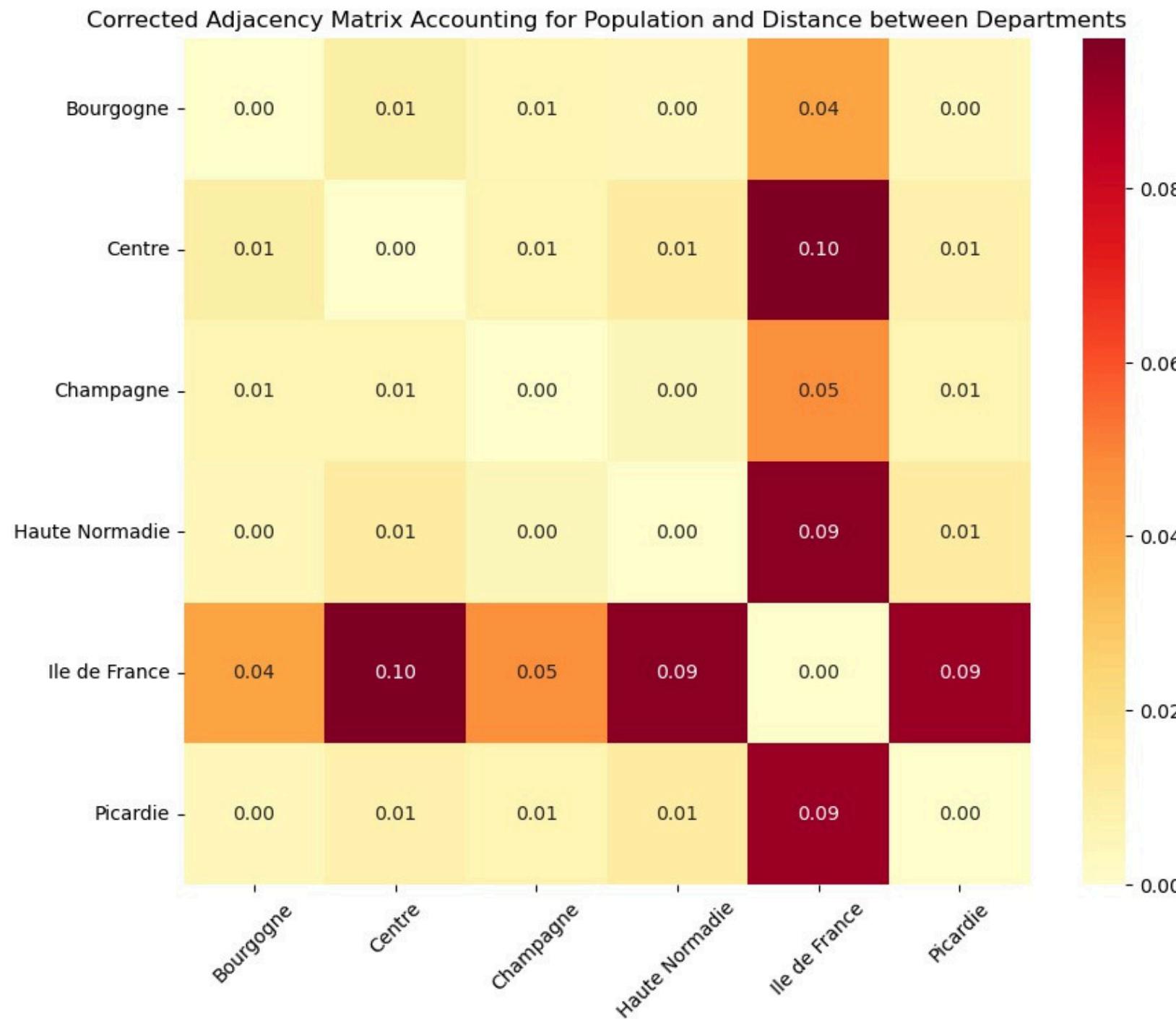
**To determine how one region's infection values is correlated to other regions**

To observe if there is a pattern between how one region affects the other

**Heatmap showing the distance and population values (adjacency matrix) between regions**

To observe if the pattern in the heatmap is similiar to the correlation plots

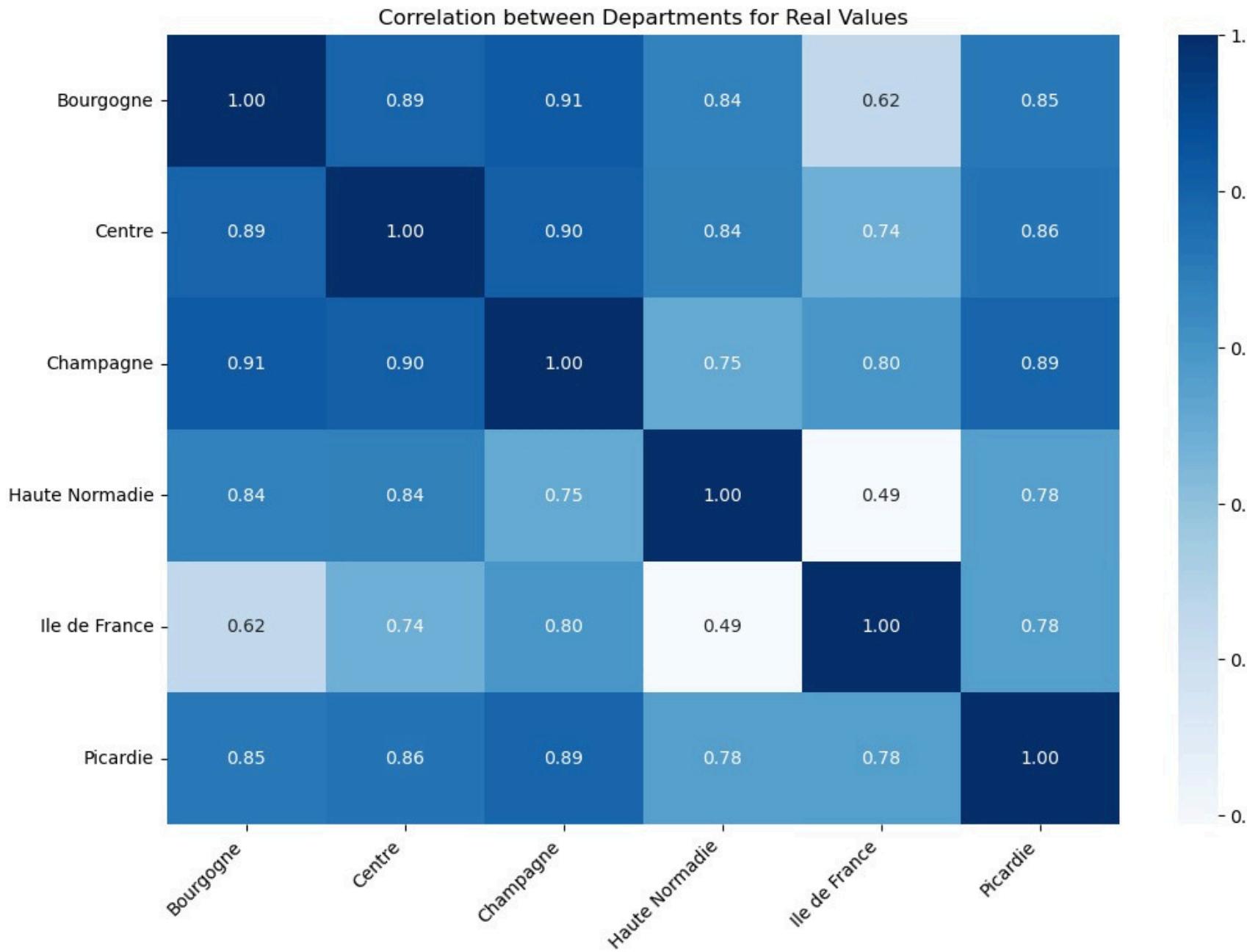
# IV/ Résultats Obtenus



**Heatmap of distance and population between regions**

Regions seem to have greater relation to Ile de France compared to other regions

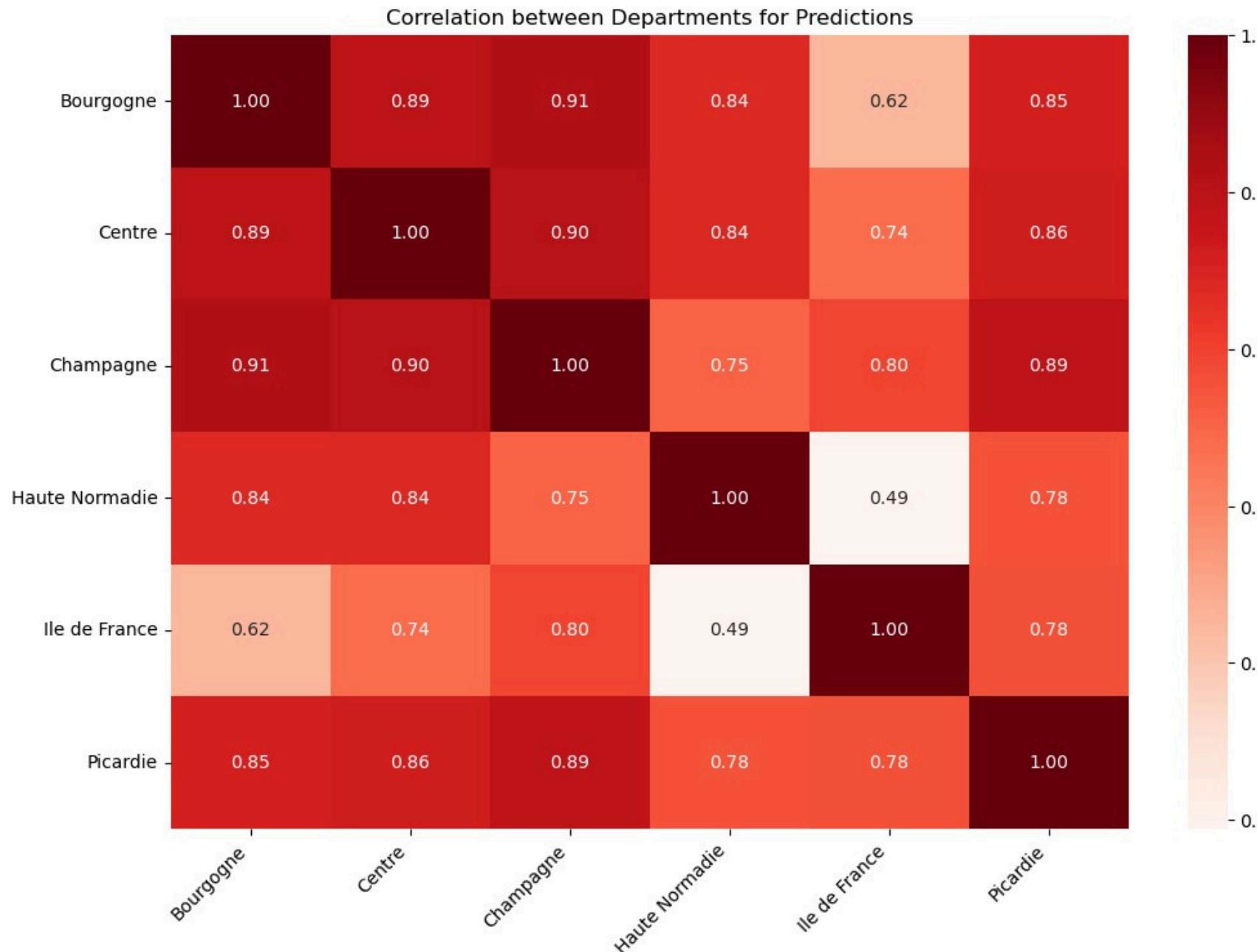
# IV/ Résultats Obtenus



**Correlation  
region's real  
values**

**Regions correlated with Ile  
de France have a less  
positive correlation**

# IV/ Résultats Obtenus



**Correlation between region's predicted infection values**

**Regions correlated with ile de france have a less positive correlation**

## IV/ Résultats Obtenus

### What can we gather:

There is a pattern surrounding Ile de France :

- Significant correlation between Ile de France and all other regions
- More “heat” (more significant values) for adjacency matrix values for regions related to Ile de France and less positive correlation for infection values

## IV/ Résultats Obtenus

### What can we gather:

- We can see that there seems to be a relationship between the distance and population and infection numbers for each region
  - Further analysis: determine if there is correlation between distance and population and infection values
    - Moran's I

## IV/ Résultats Obtenus

### Moran's I:

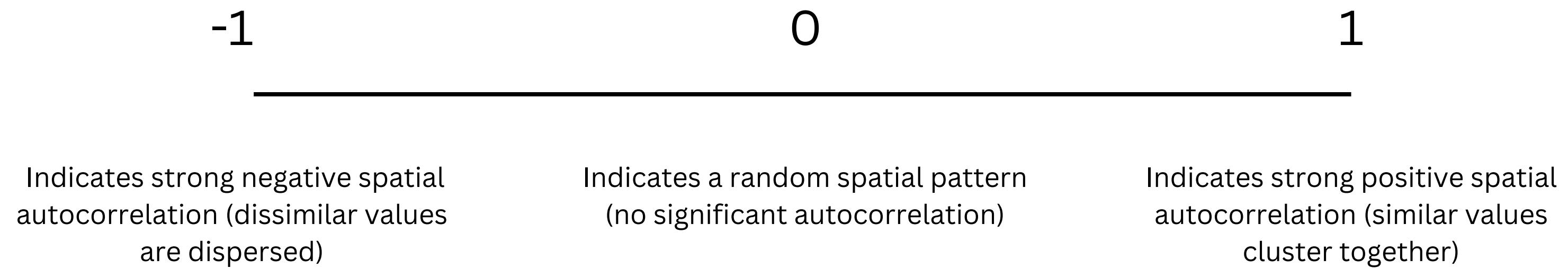
$$I = \frac{N}{W} \cdot \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2}$$

where :

- $N$  is the number of spatial units indexed by  $i$  and  $j$ ,
- $x_i$  and  $x_j$  are the observations for spatial units  $i$  and  $j$ ,
- $\bar{x}$  is the mean of all observations,
- $w_{ij}$  is the spatial weight between units  $i$  and  $j$ ,
- $W$  is the sum of all spatial weights,  $W = \sum_{i=1}^N \sum_{j=1}^N w_{ij}$ .

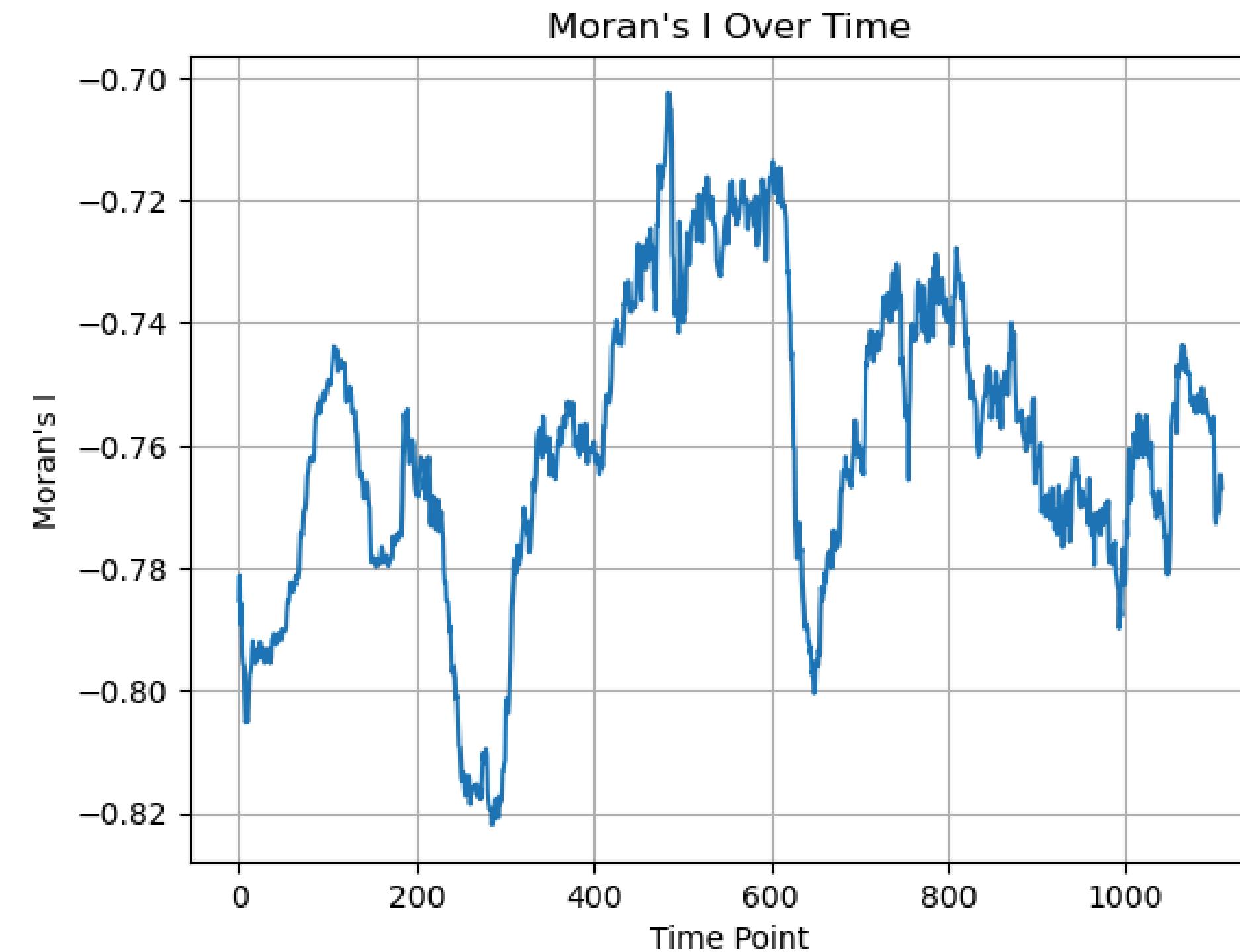
# IV/ Résultats Obtenus

## Moran's I:



# IV/ Résultats Obtenus

## Moran's I:

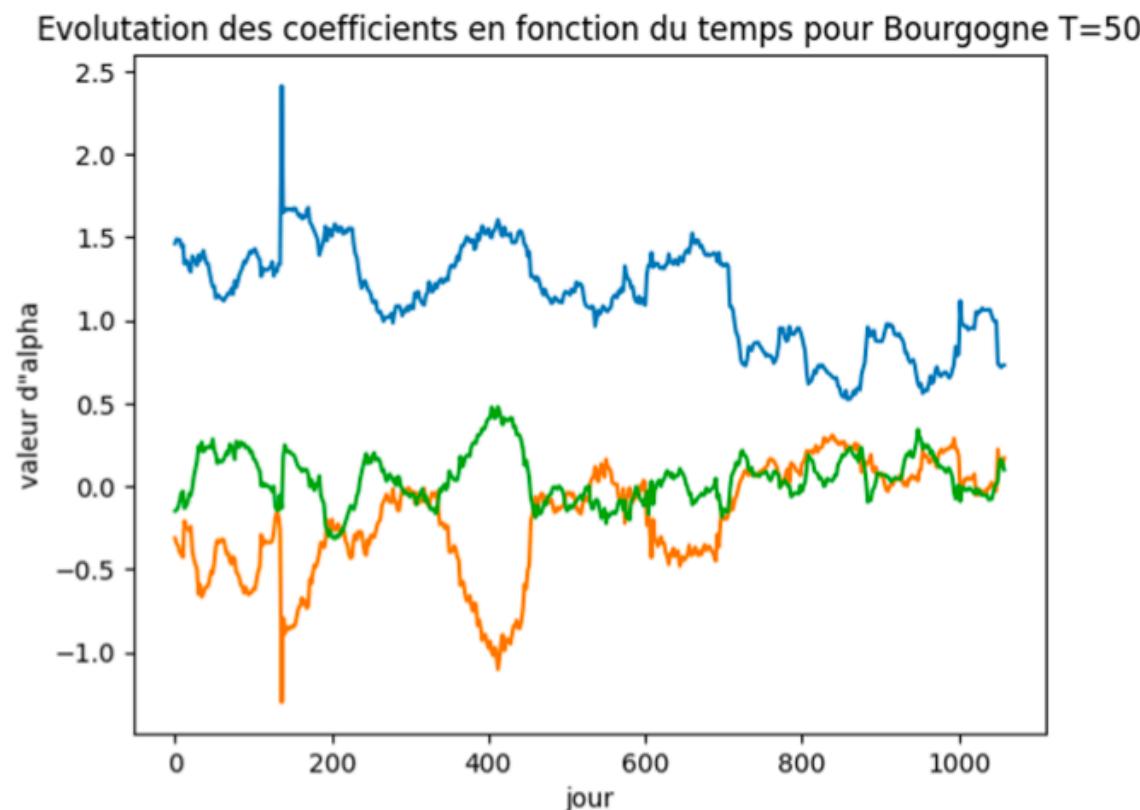


# Fenêtrage et R0

$$Z(t, c) = \sum_{k=1}^L \alpha(k, t, c) Z(t - k, c)$$

$$\hat{\alpha} = \operatorname{argmin}_{\alpha} \sum_{c=1}^6 \|Z(t, c) - \sum_{k=1}^L \alpha(k, t, c) Z(t - k, c)\|_2^2 + \lambda \sum_{c, c'} \omega(c, c') \|\alpha(t, c) - \alpha(t, c')\|_2^2$$

Fenêtre de taille  $T = 50$  : On définit des coefficients alphas sur  $[t - T/2 ; t + T/2]$



Hypothèses :

$$\alpha(k, t) = R0(t)\beta(k) \quad \text{et} \quad \sum \beta(k) = 1 \quad \text{donc} \quad R0(t) = \sum \alpha(k, t)$$

# Evolution de $R_0$

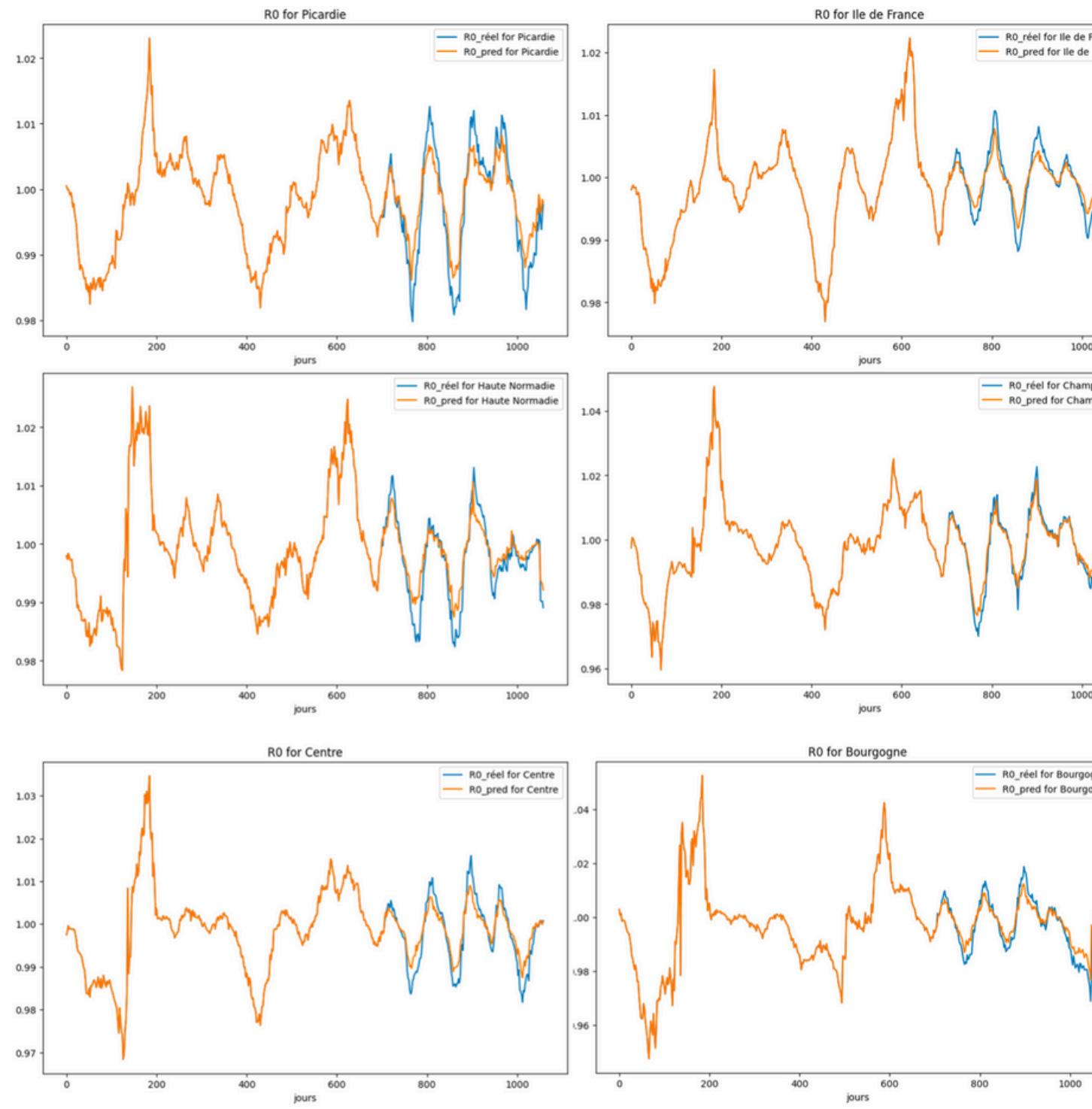


FIGURE 29 –  $R_0$  for each region

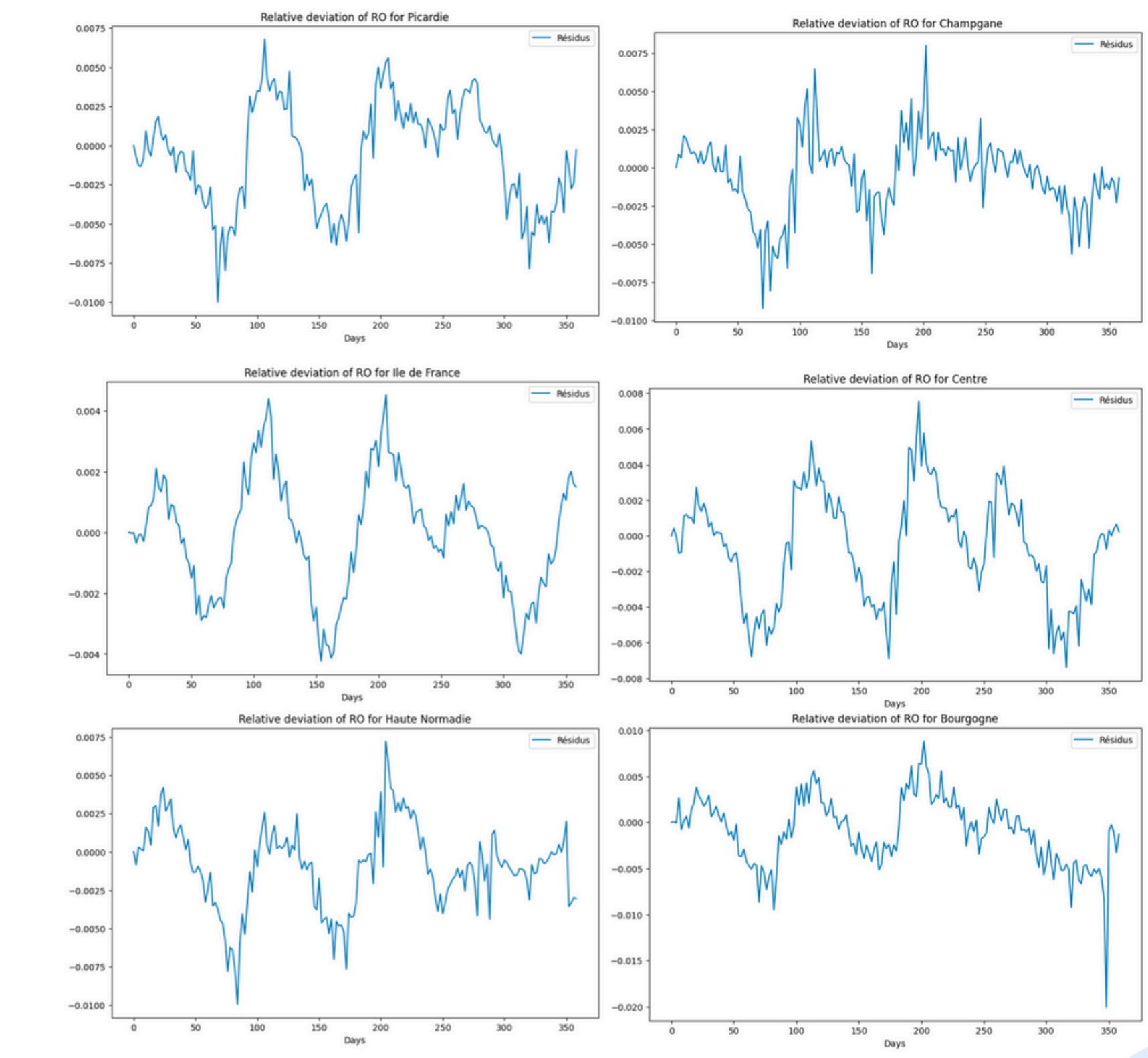


FIGURE 30 – Relative Deviation of  $R_0$  for each region

## **V/ Conclusion & Perspectives**



# V/ Conclusion et Perspectives

- Modèle Stochastique (loi de Poisson) ;
- Déterminer jusqu'à quand le modèle est vrai (ex: 15 jours) ;
- Affiner la matrice d'adjacence sur les déplacements principaux (exemple : beaucoup de trajet paris-bordeaux), population, prise en compte de la distance ;
- A l'échelle de la France ;
- Affiner départements par départements (au lieu de 6 zones géographiques, en considérer une centaine) ;
- Prédire l'évolution de alpha (non  $z_t$ ) car c'est alpha qui comprend l'information sur  $R_0$ .