Apprentissage par Renforcement pour la Modélisation de Pandémie

Modèle SEIRD & Optimisation des Politiques Sanitaires

Charles Sutty & Martin Reix

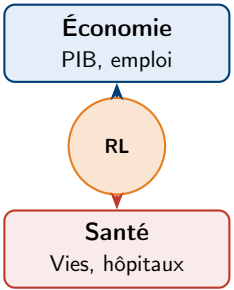
École des Mines de Nancy — 4A

Février 2026

Sommaire

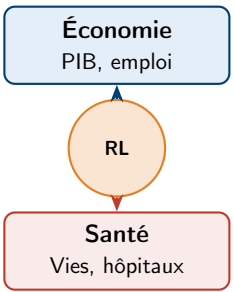
Contexte

- Pandémie du **Covid-19** : crise sanitaire mondiale



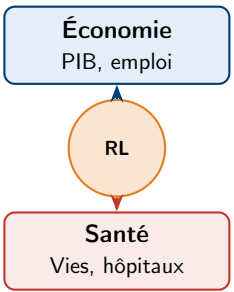
Contexte

- Pandémie du **Covid-19** : crise sanitaire mondiale
- Mesures gouvernementales : **confinement** et **vaccination**



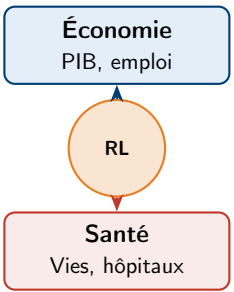
Contexte

- Pandémie du **Covid-19** : crise sanitaire mondiale
- Mesures gouvernementales : **confinement** et **vaccination**
- Compromis fondamental entre :
  - Pertes humaines
  - Pertes économiques
  - Libertés individuelles



Contexte

- Pandémie du **Covid-19** : crise sanitaire mondiale
- Mesures gouvernementales : **confinement** et **vaccination**
- Compromis fondamental entre :
  - Pertes humaines
  - Pertes économiques
  - Libertés individuelles
- Peut-on utiliser le **Reinforcement Learning** pour guider les politiques publiques ?



Objectifs du projet

Objectif principal

Comparer différentes approches de **RL** pour optimiser simultanément confinement et vaccination dans un modèle épidémiologique réaliste.

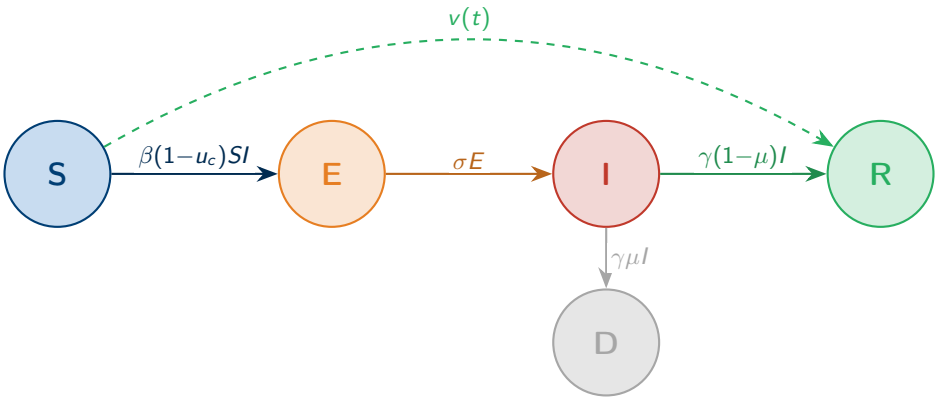
Approche

1. Modélisation SEIRD
2. Fonction de coût multi-objectif
3. Algorithmes : SARSA, Q-Learning, DP
4. Comparaison avec heuristiques

Défis

- Espace d'états continu
- Compromis économie / santé
- Stabilité des politiques
- Réalisme des paramètres

Le modèle SEIRD



S : Susceptibles **E** : Exposés
I : Infectés **R** : Rétablis
D : Décédés

Conservation : $N = S + E + I + R + D = 1$
$v(t) = \min(u_{\text{vacc}} \cdot v_{\max},\ S(t))$

Équations du modèle

Dynamique SEIRD

$$\frac{dS}{dt} = -\beta(1 - u_{\text{conf}})SI - v(t)$$

$$\frac{dE}{dt} = \beta(1 - u_{\text{conf}})SI - \sigma E$$

$$\frac{dI}{dt} = \sigma E - \gamma I$$

$$\frac{dR}{dt} = \gamma(1 - \mu)I + v(t)$$

$$\frac{dD}{dt} = \gamma\mu I$$

Paramètres :

- β : taux de transmission
- σ : taux d'incubation ($1/7$ j^{-1})
- γ : taux de guérison ($1/10$ j^{-1})
- μ : taux de mortalité (1%)

Contrôles :

- $u_{\text{conf}} \in [0, 1]$: confinement
- $u_{\text{vacc}} \in [0, 1]$: vaccination

Taux de reproduction

$$R_0 = \beta/\gamma \approx 2.7$$

Fonction de coût

On cherche à minimiser le coût total à chaque pas de temps :

$\mathcal{L}(s_t, a_t)$

$$\mathcal{L} = \underbrace{C_{\text{eco}} \times u_{\text{conf}}^2}_{\text{Coût économique}} + \underbrace{C_{\text{vacc}} \times \frac{v(t)}{v_{\max}}}_{\text{Coût logistique}} + \underbrace{C_{\text{vie}} \times \Delta D}_{\text{Pertes humaines}} + \underbrace{C_{\text{hosp}} \times I_t}_{\text{Coût hospitalier}}$$

$\mathcal{L}_{\text{eco}}$	$\mathcal{L}_{\text{log}}$	$\mathcal{L}_{\text{décès}}$	$\mathcal{L}_{\text{hosp}}$
Confinement^2	Vaccination	Décès ΔD	Infectés I

Note : Le coût du confinement est **quadratique** — un confinement modéré coûte peu, un confinement total est très coûteux (systèmes critiques).

Paramètres économiques

Estimations basées sur les données du Covid-19 en France :

Paramètre	**Valeur réelle**	**Normalisée**
C_{eco} (perte éco. / jour)	$30\% \times 6{,}5$ Md€ ≈ 2 Md€	1,0
C_{vie} (par décès)	3 M€ / personne	≈ 1000
C_{vacc} (logistique totale)	3,5 Md€	0,02
C_{hosp} (par infecté / jour)	580 € / patient / jour	≈ 20

Sources

- OFCE (2020) — impact sur le PIB
- INSEE — PIB pré-pandémie
- Valeur statistique d'une vie (gouv.)
- Coûts hospitalisation Covid

Normalisation

Toutes les valeurs sont normalisées par C_{eco} pour la stabilité numérique des algorithmes.

Discrétisation de l'espace

Discrétisation des états (bins) :

- **Susceptibles (S)** : 12 bins, découpage non-uniforme
  $[0, 0.2, 0.4, 0.6, 0.7, 0.8, 0.85, 0.9, 0.93, 0.95, 0.97, 0.985, 1.0]$
- **Exposés (E) et Infectés (I)** : 12 bins, échelle **logarithmique**
  $[0, 10^{-5}, 3\times10^{-5}, 10^{-4}, 3\times10^{-4}, 10^{-3}, 3\times10^{-3}, 10^{-2}, 3\times10^{-2}, 0.1, 0.3, 0.5, 1.0]$

Espace d'états (S, E, I)

$12^3 = 1\,728$ états

Espace d'actions

$5 \times 5 = 25$ actions par état

$u \in \{0, 0.25, 0.5, 0.75, 1\}$

Pourquoi une échelle logarithmique ?

L'épidémie évolue de manière exponentielle. Une discrétisation logarithmique pour E et I permet de capturer avec précision les dynamiques critiques lorsqu'une faible partie de la population est infectée.

Algorithmes d'apprentissage

SARSA (*on-policy*)

1. Choisir a par ϵ-greedy
2. Exécuter a, observer r, s'
3. Choisir a' par ϵ-greedy
4. Mise à jour :
   $Q(s,a) \mathrel{+}= \alpha\big[r + \gamma Q(s',a') - Q(s,a)\big]$

Met à jour avec l'action **réellement choisie**

Q-Learning (*off-policy*)

1. Choisir a par ϵ-greedy
2. Exécuter a, observer r, s'
3. Mise à jour :

$$Q(s,a) \mathrel{+}= \alpha\big[r + \gamma \max_{a'} Q(s',a') - Q(s,a)\big]$$

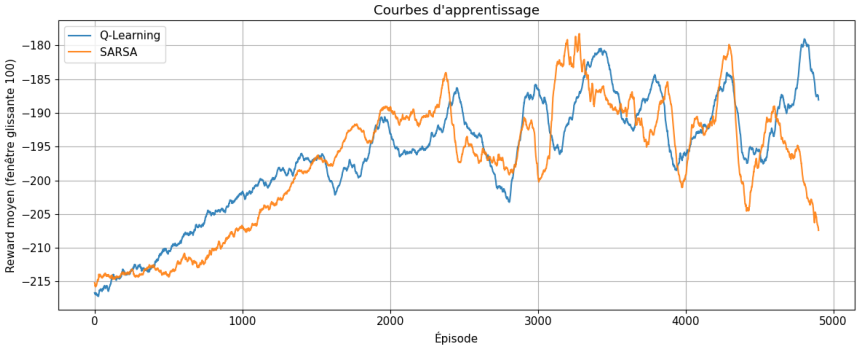
Met à jour avec l'action **optimale**

Programmation Dynamique approchée

Deux variantes : **Itération de la politique** et **Itération de la valeur** avec estimation Monte-Carlo des transitions.

$$V_t(s) = \max_a \big(\hat{\mathbb{E}}[r(s,a,S')] + \gamma\, \hat{\mathbb{E}}[V_{t+1}(S')]\big)$$

Courbes d'entraînement



- Les deux algorithmes convergent en $\sim$5 000 épisodes
- Q-Learning présente une convergence plus stable
- La récompense cumulée se stabilise, signe de convergence de la politique

Quatre politiques de référence

Aucun contrôle

$u_c = 0,\ u_v = 0$

Borne supérieure du coût. Scénario « Brésil ».

Confinement par seuil

$u_c = \mathbb{1}_{I>0.02},\ u_v = 0.5$

Réaction « tout ou rien ». Premiers confinements français.

Vaccination max

$u_c = 0,\ u_v = 1$

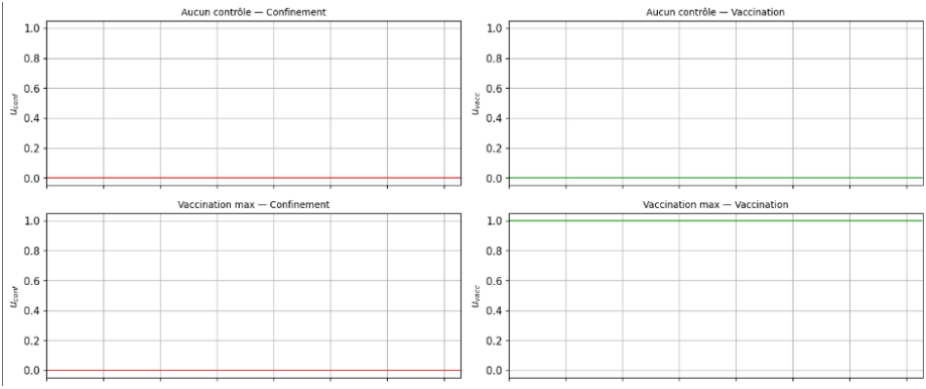
Effet de la vaccination seule. Coût logistique max.

Double seuil adaptatif

u_c gradué + u_v adapté à $S(t)$

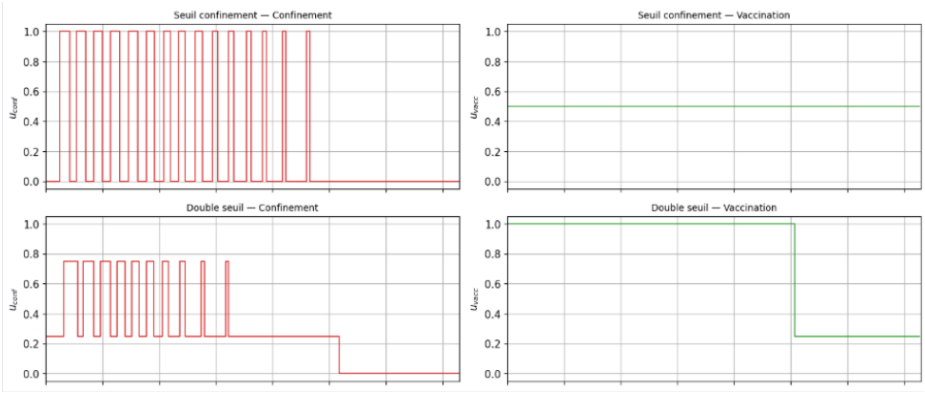
Confinement progressif + vaccination ciblée. La plus réaliste.

Politiques des heuristiques



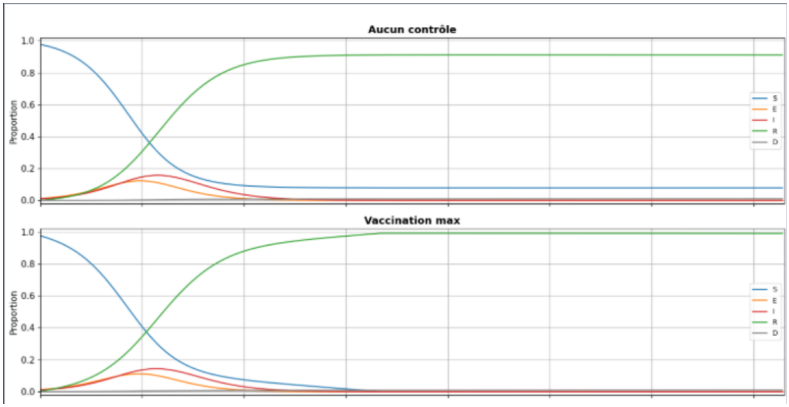
- Aucun contrôle $\rightarrow$ rien à signaler, taux nuls
- Vaccination max $\rightarrow$ vaccination à taux plein

Politiques des heuristiques



- Confinement par seuil $\Rightarrow$ oscillations de $\sim$10 jours
- Double seuil $\Rightarrow$ transitions plus souples, vaccination agressive puis relâchée

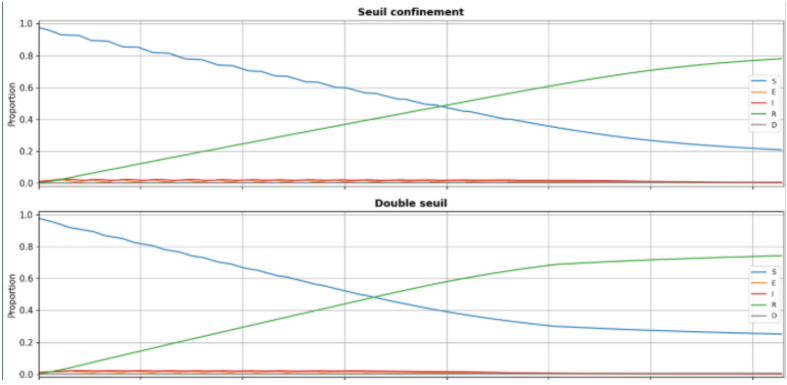
Évolution SEIRD — Heuristiques



Sans contrôle / Vacc. seule

Croissance rapide des infectés. Vacciner seul est insuffisant.

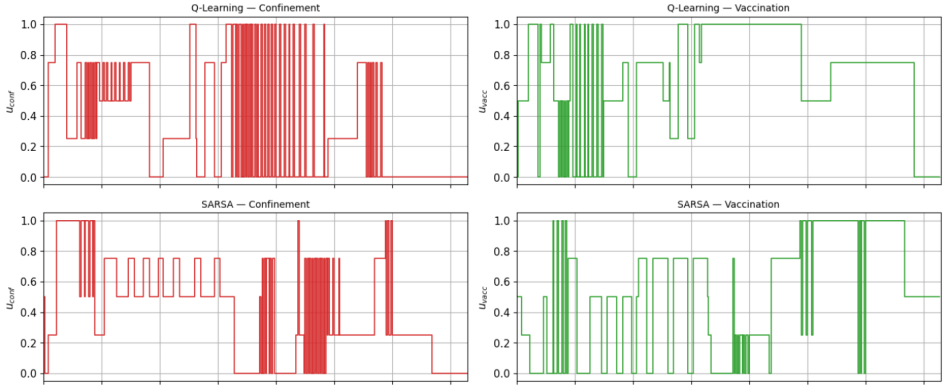
Évolution SEIRD — Heuristiques



Seuil / Double seuil

Infectés bien plus faibles, mais épidémie plus longue.

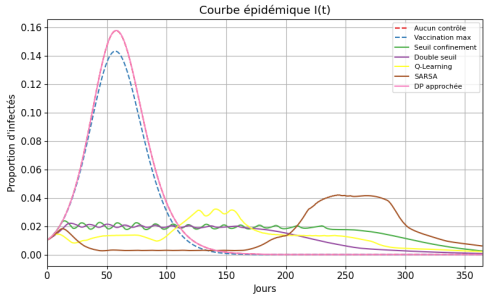
Politiques des algorithmes RL

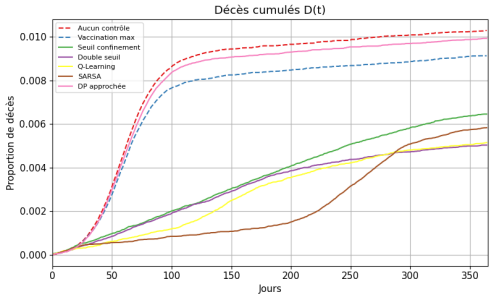


- Politiques **agressives** : oscillations entre confinement total et nul
- **SARSA** : confinement fort en début, puis vaccination intensive
- **Q-Learning** : confinement fort + vaccination plus tardive

→ Manque de contrainte de *lissage* dans l'espace d'actions

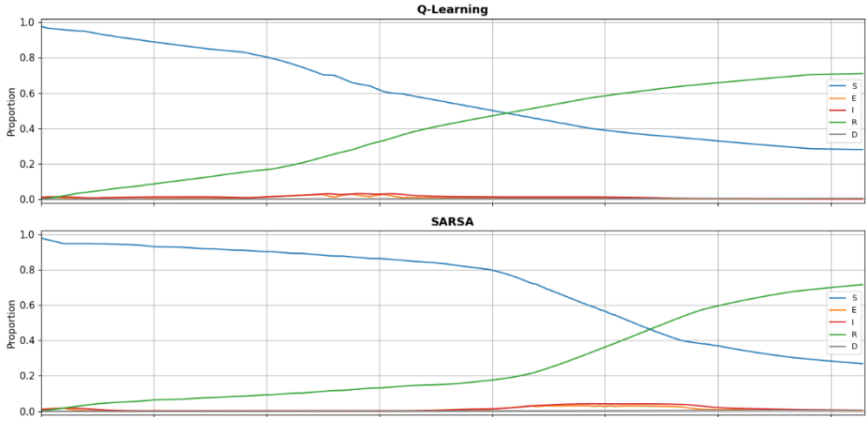
Évolution de l'épidémie — Comparaison globale





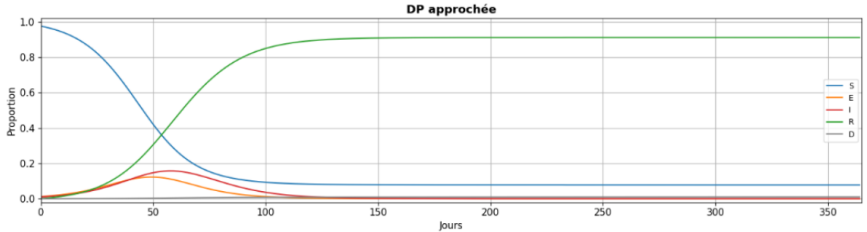
- Sans contrôle : impact le plus dévastateur
- SARSA et Q-Learning : les plus **prudents** en termes de décès
- DP approché : résultats proches du laisser-faire (complexité du problème)

Évolution SEIRD — Algorithmes



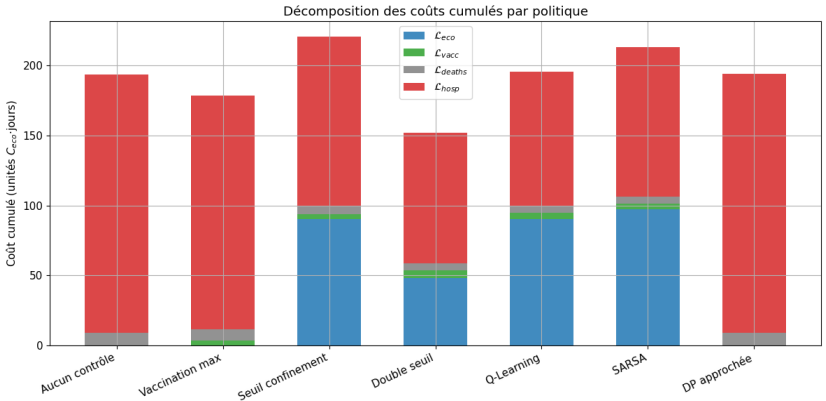
- Le pic des exposés (E) précède celui des infectés (I) $\Rightarrow$ confirmation de la dynamique d'incubation
- SARSA/Q-Learning : pertes humaines minimisées au détriment de l'économie

Évolution SEIRD — Algorithmes



- DP Learning → très similaire à aucun contrôle, probablement dû à un problème d'implémentation

Coûts financiers par scénario



Meilleur compromis

Double seuil adaptatif : limite à la fois les pertes humaines, l'hospitalisation et les pertes économiques.

RL : compromis différent

SARSA et Q-Learning privilégient les vies humaines au détriment de l'économie $\Rightarrow$ choix éthique.

Synthèse comparative

Politique	**Décès**	**Coût éco.**	**Coût hosp.**	**Réalisme**
Aucun contrôle	$\uparrow\uparrow\uparrow$	$\downarrow$	$\uparrow\uparrow\uparrow$	$\star$
Vaccination max	$\uparrow\uparrow$	$\downarrow$	$\uparrow\uparrow$	$\star\star$
Seuil	$\uparrow$	$\uparrow\uparrow$	$\uparrow$	$\star\star$
Double seuil	$\downarrow$	$\uparrow$	$\downarrow$	$\star\star\star$
SARSA	$\downarrow\downarrow$	$\uparrow\uparrow$	$\downarrow\downarrow$	$\star\star$
Q-Learning	$\downarrow\downarrow$	$\uparrow\uparrow$	$\downarrow\downarrow$	$\star\star$
DP approché	$\uparrow\uparrow$	$\downarrow$	$\uparrow\uparrow$	$\star$

$\uparrow$ = élevé (mauvais), $\downarrow$ = faible (bon)

$\star$ = réalisme de la politique

Conclusion

Résultats clés

- Le modèle SEIRD capture bien les dynamiques épidémiques
- Les algorithmes RL **minimisent les décès** mais au prix de politiques **erratiques**
- Le **double seuil adaptatif** offre le meilleur compromis global
- Le choix final relève de **décisions éthiques et politiques**

Pistes d'amélioration

- Contrainte de **lissage** des actions
- Capacité hospitalière **maximale**
- Algorithmes **Deep RL** (DQN, PPO)
- Modèle **stochastique** (aléas réels)
- Prise en compte des **variants**

Code source

`https://github.com/Charsutty/RL-pandemic-modeling`

Merci de votre attention !

Questions ?

Charles Sutty & Martin Reix

https://github.com/Charsutty/RL-pandemic-modeling

Références I

Jianhong Wu, Kathy Leung, and Gabriel Leung.

Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china : A modeling study.

Obstetrical & Gynecological Survey, 75 :399–400, 07 2020.

Fernando E Alvarez, David Argente, and Francesco Lippi.

A simple planning problem for covid-19 lockdown.

Working Paper 26981, National Bureau of Economic Research, 4 2020.

Regina Padmanabhan, Nader Meskin, Tamer Khattab, Mujahed Shraim, and Mohammed Al-Hitmi.

Reinforcement learning-based decision support system for covid-19.

Biomedical Signal Processing and Control, 68 :102676, 2021.

Ofce Observatoire Français Des Conjonctures Économiques.

Évaluation au 30 mars 2020 de l'impact économique de la pandémie de COVID-19 et des mesures de confinement en France.

OFCE Policy Brief, (65) :., March 2020.

Annexe — Paramètres épidémiologiques détaillés

Symbole	Signification	Valeur	Source
Δt	Pas de temps	1 jour	—
σ	Taux d'incubation	$1/7 \approx 0.14$	Littérature
γ	Taux de guérison	$1/10 = 0.1$	Littérature
R_0	Taux de reproduction	2.7	Littérature
β	Taux de transmission	0.27	$R_0 \times \gamma$
μ	Taux de mortalité	0.01	Littérature
$v_{\max}$	Cap. logistique vacc.	0.001 / jour	Estimation

Annexe — Double seuil adaptatif (détail)

Confinement gradué

$$u_{\text{conf}}(t) = \begin{cases} 0.75 & \text{si } I(t) > 0.02 \\ 0.25 & \text{si } 0.005 < I(t) \leq 0.02 \\ 0 & \text{sinon} \end{cases}$$

Vaccination adaptée

$$u_{\text{vacc}}(t) = \begin{cases} 1 & \text{si } S(t) > 0.3 \\ 0.25 & \text{sinon} \end{cases}$$

Principe : Confinement progressif pour limiter le coût économique + vaccination concentrée quand S est grand, réduite ensuite.