PETIT DÉJEUNER DE LA CHAIRE DIALOG

Allocation d'actifs averses au risque dans un contexte de changement climatique avec apprentissage par renforcement et modèles de Markov cachés

Le 25 septembre 2025

https://hal.science/hal-05142078

Etienne RAYNAL
travail en collaboration avec Stéphane Loisel

LABORATOIRE SAF SCIENCES ACTUARIELLE & FINANCIÈRE

UNIVERSITÉ DE LYON | Université Claude Bernard Lyon 1

//galea

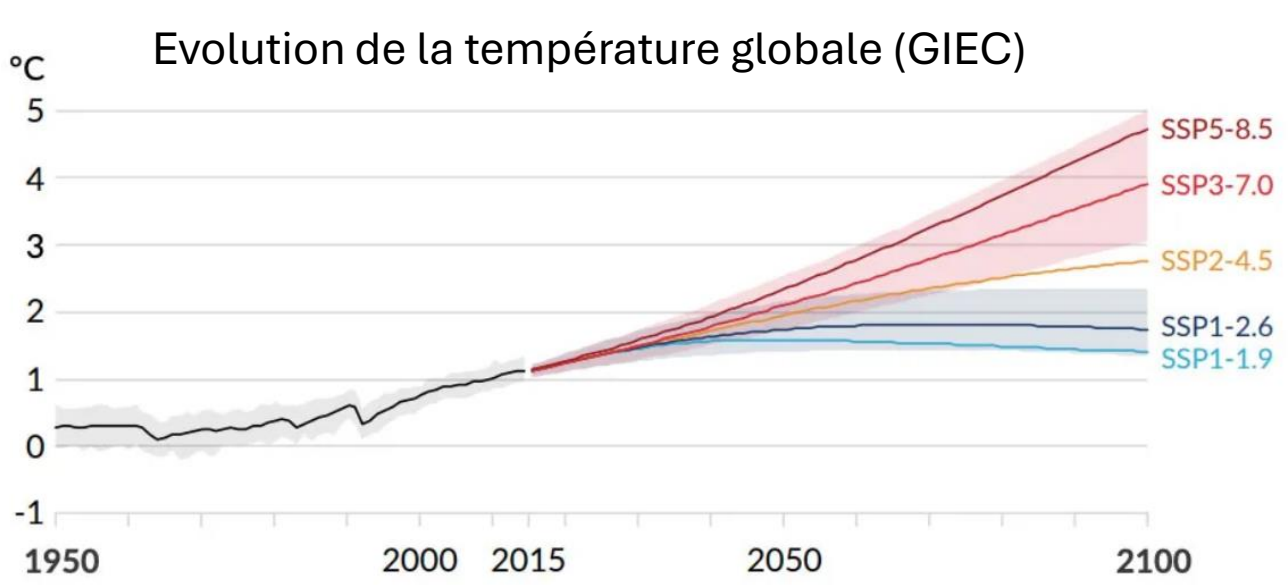
PLAN DE LA PRÉSENTATION

1. Disposer de meilleurs outils de gestion du risque climatique : une nécessité actuarielle
2. Contributions de la thèse
3. Allocation d'actifs, apprentissage par renforcement, modèles de Markov cachés et risque climatique
4. Conclusion et perspectives de recherche

1. **Disposer de meilleurs outils de gestion du risque climatique : une nécessité actuarielle**
2. Contributions de la thèse
3. Allocation d'actifs, apprentissage par renforcement, modèles de Markov cachés et risque climatique
4. Conclusion et perspectives de recherche

DISPOSER DE MEILLEURS OUTILS DE GESTION DU RISQUE CLIMATIQUE : UNE NÉCESSITÉ ACTUARIELLE

DISPOSER DE MEILLEURS OUTILS DE GESTION DU RISQUE CLIMATIQUE : UNE NÉCESSITÉ ACTUARIELLE



// Pluie diluvienne à Valence.

// Sécheresse au Brésil.

// Les conséquences du dérèglement climatique sont de plus en plus fréquentes et intenses, et le seront encore plus à mesure que la concentration en gaz à effet de serre augmente dans l'atmosphère.

// La crise climatique affecte déjà l'ensemble de l'activité d'assurance.

// La transition est également une source d'incertitude radicale.

DISPOSER DE MEILLEURS OUTILS DE GESTION DU RISQUE CLIMATIQUE : UNE NÉCESSITÉ ACTUARIELLE

Caractéristiques du risque climatique	Conséquence sur le métier d'actuaire
Le risque passé ne reflète pas le risque futur	Besoin de nouvelles données : données issues de modèles climatiques et intégrés
Les modèles actuariels classiques n'intègrent pas le risque climatique	Besoin de nouveaux modèles ou de modifications des modèles existants
L'incertitude croit de manière exponentielle et les enjeux s'étalent sur du long terme	Besoin d'un nouveau cadre conceptuel : restriction de l'étude à un petit échantillon de trajectoires plausibles
Les risques ont une hétérogénéité géographique croissante	Besoin de données plus fines à l'échelle locale
Les effets du changements climatiques peuvent être indirects en interaction systémique	Besoin d'une ouverture de la discipline actuarielle sur les autres sciences : climat, santé, …

1. Disposer de meilleurs outils de gestion du risque climatique : une nécessité actuarielle
2. **Contributions de la thèse**
3. Allocation d'actifs, apprentissage par renforcement, modèles de Markov cachés et risque climatique
4. Conclusion et perspectives de recherche

CONTRIBUTIONS DE LA THÈSE

CONTRIBUTIONS DE LA THÈSE

Chapitre 2 : Une analyse ERM (*Enterprise Risk Management*) des stress tests climatiques basés sur des scénarios en assurance.

Contributions :

- Détail de la chaîne complète de modélisation des stress tests climatiques basés sur des scénarios.
- Mise en valeur des incertitudes et des points faibles des modèles le long de cette chaîne de modélisation.
- Proposition d'un ensemble d'axes d'amélioration pour de prochains stress tests climatiques.

Article :

« Une analyse ERM des stress tests climatiques basés sur des scénarios en assurance », Raynal and Loisel, Revue Banque (accepté sous réserve de modifications mineures)

CONTRIBUTIONS DE LA THÈSE

Chapitre 3 : Allocation d'actifs averse au risque dans un contexte de changement climatique avec apprentissage par renforcement et modèles de Markov cachés.

Contributions :

- Intégration d'une chaîne de Markov cachée dans l'espace d'état d'un *Markov Decision Process* pour une stratégie d'allocation d'actifs.
- Mise en évidence des points forts et des limites d'une telle approche.
- Intégration du risque climatique à travers deux approches : une approche indicielle, une approche par scénario.

Article :

"*Risk averse asset allocation in a context of climate change with reinforcement learning and hidden markov models*" , Raynal and Loisel

- Conférence "P*erspectives on Actuarial Risks in Talks of Young Researchers*" , 2024,
- Conférence Lyon-Lausanne, 2024,
- *Annals of Finance* (soumis en mai 2025).

CONTRIBUTIONS DE LA THÈSE

Chapitre 4 : Modélisation de la mortalité pour l'évaluation des risques climatiques en France : impact des chaleurs extrêmes.

Contributions :

- Exploitations de données environnementales pour construite des régions plus homogènes en risque que des régions administratives.
- Modélisation qui combine modèles actuariels et modèles d'apprentissage statistique.
- Proposition d'un choc de mortalité, dans le climat actuel, adapté à des exercices prudentiels et reposant sur la littérature scientifique climatique.

Article :

"*Mortality modelling for short term climate stress test in France : impact of extreme heat*" , Raynal and Loisel, A*nnals of Actuarial Science* (soumis en mars 2025).

1. Disposer de meilleurs outils de gestion du risque climatique : une nécessité actuarielle
2. Contributions de la thèse
3. **Allocation d'actifs, apprentissage par renforcement, modèles de Markov cachés et risque climatique**
4. Conclusion et perspectives de recherche

ALLOCATION D'ACTIFS, APPRENTISSAGE PAR RENFORCEMENT, MODÈLES DE MARKOV CACHÉS ET RISQUE CLIMATIQUE

- Il est très difficile de modéliser explicitement la manière dont les risques climatiques se matérialiseront sur les marchés financiers.

- Les études actuelles se concentrent essentiellement sur :
  - Les secteurs les plus émetteurs,
  - Les tendances à long terme,
  - Les trajectoires issues des IAM.

- Il est toutefois possible de s'attendre à certaines conséquences possibles : crises financières systémiques se traduisant par des périodes de forte volatilité.

- Objectif de l'étude : construire des stratégies d'allocation d'actifs performantes à long terme ET résiliente pendant les périodes de crises financières.

- Cadre mathématique : un algorithme d'apprentissage par renforcement dont l'espace d'état est caractérisé par l'estimation que se fait l'agent du régime de marché dans lequel il se trouve. L'agent qui investit sur un actif risqué et un actif non risqué, fait l'hypothèse que l'actif risqué suit une dynamique *Regime Switching Log Normal* (Hardy (2001)).

// Description du modèle : modélisation financière

Un espace probabilisé $(\Omega, \mathcal{F} = \{\mathcal{F}_t, t \in \mathbb{N}\}, \mathbb{P})$

Un actif sans risque

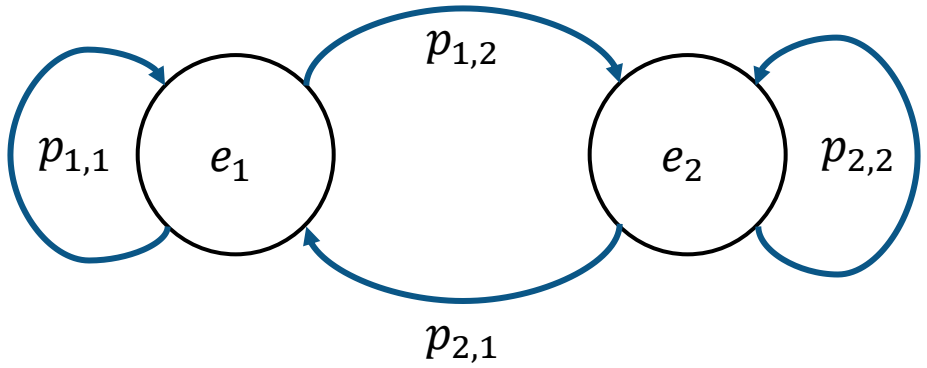
$$\begin{cases} B_t = B_{t-1}\exp(R_f) \\ B(t=0) = R_0 \end{cases}$$

Un actif risqué
(*Regime Switching Log-Normal, (RSLN)*)

$$\begin{cases} S_t = S_{t-1}\exp\left(\mu_1 - \frac{\sigma_1^2}{2} + \sigma_1\epsilon_t\right), si\ \rho_t = e_1 \\ S_t = S_{t-1}\exp\left(\mu_2 - \frac{\sigma_2^2}{2} + \sigma_2\epsilon_t\right), si\ \rho_t = e_2 \\ S(t=0) = S_0 \end{cases}$$

Une chaîne de Markov cachée

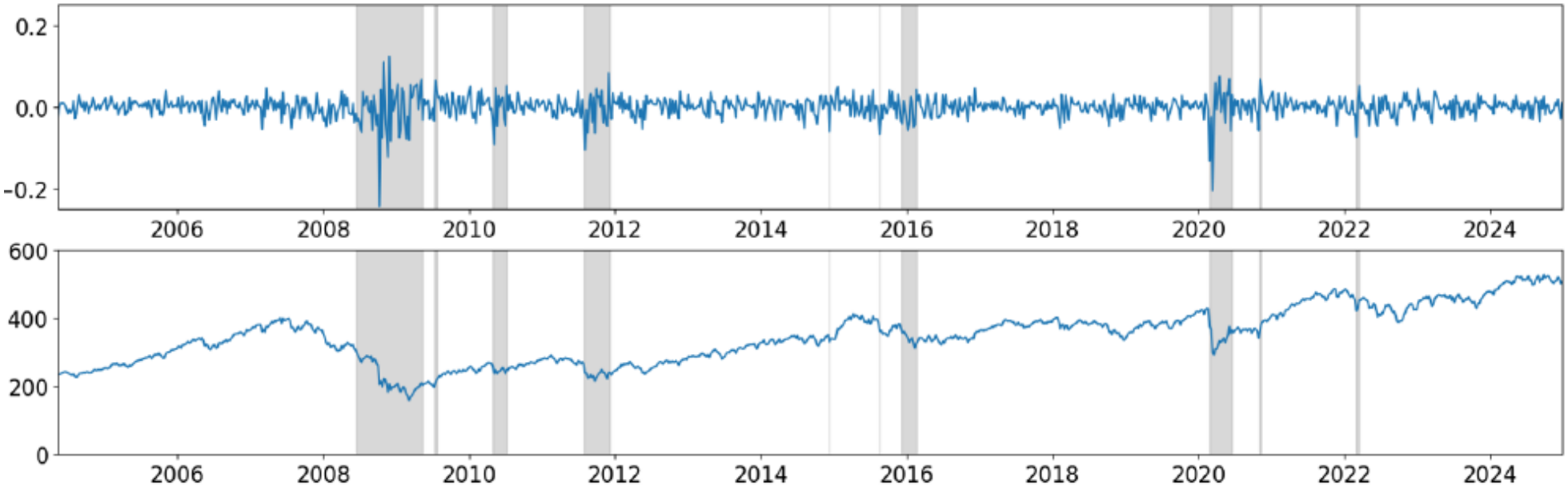
$$(\rho_t)_{t>0},\ \{e_1, e_2\},\ \mathrm{P} = \begin{bmatrix} p_{1,1} & p_{1,2} \\ p_{2,1} & p_{2,2} \end{bmatrix}$$



// Modèle RSLN sur les données de l'indice STOXXE600

Log-rendements

Valeurs de l'indice



$\rho_t = e_1$ $\rho_t = e_2$

$\mu_1 = 1{,}82e-03\,,\quad \sigma_1 = 3{,}29e-04$ $\quad p_{1,1} = 0{,}987$

$\mu_2 = -7{,}24e-03\,,\quad \sigma_2 = 2{,}97e-03$ $\quad p_{2,1} = 0{,}094$

// Justification de l'usage du modèle RSLN

Table 1 Comparison of Selection Information for STOXXE600 data

Model	Log Likelihood	AIC	BIC
RSLN 2 states	**2492**	**-4972**	**-4943**
GARCH	2468	-4929	-4909
ARCH	2371	-4738	-4723
BS	2322	-4640	-4630

// Un portefeuille de valeur Π_t composé d'un actif sans risque, en proportion τ_t, et d'un actif risqué. Rendement noté r_t^{Π} qui intègre des coûts de transaction.

Le problème d'optimisation :

$$\tau^* = \arg\max_{\tau} \mathbb{E}\left[\Pi_H^{\tau}\right]$$

$$\text{s.t.} \quad VaR_q\left(\mathbb{E}\left[\log\left(\frac{\Pi_{t+1}^{\tau}}{\Pi_t^{\tau}}\right)|\mathcal{F}_t\right]\right) \leqslant \lambda_{VAR} \quad \forall\ t$$

ALLOCATION D'ACTIFS, RL, HMM ET RISQUE CLIMATIQUE

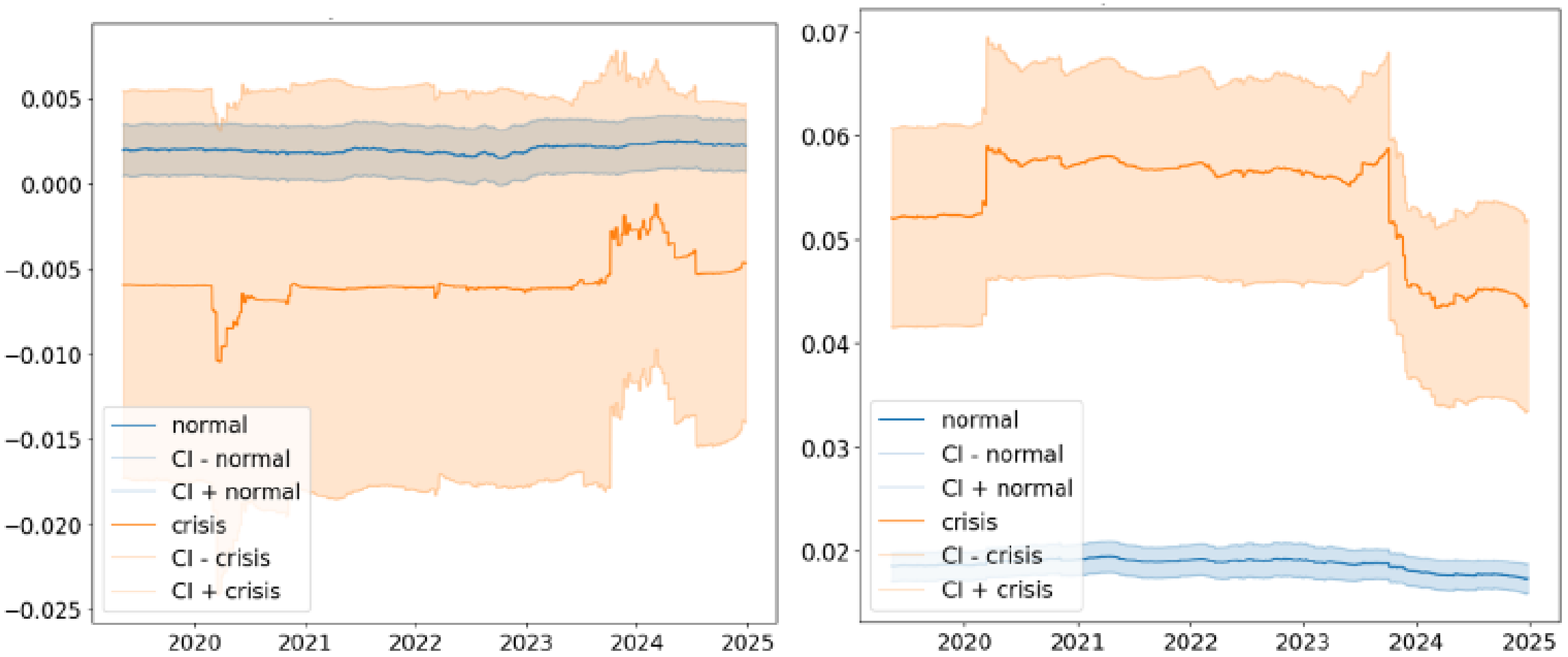


Illustration des paramètres estimés pour chaque état : e1 en bleu, et e2, l'état de crise, en orange.

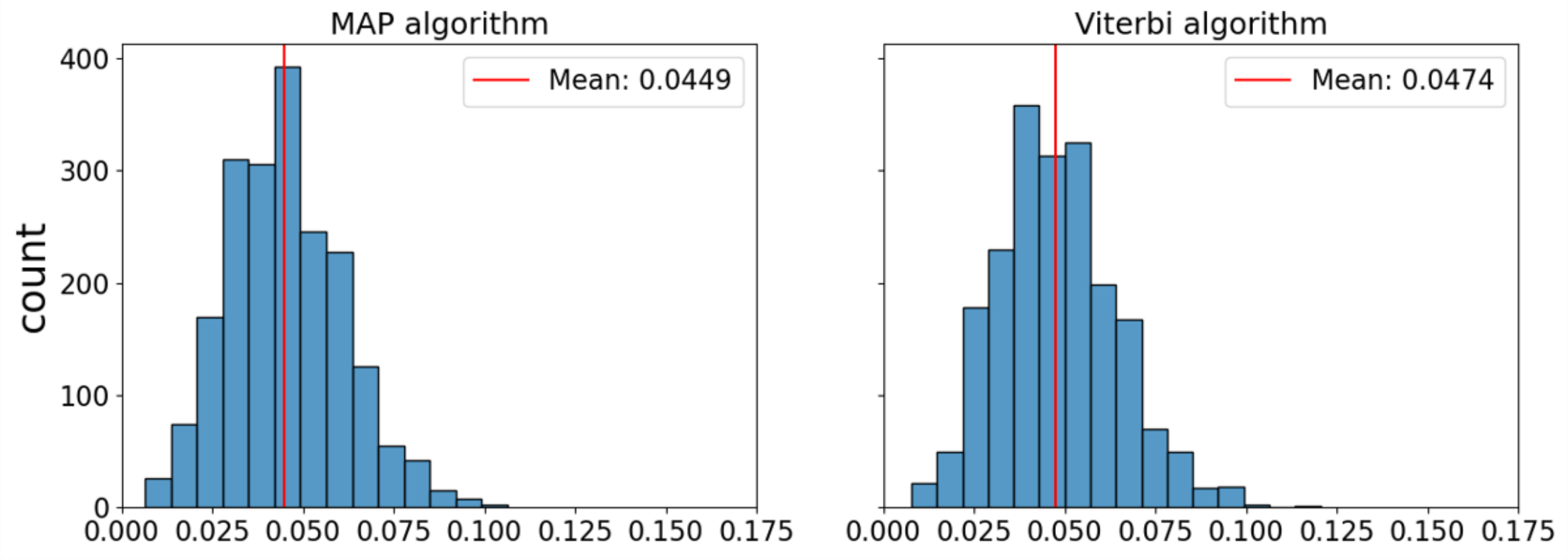
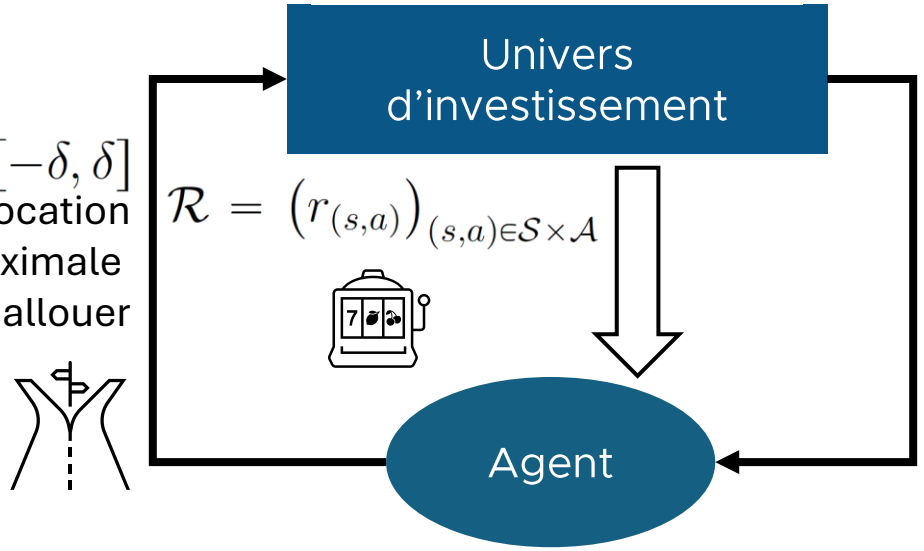


Illustration de la proportion d'erreurs faites par 2 algorithmes de décodage.

// Description du modèle : *Markov Decision Process* $(\mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{P})$

$$\mathcal{P} = (p(.|s,a))_{(s,a)\in\mathcal{S}\times\mathcal{A}}$$



$\mathcal{A} = [-\delta, \delta]$
Choix de réallocation
δ proportion maximale que l'on peut réallouer

$\mathcal{R} = (r_{(s,a)})_{(s,a)\in\mathcal{S}\times\mathcal{A}}$

$\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2$

$s_1 \in \mathcal{S}_1 = [0,1]$: Proportion d'actif sans risque

$s_2 \in \mathcal{S}_2 = [0,1]$: Probabilité $\mathbb{P}(\rho_t = e_1)$

Le problème d'optimisation : trouver une stratégie π qui permet de maximiser

$$Q_h^\pi(s,a) = E^\pi\left[\sum_{t=h}^{H} \gamma^{t-1} r_t(s_t,a_t) | \, s_h = s, a_h, a\right], \forall s, a$$

Dans notre cas la récompense est : $r_t = r_t^\Pi - \lambda\,[\,\widehat{VaR}(s_1, s_2, \mu_1, \sigma_1, \mu_2, \sigma_2) - \lambda_{VaR}]^+$

- Problème d'optimisation initial : $$\tau^* = \arg\max_{\tau} \mathbb{E}\left[\Pi_H^\tau\right]$$
$$\text{s.t.} \quad VaR_q\left(\mathbb{E}\left[\log\left(\frac{\Pi_{t+1}^\tau}{\Pi_t^\tau}\right)|\mathcal{F}_t\right]\right) \leqslant \lambda_{VAR} \quad \forall\, t$$

- Récompense dans le MDP :
$$r_t = \text{portfolio\_log\_return}_t - \lambda\left[\widehat{VaR}(s_1, s_2, \mu_1, \sigma_1, \mu_2, \sigma_2) - \lambda_{VaR}\right]^+$$

- Maximiser la somme des log-rendements est équivalent à maximiser la valeur finale du portefeuille.

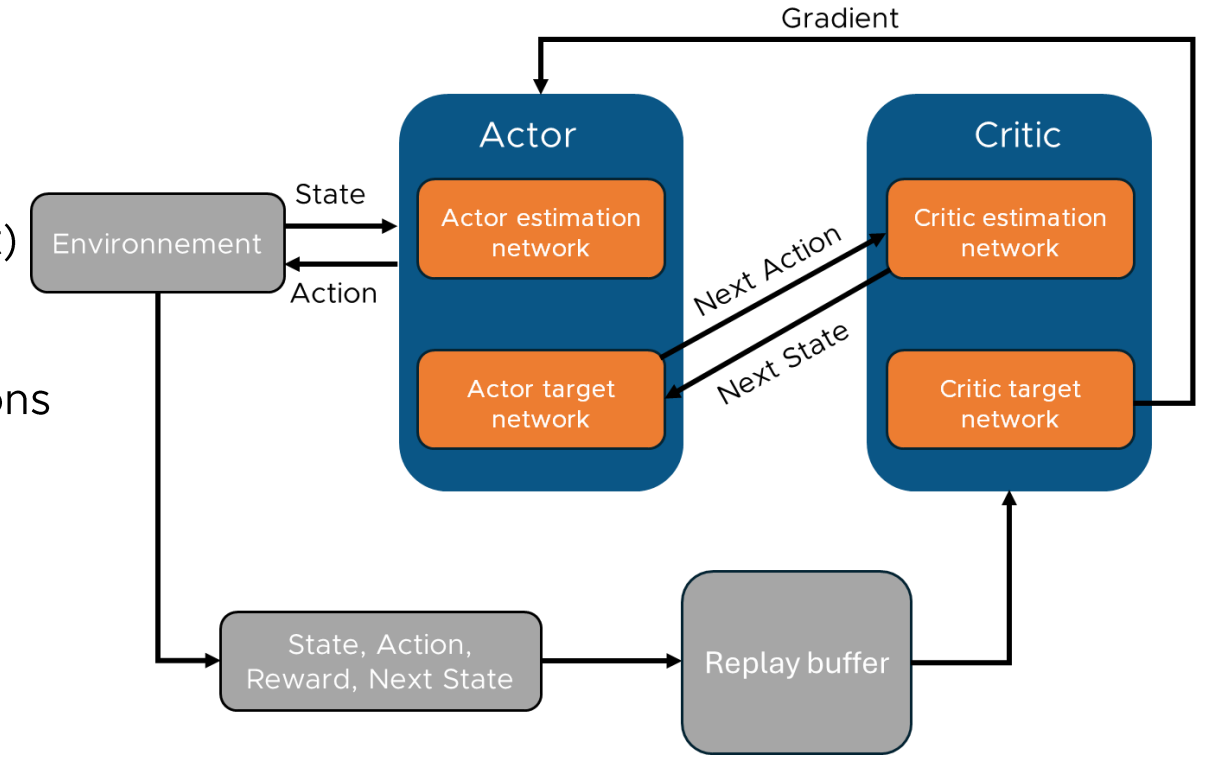
- Appliquer une pénalité avec un poids λ fournit un retour d'information régulier, permettant à l'agent d'apprendre progressivement le compromis risque/rendement à proximité de la limite de la contrainte.

- La récompense basée sur la pénalité se rapproche d'une relaxation lagrangienne du problème contraint avec un paramètre d'aversion au risque réglable
$$\max_{\tau}\left(\mathbb{E}\left[\Pi_H^\tau\right] - \lambda \max_t \left(VaR_q(r_t^\Pi) - \lambda_{VaR}\right)^+\right)$$

// Principes généraux :
- Algorithme actor-critic hors-politique (off-policy)
- Adapté à des espaces d'état continus
- Apprentissage d'une politique déterministe : $a = \mu(s)$
- Basé sur DQN (Deep Q-Network) et PG (Policy Gradient)

// Deux réseaux neuronaux principaux :
- Actor : apprend la politique $\mu(s \mid \theta^{\mu})$ qui génère les actions
- Critic : approxime la fonction $Q(s, a \mid \theta^{Q})$

// Mémoire d'expérience (Replay Buffer) :
- Stocke les transitions (s_t, a_t, r_t, s_{t+1})
- Permet un entraînement stable et échantillonné



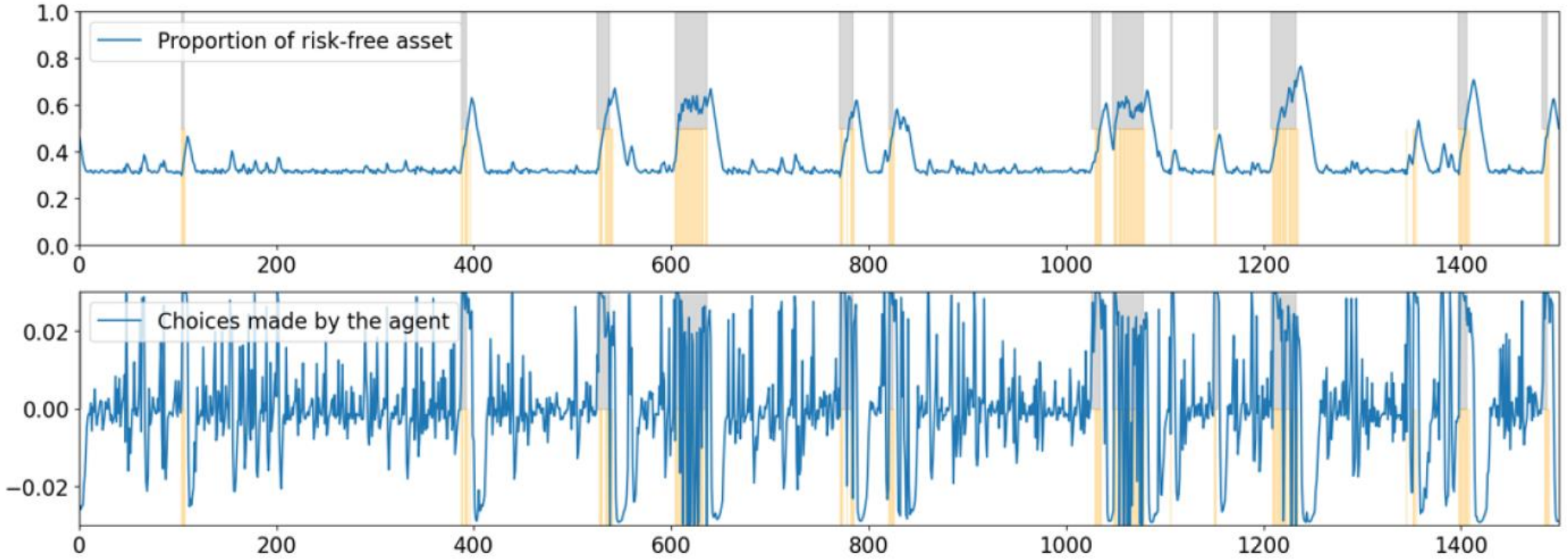
// Exploration avec bruit :
- Bruit ajouté à l'action pour explorer : $a_t = \mu(s_t) + \epsilon$ (par exemple un bruit d'Ornstein-Uhlenbeck pour la corrélation

// Cibles lentes (Target Networks) :
- Réseaux cibles pour actor et critic μ' et Q' avec mise à jour lente $\theta' = \tau\theta + (1 - \tau)\theta'$

// Résultats obtenus avec DDPG entraîné sur des données réelles, testé sur des données synthétiques

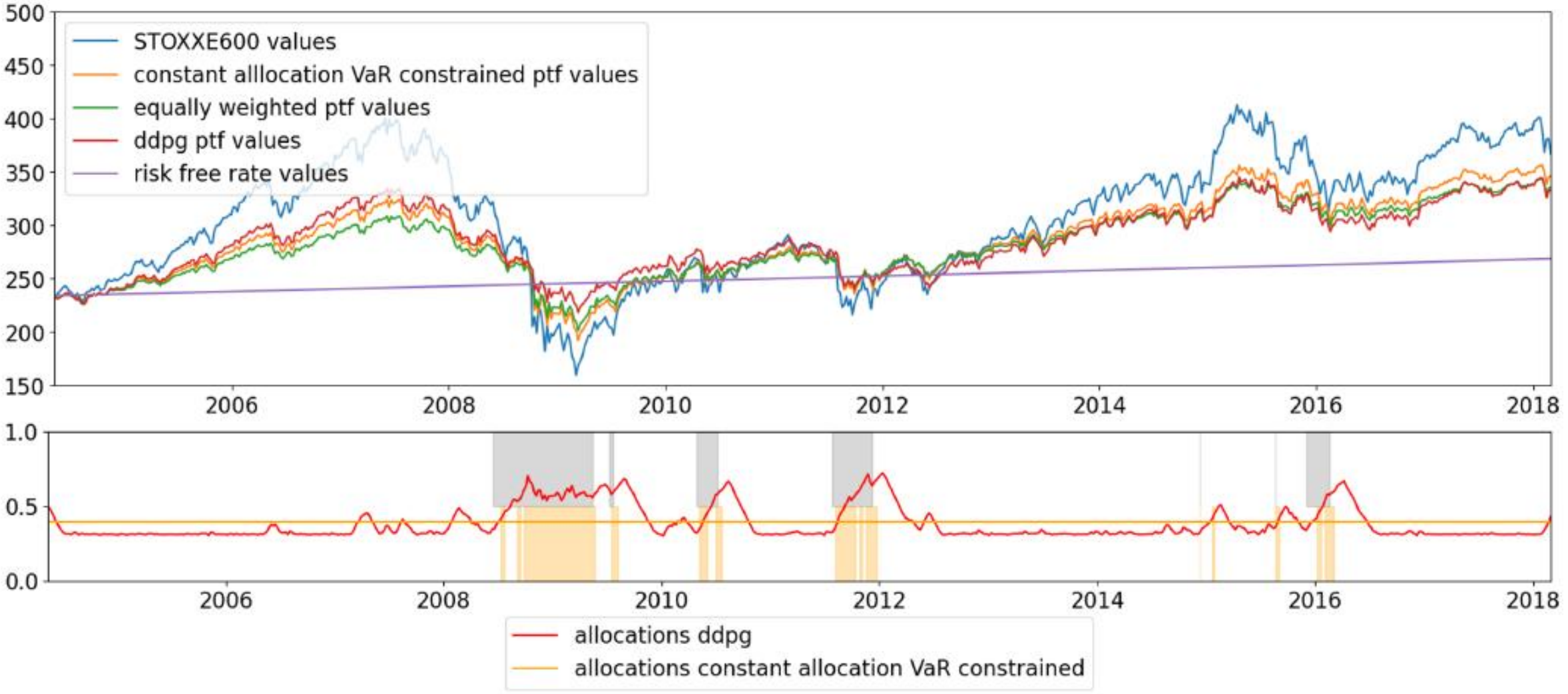
Proportion d'actif sans risque

Choix de réallocation



$\rho_t = e_1$ $\rho_t = e_2$ $\widehat{\rho_t} = e_2$

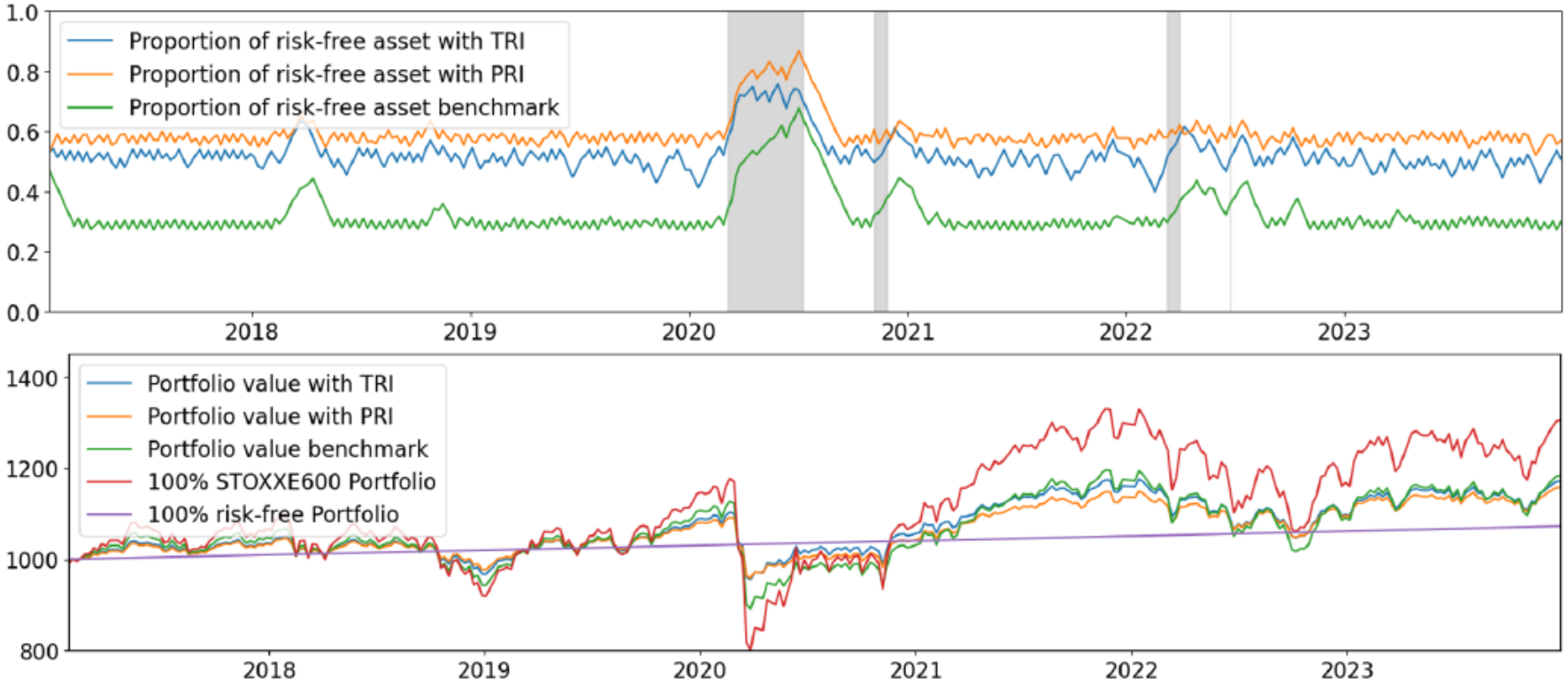
// Résultats obtenus avec DDPG entraîné sur des données réelles, testé sur des données réelles



$\rho_t = e_1$ $\rho_t = e_2$ $\hat{\rho}_t = e_2$

ALLOCATION D'ACTIFS, RL, HMM ET RISQUE CLIMATIQUE

// Intégration du risque climatique avec des indices $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2 \times \mathcal{S}_3$ s_3: valeur d'un indice climatique



Indices issus de (Bua et al. (2024))

TRI : Transition Risk Index

PRI : Physical Risk Index

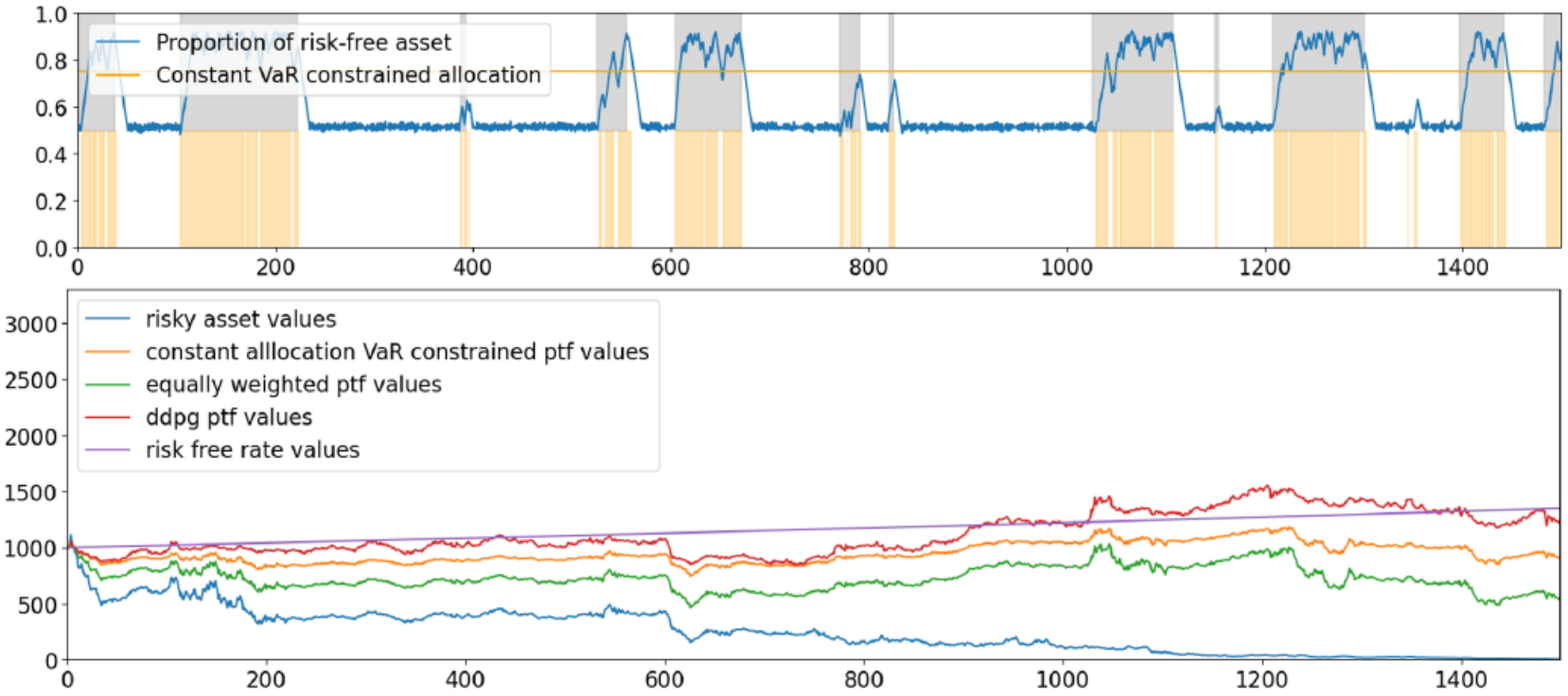
	PRI index	TRI index	Benchmark without climate index
Alpha	2.60e-4	2.25e-4	2.85e-4

Performance par rapport au benchmark calculé avec le α de Jensen

// Intégration du risque climatique avec une déformation des régimes de marché

- Scénario 1 : Amplification de l'état de crise : e_3 remplace e_2 avec $\mu_3 < \mu_2$ et $\sigma_3 > \sigma_2$
- Scénario 2 : Plus forte probabilité de rester dans l'état de crise : e_3 remplace e_2 avec $p_{3,1} < p_{2,1}$
- Scénario 3 : Etat de crise permanent

Illustration dans le cas du scénario 2



1. Disposer de meilleurs outils de gestion du risque climatique : une nécessité actuarielle
2. Contributions de la thèse
3. Allocation d'actifs, apprentissage par renforcement, modèles de Markov cachés et risque climatique
4. **Conclusion et perspectives de recherche**

CONCLUSION ET PERSPECTIVES DE RECHERCHE

CONCLUSION

Conclusion

// L'objectif de l'étude était de montrer comment construire des stratégies d'allocation d'actif à long terme.

// Les dynamiques financières sont pour l'instant décorrélées des nouvelles sur le climat, mais cela pourrait changer.

// L'agent superforme les stratégies classiques lorsque les régimes de marché évoluent. Nous avons mis en évidence les capacités d'adaptation d'un agent entraîné par renforcement.

Perspectives de recherche

// Méthodologie pour mieux calibrer les hyperparamètres.

// Utilisation d'autres processus stochastiques

// Améliorer les stratégies pour mieux pallier au décalage entre les régimes.

// Tests avec d'autres indices pouvant refléter les risques climatiques.

// Autres algorithmes de renforcement.

ANNEXES

ANNEXE : DDPG

Algorithme 2 : Deep Deterministic Policy Gradient (DDPG)

Randomly initialize critic network $Q(s, a|\theta^Q)$ and actor $\mu(s|\theta^\mu)$ with weights θ^Q and θ^μ.;

Initialize target network Q' and μ' with weights $\theta^{Q'} \leftarrow \theta^Q$, $\theta^{\mu'} \leftarrow \theta^\mu$;

Initialize replay buffer R;

for *each episode* **do**

  Initialize a random process $\mathcal{N}$ for action exploration;

  Receive initial observation state s_0;

  for *each time step t* **do**

    Select action $a_t = \mu(s_t|\theta^\mu) + \mathcal{N}_t$ according to the current policy and exploration noise Execute action a_t and observe reward r_t and observe new state s_{t+1} Store transition (s_t, a_t, r_t, s_{t+1}) in R Sample a random minibatch of N transitions (s_i, a_i, r_i, s_{i+1}) from R Set :

$$y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'}) \tag{3.9}$$

ANNEXE : DDPG

Update critic by minimizing the loss:

$$L = \frac{1}{N}\sum_i (y_i - Q(s_i, a_i|\theta^Q))^2 \tag{3.10}$$

Update the actor policy using the sampled policy gradient:

$$\nabla_{\theta^\mu} J \approx \frac{1}{N}\sum_i \nabla_a Q(s, a|\theta^Q)|_{s=s_i, a=\mu(s_i)} \nabla_{\theta^\mu}\mu(s|\theta^\mu)|_{s_i}$$

Update the target networks:

$$\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}$$

$$\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}$$

end

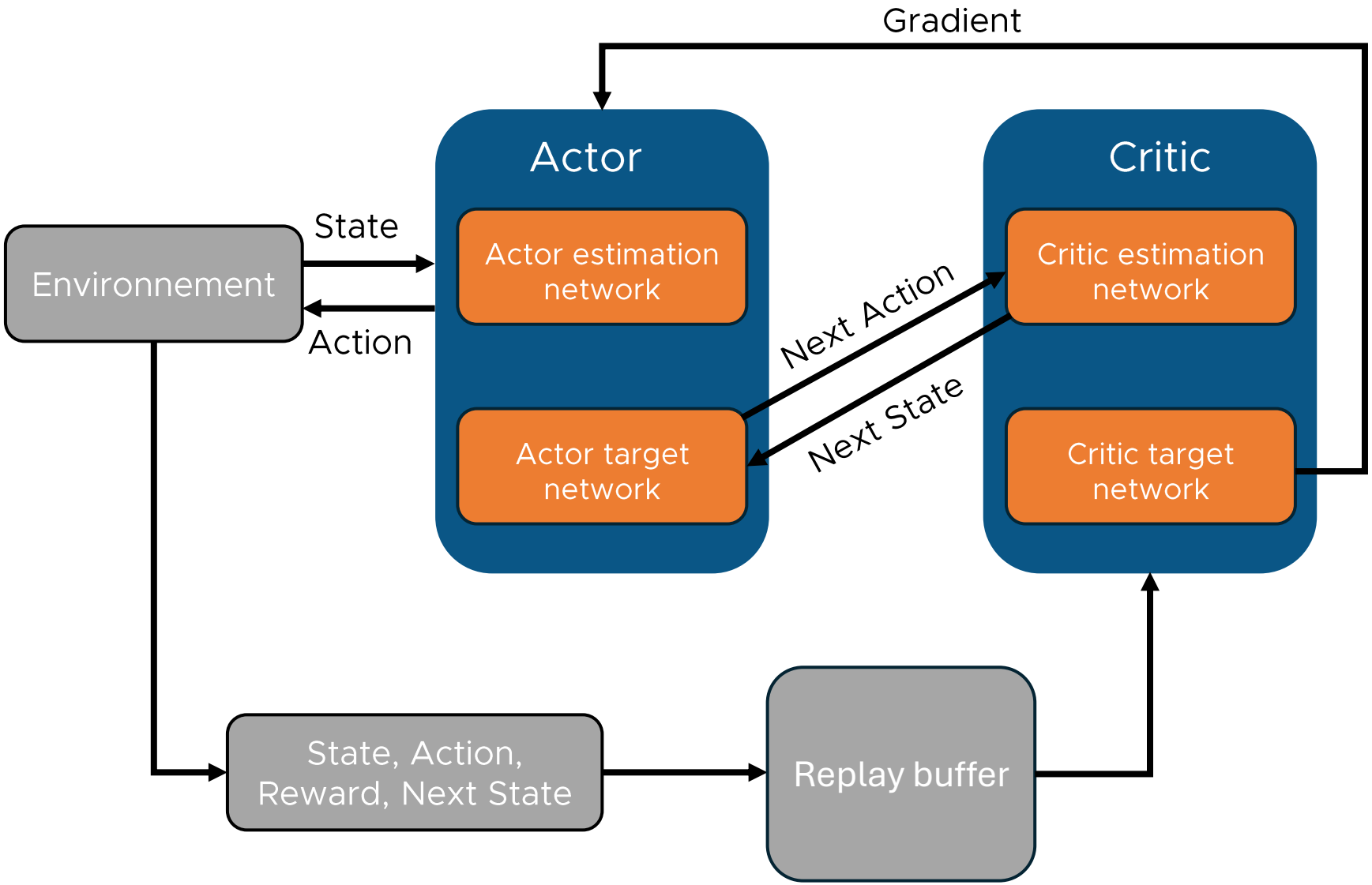
end

Source : Lillicrap et al. (2015)

ANNEXE : DDPG

Réseaux neuronaux

- taux d'apprentissage de l'acteur et du critique
- taille et nombre de couches cachées des réseaux
- fonction d'activation (ReLU, tanh, etc.)


Gradient
Actor
Critic
State
Environnement
Actor estimation network
Critic estimation network
Next Action
Action
Next State
Actor target network
Critic target network
State, Action, Reward, Next State
Replay buffer