



**HAL**  
open science

# Développement d'un protocole de comparaison d'algorithmes d'apprentissage par renforcement pour la gestion optimale temps-réel de flux de puissance dans un micro-réseau électrique

Rima Oulhaj, Pierre Garambois, Lionel Roucoules

## ► To cite this version:

Rima Oulhaj, Pierre Garambois, Lionel Roucoules. Développement d'un protocole de comparaison d'algorithmes d'apprentissage par renforcement pour la gestion optimale temps-réel de flux de puissance dans un micro-réseau électrique. S-MART 2023: 18ème Colloque national S.mart, Arts et Métiers Paristech ENSAM Aix-en-Provence, Université de Toulon [UTLN], Apr 2023, Carry-le-Rouet, France. hal-04278704

**HAL Id: hal-04278704**

**<https://hal.science/hal-04278704v1>**

Submitted on 10 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Développement d'un protocole de comparaison d'algorithmes d'apprentissage par renforcement pour la gestion optimale temps-réel de flux de puissance dans un micro-réseau électrique

R. Oulhaj <sup>a,\*</sup>, P. Garambois <sup>a</sup>, L. Roucoules <sup>a</sup>

<sup>a</sup> Arts et Métiers Institut de Technologie, HESAM Université, LISPEN, 13617 Aix-en-Provence, France

\* e-mail : rima.oulhaj@ensam.eu

## 1. INTRODUCTION

Les micro-réseaux électriques ont permis l'intégration de sources énergétiques distribuées, ainsi que l'amélioration de l'autosuffisance énergétique et de la pénétration d'énergies renouvelables grâce à l'utilisation de systèmes de stockage et la possibilité de gestion et de prise de décision indépendantes. En effet, l'installation d'un micro-réseau électrique dans un territoire donné permet à celui-ci la gestion des producteurs et des stockeurs locaux de manière à maximiser les intérêts environnementaux (e.g., diminution de l'impact carbone), sociaux (e.g., équité énergétique) et économiques (e.g., coûts énergétiques) de ce territoire. Il s'agit donc d'assurer une gestion (i.e., un pilotage) des technologies du micro-réseau pouvant maximiser ces intérêts tout en garantissant un approvisionnement de qualité. L'utilisation d'algorithmes d'apprentissage-machine est un point focal dans les travaux de recherche sur cette problématique, notamment en optimisation temps-réel de flux de puissance, et l'un des défis importants soulignés par les auteurs de cet article a été la difficulté de comparaison des algorithmes utilisés dans lesdits travaux puisque l'on doit réimplémenter (recoder et donc déboguer) les modèles de micro-réseaux qui y sont inclus afin d'émuler leurs résultats. De plus, il est nécessaire d'interfacer ces modèles aux algorithmes spécifiques afin de prendre en compte la terminologie adaptée, e.g., pour des algorithmes d'apprentissage par renforcement (AR) il faut coder les processus de transition et de récompense (plus de détails sont présents dans les parties 3.1.1, 3.1.2 et 3.1.3). Les travaux existants sur la normalisation de modèles de micro-réseaux incluent le CIGRE Task Force

C6.04.02 et le CIGRE Working Group 14.02 [2, 3, 4] ainsi que le IEEE European LV microgrid benchmark network [5]. De plus, les systèmes de bus de IEEE peuvent être considérés comme une normalisation de réseaux réels (le bus IEEE-14 [6] représente une approximation du système électrique américain).

Dans cet article, nous tentons de développer un protocole d'évaluation et de comparaison ou « benchmark » afin de comparer des algorithmes d'AR appliqués au problème d'optimisation temps-réel de flux de puissance dans un micro-réseau électrique. Pour cela, et au vu du développement d'un outil informatique permettant à différents utilisateurs d'effectuer cette comparaison, nous avons comme perspective de proposer différentes configurations « intégrées » de micro-réseaux ainsi que des outils spécifiques à la simulation du processus de décision markovien lié au problème d'optimisation afin d'implémenter les algorithmes d'AR. Un schéma de cet outil est donné dans la figure 3.

## 2. FORMALISATION DU BENCHMARK

Au vu du développement de notre protocole de comparaison d'algorithmes d'AR pour la résolution du problème d'optimisation temps-réel de flux de puissance dans un micro-réseau électrique, il est important de définir les éléments suivants [1] :

### Objet d'étude

Notre objet d'étude, ou objet de comparaison, correspond aux algorithmes d'AR dont on étudie la performance dans la résolution du problème spécifique

d'optimisation temps-réel de flux de puissance dans un micro-réseau électrique, ce problème étant formalisé sous forme de processus décisionnel de Markov (section 3.1.1). On s'intéresse donc à comparer leur capacité à effectuer une tâche précise. Notons qu'il s'agit ici du problème temps-réel, ce qui signifie que durant la phase de test, les algorithmes ont le droit à un seul passage sur la période de simulation (un seul épisode) et que les données de consommation et de production renouvelable non contrôlable sont différentes de celles utilisées dans la phase d'apprentissage.

Tout algorithme est défini par une donnée d'entrée, une donnée de sortie et par l'ensemble d'instructions à exécuter et qui, pour des raisons de lisibilité, sont généralement formalisées sous forme d'un pseudo-code. Plusieurs classes d'algorithmes nécessitent en plus de ces éléments le choix d'« hyperparamètres » régissant le fonctionnement d'une ou plusieurs instructions. Dans le cas des algorithmes d'AR, les instructions varient selon la classe de l'algorithme : on parle d'algorithmes d'AR « off-line » ou « on-line » et « on-policy » ou « off-policy ». Les hyperparamètres comprennent notamment le nombre d'itérations d'apprentissage, la longueur des épisodes, le coefficient d'actualisation dans l'équation de Bellman, etc. Dans le cas des algorithmes d'AR « profonds » qui font usage de réseaux de neurones, les hyperparamètres comprennent également la configuration de ces réseaux (nombre et connexité des couches, fonctions d'activation, etc.).

### Caractéristiques de l'objet d'étude

On s'intéresse à mesurer et à comparer les caractéristiques d'aptitude à réaliser la tâche (récompense cumulée à l'issue de l'optimisation de flux), de coût de calcul en termes de temps et de taille de données nécessaires à l'apprentissage ainsi que de fiabilité. Shengren et al. [17] comparent des algorithmes d'AR pour le pilotage de micro-réseaux, mais ils s'intéressent uniquement à la performance en termes d'aptitude à réaliser la tâche (récompense cumulée).

### Mesures

Des métriques permettant la quantification des caractéristiques étudiées ont été choisies pour comparer les algorithmes d'AR appliqués à notre problème d'optimisation (section 3.2).

### Collections expérimentales

Il s'agit des configurations de micro-réseau (technologies et leurs paramètres) qui feront l'objet des simulations de flux de puissance et qui seront pilotées à l'aide des algorithmes d'AR que l'on souhaite comparer. Ces simulations reposent sur une modélisation que nous proposons. Un exemple est présent dans la section 3.3.1.

### Jeux de données

Les jeux de données consistent en des profils de demande et des données météorologiques régissant la génération d'énergie renouvelable non contrôlable (section 3.4).

## 3. PROTOCOL EXPERIMENTAL

### 3.1. Objet d'étude : qu'est-ce que l'apprentissage par renforcement

#### 3.1.1. Le processus de décision markovien

Dans notre problème de gestion temps-réel de flux de puissance on souhaite déterminer pour chaque pas de temps les flux de puissance optimaux, comme la puissance à charger/décharger de chaque stockeur, la puissance à générer par les producteurs pilotables, etc. Il s'agit de maximiser une fonction de récompense à l'issue de la période de simulation, plus ou moins longue, en choisissant à chaque pas de temps l'action adéquate. Ce choix ne se base pas uniquement sur la récompense immédiate mais bien sur celle cumulée tout au long de la période de simulation en question. Chaque action a un impact sur l'état interne du système (ex : les niveaux d'énergie des stockeurs). Nous pouvons traduire ce problème par un Processus de Décision Markovien (Markov Decision Process ou MDP).

Un MDP (figure 1) est un tuple  $(S, A, T, R, \gamma)$  où  $S$  est l'espace des états du système pouvant être discret ou continu (dans le cas des micro-réseaux électriques les états sont continus puisqu'il s'agit de quantités physiques comme l'état de charge des stockeurs). L'élément  $A$  correspond à l'espace des actions qui lui aussi peut être discret (l'agent choisit ses actions dans une liste discrète) ou continu. Le dernier cas correspond à la gestion de flux de puissance puisqu'une action est un choix de valeurs de ces différents flux.  $A$  et  $S$  sont donc des sous-ensembles de deux espaces euclidiens multi-dimensionnels:  $A \subset \mathbb{R}^{D_A}$  et  $S \subset \mathbb{R}^{D_S}$ ,  $D_A$  et  $D_S$  étant les dimensions des espaces  $A$  et  $S$  respectivement. La fonction de transition  $T(s, a, s')$  donne la probabilité de transition à l'état  $s'$  lorsque l'action  $a$  est choisie à l'état  $s$ . La fonction  $R(s, a, s')$  représente la récompense obtenue lorsqu'on arrive à l'état  $s'$  en choisissant l'action  $a$  à l'état  $s$ . Si  $T$  et/ou  $R$  ne sont pas déterministes alors le MDP est dit stochastique. C'est cette stochasticité des transitions et des récompenses qui appelle à l'utilisation d'algorithmes d'AR.

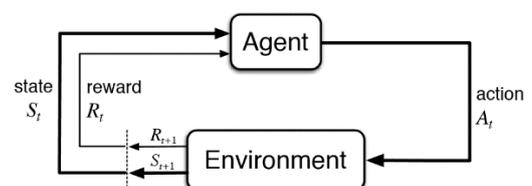


Figure 1 – Processus décisionnel markovien

### 3.1.2. Analogie avec le problème de gestion optimale de flux de puissances

Une analogie est possible entre le MDP tel que nous l'avons formalisé et le problème de gestion de flux de puissance dans un micro-réseau. Le système dont les états sont observés dans le MDP est ici les technologies que l'on souhaite piloter. Limitons-nous au cas où les seules technologies pilotables représentent une collection de stockeurs (micro-réseau de la figure 2).

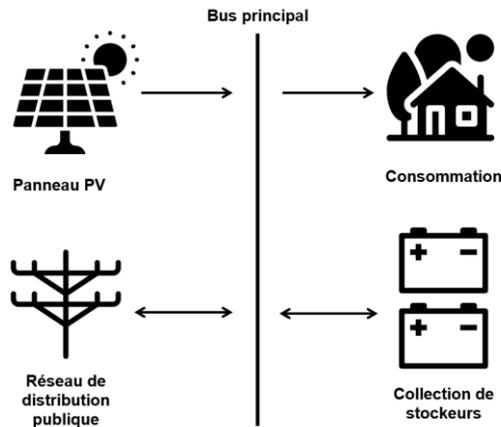


Figure 2 - Vanilla Microgrid

**Espace d'états du système** : un état du système, qui est dans notre exemple un micro-réseau comprenant une seule technologie pilotable (des stockeurs), est donné par les niveaux de charge de ceux-ci ( $\mathbf{W}_t^s$ )<sub>s∈S</sub> (d'autres éléments peuvent être considérés comme le niveau d'usure) ainsi que la génération renouvelable non pilotable  $RE_t$  (panneau PV) et les flux tarifaires du réseau de distribution publique :  $GP_t$  (prix d'achat) et  $FIT_t$  (prix de rachat) ;

**Espace d'actions et transitions** : une action  $a_t$  est une puissance de charge ou de décharge des stockeurs choisie à l'instant  $t$ . Cette action est contrainte par le niveau de charge des stockeurs, ce qui veut dire que l'intervalle des actions possibles à chaque instant  $t$  est contraint par l'état du système  $s_t$ . L'espace des actions est donc fonction de l'état du système et l'on a  $\mathbf{A}(s) = \{a \in \mathbb{R} \mid \mathbf{T}(s, a) \in \mathbf{S}\}$ . La fonction de transition  $\mathbf{T}(s, a)$  comprend une partie déterministe (charge/décharge) et une partie stochastique (flux tarifaires et génération renouvelable non contrôlable) ;

**Récompense** : La récompense peut être donnée par la somme de tous critères (souhaitables) d'intérêt dans notre cas d'étude qui peuvent être évalués après chaque transition. Elle est calculée pour chaque pas de temps, et donc pour chaque transition (voir fonctions-objectifs dans la section 3.3.2.). Comme la récompense est influencée par les flux tarifaires et la génération renouvelable non contrôlable, elle est également stochastique, d'où la nécessité d'algorithmes d'AR.

### 3.1.3. Les équations d'optimalité

Dans le cas où  $\mathbf{A}$  et  $\mathbf{S}$  sont continus, une politique de sélection d'actions  $\pi(s)$  est une fonction de densité de probabilité sur  $\mathbf{S}$  qui correspond à la probabilité de sélection d'une action étant donné l'état courant.

Dans un MDP, un problème de contrôle correspond à trouver la politique de sélection d'actions  $\pi$  qui maximise la récompense future actualisée. Les équations de Bellman nous donnent les fonctions de valeur (c'est-à-dire la récompense future actualisée) optimales à un état  $s$  en choisissant une action  $a$  :

$$Q^*(s, a) = \int_{\mathbf{S}} \mathbf{T}(s, a, s') (\mathbf{R} + \gamma \max_{a'} Q^*(s', a')) ds'$$

Où  $\gamma$  est le coefficient d'actualisation. Intuitivement, la fonction  $Q^*: \mathbf{S} \times \mathbf{A} \rightarrow \mathbf{R}$  donne la valeur obtenue si en étant en  $s$  et en choisissant  $a$ , on suit les actions optimales à tous les états futurs (d'où le terme  $\max_{a'} Q^*(s', a')$ ). La seconde équation de Bellman concerne la fonction de valeur  $V^*: \mathbf{S} \rightarrow \mathbf{R}$  qui calcule la récompense future actualisée optimale étant donné l'état actuel. Dans ce cas, l'équation d'optimalité s'écrit :

$$V^*(s) = \max_a \int_{\mathbf{S}'} \mathbf{T}(s, a, s') (\mathbf{R}(s, a, s') + \gamma V^*(s')) ds'$$

Des exemples d'algorithmes d'AR résolvant ce problème d'optimalité, notamment des algorithmes « profonds » utilisant des réseaux de neurones sont présentés dans [7, 8, 9, 10]

## 3.2. Mesures

Nous avons identifié un nombre de critères d'évaluation afin de comparer les algorithmes d'AR appliqués à notre problème d'optimisation temps-réel de flux de puissance. Le tableau 2 en donne le détail.

## 3.3. Collections expérimentales : les micro-réseaux électriques

### 3.3.1. Modélisation

On considère un modèle simplifié du micro-réseau électrique (figure 2) que l'on va nommer VanillaMicrogrid (VMG) et dans lequel le souci principal est celui d'assurer la gestion de flux de puissance entre les différents composants du micro-réseau. On ne se soucie donc pas de la régulation de fréquence et de voltage, puisqu'on considère que le micro-réseau est l'agrégation de composants qui injectent de la puissance dans le réseau et d'autres qui en puisent. Les générateurs renouvelables (panneaux PV) sont non-contrôlables puisque leur flux de puissance injecté dans le VMG dépend de l'irradiation solaire. La demande est modélisée par un niveau de puissance demandé à chaque pas de temps. Le stockage

est modélisé par un système à flux de puissance bidirectionnel, ce qui signifie que celui-ci peut à la fois injecter et puiser la puissance du réseau. Le problème revient à planifier la puissance à charger/décharger pour chacun des stockeurs et celle à puiser / réinjecter dans le réseau de distribution publique (le VMG fonctionne en mode connecté).

### 3.3.2. Problème d'optimisation de flux de puissance dans le VMG

Dans le problème d'optimisation des flux de puissance dans le VMG, les flux de puissance du stockage et du réseau de distribution publique sont les variables que l'on souhaite optimiser. Cependant, comme les flux de puissance dans le VMG sont toujours équilibrés, les flux de puissance du réseau de distribution peuvent être déduits de ceux du stockage. Les deux constitueraient donc des variables redondantes. Les seules variables retenues sont donc les flux de puissance du stockage.

#### Variables de décision :

Pour chaque système de stockage  $s$  et pour chaque pas de temps  $t$  soit  $x_t^s$  la puissance de charge/décharge de  $s$ . Les variables de décision sont  $(x_t^s)_{t \in [0, \dots, T], s \in S}$  où  $T$  est le nombre total de pas de temps et  $S$  l'ensemble de stockeurs dans le VMG. Puisque par choix de notre part on ne peut pas charger et décharger les stockeurs sur le même pas de temps,  $x_t^s$  est positive si le stockeur est en décharge et négative s'il est en charge.

#### Contraintes :

Les contraintes sont relatives à la capacité des stockeurs. Pour chaque stockeur  $s$  et pour chaque pas de temps  $t$  soit  $W_t^s$  l'énergie stockée dans  $s$ . Nous pouvons obtenir  $W_t^s$  à partir de  $W_{t-1}^s$  en appliquant l'équation :

$$W_t^s = W_{t-1}^s * (1 - \sigma_{SD}^s) + (P_{C,t}^s * \eta_C^s - \frac{P_{D,t}^s}{\eta_D^s}) * \Delta t$$

Où  $\sigma_{SD}^s$  est le coefficient d'autodécharge,  $P_{C,t}^s$  est la puissance de charge,  $P_{D,t}^s$  est la puissance de décharge,  $\eta_C^s$  est le coefficient d'efficacité de charge et  $\eta_D^s$  le coefficient d'efficacité de décharge. Par définition :

$$P_{D,t}^s = \max(0, x_t^s) \text{ et } P_{C,t}^s = -\min(0, x_t^s)$$

On définit l'état de charge par :

$$SOC_t^s = \frac{W_t^s}{W_{NC}^s}$$

Où  $W_{NC}^s$  est la capacité nominale du stockeur. La contrainte sur  $SOC_t^s$  est la suivante :

$$SOC_{\min}^s \leq SOC_t^s \leq SOC_{\max}^s$$

Les SOC minimum et maximum sont des caractéristiques du stockeur. La résolution du système composé de ces deux contraintes peut être traduite par une seule inéquation :  $x_t^{s,\min} \leq x_t^s \leq x_t^{s,\max} \forall s, t$

Cette inéquation est généralisable à toute configuration de micro-réseau. Si celui-ci inclut des générateurs contrôlables alors les variables de décision comprennent les puissances de génération de ceux-ci est des contraintes éventuelles s'ajoutent au problème.

#### Fonctions-objectifs :

Le choix des fonctions-objectifs dépend de la partie prenante qui souhaite piloter le micro-réseau et peut concerner les intérêts économiques, environnementaux ou sociaux. Considérons la fonction  $F_1$  qui sert à maximiser l'intégration de l'énergie renouvelable :

$$\max_{(x_t^s)_{t,s}} F_1((x_t^s)_{t,s}) = \max_{(x_t^s)_{t,s}} \sum_{t=0}^T RE_t - GL_t^+$$

Et la fonction  $F_2$  qui sert à minimiser le coût énergétique et à maximiser le gain de revente d'énergie :

$$\min_{(x_t^s)_{t,s}} F_2 = \min_{(x_t^s)_{t,s}} \sum_{t=0}^T GL_t^+ * GP_t + GL_t^- * FIT_t$$

Où  $GL_t^+$  est la puissance achetée au réseau de distribution publique (positive ou nulle) et  $GL_t^-$  est la puissance réinjectée dans celui-ci (négative ou nulle).  $GP_t$  est le prix d'achat d'énergie et  $FIT_t$  est le prix de rachat de celle-ci par le réseau de distribution publique. Comme par choix nous ne pouvons pas acheter et réinjecter au même pas de temps, on définit le flux de puissance du réseau de distribution  $GL_t$  par :

$$GL_t^+ = \max(0, GL_t) \text{ et } GL_t^- = -\min(0, GL_t)$$

Soit  $RE_t$  la somme de la génération renouvelable totale et  $PL_t$  la demande (power load) à l'instant  $t$ , on a :

$$GL_t = PL_t - RE_t - \sum_{s \in S} x_t^s$$

### 3.4. Jeux de données

Nous avons identifié des bases de données open-source comprenant des données de consommation ainsi que des données météorologiques. Le tableau 1 en donne le détail. Ces bases de données permettent de générer les profils de demande et de génération renouvelable non-contrôlable qui seront donnés en entrée de l'algorithme d'AR pour calculer les transitions stochastiques. La figure 4 permet de visualiser un exemple de ces profils.

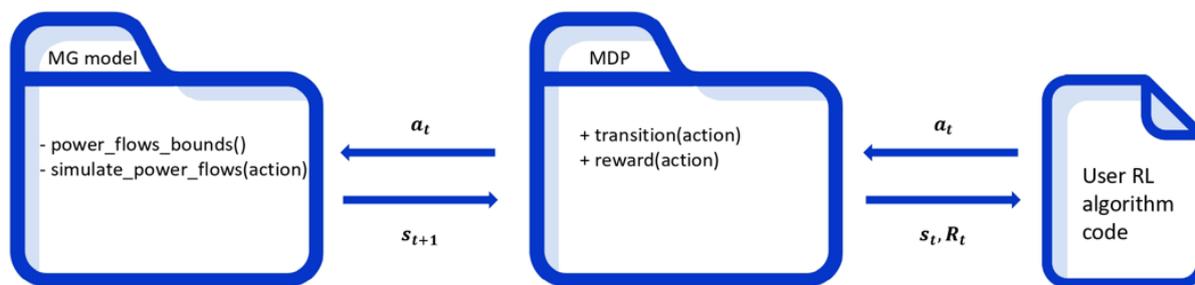


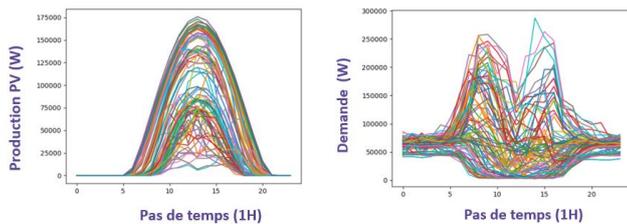
Figure 3 – Schéma de l'outil permettant l'interfaçage entre les algorithmes d'AR et le MDP

Type de données	Base de données	Granularité temporelle	Rayon géographique	Description
Consommation	ODRE [11]	15 min jusqu'à 1 an	Métropoles, régions	La plateforme Open Data Réseaux Energies (ODRE) donne accès à des données de consommation et de production (électrique et gaz) mises à disposition par des distributeurs français (GRTgaz, RTE, ...) ainsi qu'à des données de production renouvelable et de conditions météorologiques, etc. Ces données doivent être rééchelonnées pour s'adapter à un micro-réseau de plus petite taille
Rayonnement solaire (W/m <sup>2</sup> ) et vitesse du vent	ODRE	1 jour	Régions	-
Consommation	Catalogue Dido [12]	1 an	Villes, communes	Cette base de données fournit des données locales de consommation d'électricité, de gaz naturel et de chaleur et de froid. Cette base de données est fournie par le SDES [13]
Rayonnement solaire (W/m <sup>2</sup> )	Solcast [14]	30 min	Toute localisation (coordonnées de latitude et longitude)	Solcast est une API qui fournit des données réelles et prévisionnelles de rayonnement solaire pour toute coordonnée (en latitude et longitude). L'accès à titre gratuit est possible pour les chercheurs via un compte dédié

Tableau 1 – Bases de données open-source pour les données de consommation et les données météorologiques

Caractéristique	Métrique / mesure	Description
Aptitude à réaliser la tâche	Récompense cumulée (RC)	$\frac{1}{n} \sum_t^N r_t(a_t, s_t)$
	Moyenne interquartile de la récompense cumulée	Afin de résumer la performance de l'algorithme pour plusieurs configurations de micro-réseau et pour différents jeux de données, on utilise la moyenne interquartile pour éviter la sensibilité aux valeurs aberrantes tout en reflétant la distribution des performances (en termes de récompense cumulée)
Coût de calcul	Courbe de RC en fonction des itérations d'apprentissage	Il s'agit d'observer la vitesse à laquelle l'algorithme apprend la politique optimale
	Coût de l'ajustement des hyperparamètres et son effet sur la performance	Quelle taille de données est nécessaire pour l'optimisation des hyperparamètres (e.g., optimiser la configuration des réseaux de neurones utilisés dans les algorithmes d'AR profond) et à quelle point la performance est améliorée à l'issue de cette optimisation [15]
Fiabilité	SRT, LRT, DR, RR	Short-term Risk across Time, Long-term Risk across Time, Dispersion across Runs, Risk across Runs [16]. Ces métriques reflètent différents aspects de fiabilité comme la reproductibilité, la stabilité, la dispersion et le risque.

Tableau 2 – Mesures permettant la comparaison des algorithmes d'AR



**Figure 4 – Profils journaliers de génération PV (simulés à l'aide de profils d'irradiation obtenus sur Solcast) et de demande (ODRE) sur l'année 2021-2022**

## 4. CONCLUSION

Pour répondre au problème d'optimisation temps-réel de flux de puissance dans un micro-réseau électrique, on peut, en modélisant ce problème sous forme de MDP, le résoudre en utilisant des algorithmes d'apprentissage par renforcement (AR).

Dans cette étude, un protocole d'évaluation et de comparaison d'algorithmes d'AR appliqués à notre problème d'optimisation est proposé à la communauté scientifique. Ce protocole comprend des jeux de données ainsi que des métriques d'évaluation. Nous avons également comme perspective d'inclure des réseaux normalisés (e.g., réseaux CIGRE [2, 3, 4], IEEE [5, 6], ...) dans les configurations de micro-réseaux comprises dans nos collections expérimentales.

L'implémentation de ce protocole et sa mise à disposition de la communauté scientifique sous forme d'outil informatique nous paraît de la plus grande importance afin de créer une « éprouvette » commune permettant à d'autres chercheurs dans le domaine des micro-réseaux d'évaluer leurs algorithmes et de les comparer à l'état de l'art.

## Références

- [1] Benchmarks in Industrial Engineering Research, URL : <https://github.com/GIS-S-mart/Welcome>
- [2] K. Strunz and S. Barsali, "Cigre task force c6. 04.02: Developing benchmark models for integrating distributed energy resources" 01 2005
- [3] "Working group 14.02 (control in hvdc systems) of study committee 14. The cigre hvdc benchmark model — a new proposal with revised parameters," ELECTRA, pp. 60–66, 12 1994
- [4] K. Strunz and E. K. Brock, "Hybrid plant of renewable stochastic source and multi-level storage for emission free deterministic power generation," CIGRE/IEEE PES International Symposium on Quality and Security of Electric Power Delivery Systems, Montreal, Canada, 10 2003.
- [5] M. S. Ayaz, R. Azizpanah-Abarghoee, and V. Terzija, "European lv microgrid benchmark network: Development and frequency response analysis," in 2018 IEEE International Energy Conference (ENERGYCON), 2018, pp. 1–6.
- [6] IEEE 14-bus system - Illinois center for a smarter electric grid (ICSEG). [Online]. Available: <https://icseg.iti.illinois.edu/ieee-14-bus-system/>
- [7] Mohammed H. Alabdullah et Mohammad A. Abido. « Microgrid energy management using deep Q-network reinforcement learning ». In: Alexandria Engineering Journal 61.11 (1er nov. 2022), p. 9069-9078. issn: 1110-0168. doi : 10.1016/j.aej.2022.02.042. url : <https://www.sciencedirect.com/science/article/pii/S1110016822001284>
- [8] Volodymyr Mnih et al. « Human-level control through deep reinforcement learning ». In : Nature 518.7540 (fév. 2015). Number: 7540. Publisher : Nature Publishing Group, p. 529-533. issn : 1476-4687. doi : 10.1038/nature14236. url : <https://www.nature.com/articles/nature14236>
- [9] Mohammad Amin Soleimanzade, Amit Kumar et Mohtada Sadrzadeh. « Novel data driven energy management of a hybrid photovoltaic-reverse osmosis desalination system using deep reinforcement learning ». In : Applied Energy 317 (2022), p. 119184. issn : 0306-2619. doi : <https://doi.org/10.1016/j.apenergy.2022.119184>. url : <https://www.sciencedirect.com/science/article/pii/S0306261922005566>
- [10] Xiaohan Fang et al. « Multi-agent Deep Reinforcement Learning for Distributed Energy Management and Strategy Optimization of Microgrid Market ». In : Sustainable Cities and Society 74 (2021), p. 103163. issn : 2210-6707. doi : <https://doi.org/10.1016/j.scs.2021.103163>. url : <https://www.sciencedirect.com/science/article/pii/S2210670721004443>
- [11] Open data réseaux Énergies (ODRE). [Online]. Available: <https://opendata.reseaux-energies.fr/>
- [12] Catalogue dido | données et études statistiques. [Online]. Available: [https://www.statistiques.developpement-durable.gouv.fr/catalogue?page=explore&type=dataset&agination=1&sortField=last\\_modified&sortOrder=desc](https://www.statistiques.developpement-durable.gouv.fr/catalogue?page=explore&type=dataset&agination=1&sortField=last_modified&sortOrder=desc)
- [13] Service des données et études statistiques. [Online]. Available: <https://www.statistiques.developpement-durable.gouv.fr>
- [14] Solcast | solar api and solar weather forecasting tool. [Online]. Available: <https://solcast.com/>
- [15] M. Lucic, K. Kurach, M. Michalski, S. Gelly, and O. Bousquet, "Are gans created equal? a large-scale study," Advances in Neural Information Processing Systems 31., p. 698–707, 2018.
- [16] S. C. Y. Chan, S. Fishman, J. Canny, Korattikara Anoop, and G. Sergio, "Measuring the reliability of reinforcement learning algorithms."
- [17] Hou Shengren, Edgar Mauricio Salazar, Pedro P. Vergara, Peter Palensky, "Performance Comparison of Deep RL Algorithms for Energy Systems Optimal Scheduling". Doi: <https://doi.org/10.48550/arXiv.2208.00728>