POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Modélisation, simulation et contrôle par apprentissage par renforcement d'un robot de type dirigeable

ALICE LEMIEUX-BOURQUE
Département de génie informatique et génie logiciel

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
Génie informatique

Mai 2025

© Alice Lemieux-Bourque, 2025.

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Ce mémoire intitulé :

Modélisation, simulation et contrôle par apprentissage par renforcement d'un robot de type dirigeable

présenté par **Alice LEMIEUX-BOURQUE**
en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
a été dûment accepté par le jury d'examen constitué de :

Benjamin DE LEENER, président
Giovanni BELTRAME, membre et directeur de recherche
David SAUSSIÉ, membre

REMERCIEMENTS

Je suis reconnaissante d'avoir pu apprendre et évoluer au sein d'un environnement comme le MISTLab. Je tiens à remercier Giovanni Beltrame pour l'accueil et la confiance qu'il m'a accordée.

I would like to thank all the members of MISTLab for their time and guidance throughout my Master's program. I am especially grateful to Hassan Fouad, who served as a mentor and was always available to support me.

I would also like to thank Professor Ely Carneiro De Paiva for the many helpful insights and the time he took to share his expertise with me. Obrigada !

Je veux souligner l'appui financier des Fonds de recherche du Québec - Nature et technologie (#331556) et du Conseil de recherche en sciences naturelles et en génie du Canada. I would like to highlight the financial support of the Fonds de recherche du Québec - Nature et technologie (#331556) and the National Sciences and Engineering Research Council of Canada.

Je tiens aussi remercier Polytechnique Montréal et toutes les personnes que j'y ai cotoyées depuis mon arrivée en 2017.

Finalement, merci à mes amies et amis, ma famille et mon partenaire Charles-Étienne pour leur amour et leur soutien inconditionnels.

RÉSUMÉ

Le secteur des transports est un des plus polluants dans le monde. Que ce soit pour le transport de personnes par avion d'un côté à l'autre de l'océan ou pour le transport de marchandises par camion, l'industrie produit près de 15% des émissions de gaz à effet de serre au niveau mondial. Au-delà de l'électrification des transports, certains rêvent d'un grand retour des dirigeables. Ces aéronefs particuliers se présentent en effet comme un candidat idéal pour pallier aux problèmes de l'industrie du transport : ils ont une très faible consommation d'énergie puisque la majorité de la force de portance est générée par le gaz de levage leur permettant de flotter, ils sont peu bruyants et ils peuvent rester en vol stationnaire, ce qui leur permet d'atteindre des zones où les avions ou les camions ne se rendent pas. Toutefois, l'utilisation des dirigeables comporte plusieurs défis. Parmi ceux-ci se trouve la difficulté de créer des contrôleurs de vols automatiques sécuritaires. En effet, en raison de leur grande surface, les dirigeables sont très sensibles aux perturbations externes comme les rafales de vent. Les forces aérodynamiques sur le dirigeable étant difficiles à modéliser, les méthodes classiques peinent à contrôler adéquatement les dirigeables. Une nouvelle approche consiste à utiliser l'apprentissage automatique, et plus spécifiquement l'apprentissage par renforcement, pour créer des politiques de contrôles applicables aux dirigeables. Toutefois, il est difficile de créer des modèles adéquats puisqu'il est presque impossible de les tester sur des systèmes de dirigeables réels. Dans ce mémoire, une plateforme de test complète permettant l'entraînement de politiques d'apprentissage par renforcement sur un petit dirigeable est développée. Elle comprend notamment la conception et la création d'un robot dirigeable de deux mètres de long pour utilisation à l'intérieur. Le robot a quatre moteurs et 2 gouvernes à l'arrière. Pour accompagner le robot, un simulateur a été créé avec MuJoCo, un moteur physique populaire en robotique. MuJoCo ne possédant pas de modélisation des interactions avec les fluides adéquate pour les dirigeables, un modèle d'aérodynamique a été développé afin d'être intégré à l'engin physique.

Finalement, des algorithmes d'apprentissage par renforcement sont testés sur le simulateur. Deux modèles sont créés : un permettant de contrôler l'altitude du dirigeable et le garder en vol stationnaire, et un deuxième permettant de garder une certaine vitesse à une altitude constante. Les résultats obtenus valident la capacité du système à apprendre à être contrôlé grâce à l'apprentissage par renforcement.

ABSTRACT

The transportation sector is one of the most polluting industries in the world. Whether it is the transportation of people by plane across the ocean or the transportation of goods by truck, the industry accounts for nearly 15% of global greenhouse gas emissions. Beyond the electrification of transportation, some dream of a great return of airships. These particular aircrafts are indeed seen as an ideal candidate to address the problems of the transportation industry: they have a very low energy consumption since the majority of their lift is generated by the lifting gas that allows them to float, they are relatively quiet, and they can hover, which allows them to reach areas where airplanes or trucks cannot go.

However, the use of airships comes with several challenges. Among these is the difficulty of creating safe automatic flight controllers. Indeed, due to their large surface area, airships are very sensitive to external disturbances such as wind gusts. The aerodynamic forces acting on airships are difficult to model, and traditional control methods struggle to adequately control them. A new approach involves using machine learning, and specifically reinforcement learning, to create control policies for airships. However, it is challenging to develop adequate models since it is nearly impossible to test them on real airship systems.

In this thesis, a complete testing platform is developed for training reinforcement learning policies on a small airship. It includes the design and creation of a two-meter-long indoor airship robot. The robot has four motors and two rudders at the rear. To accompany the robot, a simulator was created using MuJoCo, a popular physics engine in robotics. Since MuJoCo does not provide an adequate fluid interaction model for airships, an aerodynamic model was developed to be integrated with the physical engine.

Finally, reinforcement learning algorithms are tested on the simulator. Two models are created: one for controlling the altitude of the airship and keeping it stationary, and a second one for maintaining a certain speed at a constant altitude. The results validate the system's ability to learn how to be controlled through reinforcement learning.

TABLE DES MATIÈRES

REMERCIEMENTS iii

RÉSUMÉ iv

ABSTRACT v

TABLE DES MATIÈRES vi

LISTE DES TABLEAUX viii

LISTE DES FIGURES ix

LISTE DES SIGLES ET ABRÉVIATIONS x

LISTE DES ANNEXES xi

CHAPITRE 1 INTRODUCTION 1
- 1.1 Définitions et concepts de base 1
  - 1.1.1 Dirigeables 1
  - 1.1.2 Opération d'un dirigeable 2
  - 1.1.3 Masse ajoutée 2
  - 1.1.4 Apprentissage par renforcement 3
- 1.2 Éléments de la problématique 4
  - 1.2.1 Conception d'un dirigeable pour utilisations dans des espaces fermés . 5
  - 1.2.2 Modélisation aérodynamique 5
  - 1.2.3 Limitation des simulateurs 5
  - 1.2.4 Déploiement d'algorithmes d'apprentissage par renforcement 6
  - 1.2.5 Apprentissage par renforcement sur des systèmes réels 6
- 1.3 Objectifs de recherche 6
- 1.4 Plan du mémoire 7

CHAPITRE 2 REVUE DE LITTÉRATURE 8
- 2.1 Dirigeables de petite et moyenne taille 8
- 2.2 Modélisation dynamiques des dirigeables 9
- 2.3 Simulateurs pour apprentissage par renforcement 10
- 2.4 Apprentissage par renforcement appliqué aux dirigeables 11

CHAPITRE 3 DÉTAILS DE LA SOLUTION 13
3.1 Le prototype physique 13
3.1.1 L'enveloppe 13
3.1.2 Le système de commande 14
3.1.3 La gondole 15
3.1.4 Le centre de masse 15
3.1.5 Le système de communication 16
3.1.6 La conception électronique 16
3.1.7 Les capteurs 16
3.1.8 Le prototype final 18
3.2 Le modèle mathématique 19
3.2.1 Cadres référentiels 19
3.2.2 Matrice de masses 20
3.2.3 Forces dynamiques 21
3.2.4 Forces de propulsion 21
3.2.5 Forces aérodynamiques 22
3.2.6 Forces de flottabilité et gravité 26
3.3 Le simulateur 26

CHAPITRE 4 RÉSULTATS THÉORIQUES ET EXPÉRIMENTAUX 28
4.1 Les environnements 28
4.1.1 Vol stationnaire à altitude spécifiée 29
4.1.2 Vol de croisière à vitesse spécifiée 31
4.2 Résultats 32
4.2.1 Vol stationnaire à altitude spécifiée 33
4.2.2 Vol de croisière à vitesse spécifiée 38
4.3 Conclusion 43

CHAPITRE 5 CONCLUSION 44
5.1 Synthèse des travaux 44
5.2 Limitations de la solution proposée 44
5.3 Améliorations futures 45

RÉFÉRENCES 46

ANNEXES 50

LISTE DES TABLEAUX

Tableau 3.1 Coefficients utilisés pour calculer les forces aérodynamiques 24
Tableau 4.1 Table de correspondance des actions pour la tâche de vol stationnaire 29
Tableau 4.2 Table de correspondance des actions pour la tâche de vol de croisière 31
Tableau 4.3 Paramètres de l'environnement pour la tâche de vol stationnaire . . . 33
Tableau 4.4 Paramètres de l'environnement pour la tâche de vol de croisière . . . 38
Tableau A.1 Tableau des définitions des variables utilisées dans le modèle dynamique du dirigeable. 51

LISTE DES FIGURES

Figure 1.1 Apprentissage par renforcement 4
Figure 3.1 Électronique 17
Figure 3.2 Design 3D de RAFALE 18
Figure 3.3 Dirigeable physique 18
Figure 3.4 Axes de référence du dirigeable 20
Figure 3.5 Intégration de l'aérodynamique avec MuJoCo 27
Figure 4.1 Courbes d'entraînement pour la tâche de vol stationnaire à altitude donnée 34
Figure 4.2 Évaluation des modèles à différents stades d'entraînement (cibles entre 2 et 4 m) 35
Figure 4.3 Évaluation des modèles entraînés avec PPO à différents stades d'entraînement (cibles entre 2 et 4 m) 36
Figure 4.4 Observations selon le pas de temps pour différentes altitudes 37
Figure 4.5 Résultats de l'entraînement pour la tâche de vol en mode croisière . . 39
Figure 4.6 Visualisation des observations générées par le modèle entraîné : 4000 pas de temps et vitesse cible de 0.1 m/s 40
Figure 4.7 Visualisation des observations générées par le modèle entraîné : 4000 pas de temps et vitesse cible de 0.5 m/s 41
Figure 4.8 Visualisation des observations générées par le modèle entraîné : 4000 pas de temps et vitesse cible de 1 m/s 42
Figure 4.9 Mouvements de la gouverne de profondeur du dirigeable selon l'angle de tangage 43

LISTE DES SIGLES ET ABRÉVIATIONS

A2C	Advantage Actor Critic
BEC	Régulateur de tension électrique
CFD	Mécanique de fluide numérique
DQN	Deep Q Network
DRL	Apprentissage par renforcement profond
ESC	Variateur électronique de vitesse
IMU	Centrales inertielles
PPO	Proximal policy optimization
RL	Apprentissage par renforcement
RTF	Facteur temps réel

LISTE DES ANNEXES

Annexe A Définitions des variables du modèle dynamique du dirigeable 50

CHAPITRE 1 INTRODUCTION

En 2023, le secteur des transports se classait au deuxième rang des émetteurs de gaz à effet de serre, représentant 15% des émissions au niveau mondial : 11% pour le transport routier et 2% pour l'aviation [1]. Différentes initiatives visant à réduire l'empreinte carbone de ce secteur ont émergé, notamment le développement de carburants à zéro émission et l'électrification des transports. Une idée soutenue par certaines entreprises est de repenser les modes de transport eux-mêmes et d'envisager le développement et l'utilisation de dirigeables comme une alternative pour le transport de marchandises. Leur faible consommation énergétique pourrait faire des dirigeables une option pour concurrencer le transport par avion, mais aussi le transport routier sur de longues distances [2].

À plus petite échelle, les robots dirigeables pourraient aussi être utilisés dans les tâches de longue durée, leur autonomie prolongée et leur capacité à rester en vol stationnaire les rendant efficaces dans des applications telles que la collecte de données [3] [4].

La construction de dirigeables soulève toutefois plusieurs défis. Parmi ceux-ci, on peut citer leur faible acceptabilité sociale, en partie attribuable à l'accident du Hindenburg en 1937, ainsi que la complexité de leur contrôle, notamment en raison de leur grande sensibilité aux conditions météorologiques et perturbations externes.

Ce travail a été réalisé dans le cadre d'une collaboration entre le MIST Lab de Polytechnique Montréal, Thales et Flying Whales Québec. L'objectif global de cette collaboration est la création de politiques d'apprentissage par renforcement certifiables pour des aéronefs comme les dirigeables. Ce mémoire décrit la conception d'une plateforme de test et d'un simulateur permettant d'appliquer des politiques d'apprentissage par renforcement sur un robot de type dirigeable. Les travaux ont été menés entre août 2022 et avril 2025.

1.1 Définitions et concepts de base

1.1.1 Dirigeables

Un dirigeable est un aéronef motorisé utilisant un gaz de levage, généralement de l'hélium ou de l'hydrogène, pour générer une flottabilité permettant de réduire son besoin de propulsion verticale. Il se distingue entre autres par sa faible consommation d'énergie pour rester en vol, lui permettant d'être en l'air pour de longues périodes. Le dirigeable possède à la fois certains avantages des avions, notamment la capacité à parcourir de longues distances et à transporter de lourdes charges, et certains avantages des hélicoptères, notamment la possibilité de décoller

et atterrir verticalement, ainsi que rester en vol stationnaire. Il a traditionnellement une forme allongée qui permet de réduire la force de traînée lors du vol.

Les dirigeables sont généralement classés en trois grandes catégories : les dirigeables rigides, qui possèdent une structure légère limitant les déformations, les dirigeables souples, qui ne possèdent pas de structure interne et dont la forme est déterminée par la pression du gaz de levage à l'intérieur, et les dirigeables semi-rigides, qui se situent quelque part entre les deux autres types.

Les dirigeables peuvent également être différenciés en fonction de leur taille. Les plus grands dirigeables, comme le LCA60T de Flying Whales, peuvent transporter des charges lourdes sur de longues distances. Les plus petits peuvent quant à eux être utilisés pour la collecte de données, la publicité, voire des applications à l'intérieur de bâtiments ou d'espaces restreints.

Les principales composantes d'un dirigeable sont : l'enveloppe, qui contient le gaz de levage et contribue largement à la traînée aérodynamique ; la gondole, qui peut contenir les systèmes de commandes, les passagers et la charge utile ; le système de propulsion et le système d'empennage.

1.1.2 Opération d'un dirigeable

Le contrôle des dirigeables repose sur différents systèmes adaptés aux différentes phases de vol. À basse vitesse ou en vol stationnaire, le système de propulsion regroupant les différents moteurs permet le contrôle de la direction, la vitesse et l'altitude du véhicule. C'est le même système qui est utilisé pour les manœuvres de décollage et d'atterrissage. En mode croisière, donc à des vitesses plus élevées, le contrôle combine le système de propulsion et le système de gouvernes, qui permet une meilleure stabilité. Pour les plus grands dirigeables, des systèmes de ballonnets ou de relâchement de gaz peuvent être utilisés pour contrôler l'altitude [5].

1.1.3 Masse ajoutée

La masse ajoutée, ou masse virtuelle, représente la masse supplémentaire qu'un objet doit déplacer lorsqu'il se déplace dans un fluide. Cette masse ajoutée est due à l'inertie du fluide environnant qui résiste au mouvement de l'objet : c'est l'équivalent de la masse de l'air qui est déplacée avec l'objet.

La masse ajoutée intervient dans toutes les situations où un objet accélère ou décélère dans un fluide, mais elle est particulièrement importante dans le cas des dirigeables : en raison de leur faible densité, la masse ajoutée est proportionnellement plus importante. [6].

Afin d'estimer les forces et moments supplémentaires générés par le fluide en mouvement autour du dirigeable, la masse ajoutée est généralement modélisée à l'aide de coefficients d'inertie, comme ceux proposés par Lamb pour des formes ellipsoïdales [7].

Les masses ajoutées et les inerties associées peuvent être estimées comme suit :

$$\begin{aligned}
X_{\dot{u}} &= -k_1 \frac{B}{g} \\
Y_{\dot{v}} &= -k_2 \frac{B}{g} \\
Z_{\dot{w}} &= Y_{\dot{v}} \\
M_{\dot{q}} &= -k' \frac{B}{g} \left[\frac{l^2 + d^2}{20} \right] \\
N_{\dot{r}} &= M_{\dot{q}} \\
L_{\dot{p}} &= 0
\end{aligned} \tag{1.1}$$

où B est la force de flottabilité, g est l'accélération gravitationnelle, k_1, k_2 et k' sont les ratios d'inertie de Lamb, l est la longueur du dirigeable et d est son diamètre maximal. Les équations $Z_{\dot{w}} = Y_{\dot{v}}$, $N_{\dot{r}} = M_{\dot{q}}$ et $L_{\dot{p}} = 0$ sont de bonnes approximations si on ignore les effets de la gondole et de l'empennage, ce qui est une pratique courante, car l'effet de la masse ajoutée est principalement attribuable à l'enveloppe du dirigeable.

1.1.4 Apprentissage par renforcement

L'apprentissage par renforcement, ou RL (*Reinforcement Learning*), est un sous-domaine de l'apprentissage automatique. C'est une méthode fonctionnant par essai et erreur où un agent apprend à partir de ses interactions avec l'environnement. L'agent reçoit une observation de l'environnement, choisit une action à effectuer puis reçoit une récompense ou une punition en fonction de l'action sélectionnée. Le signal de récompense reçu est utilisé pour modéliser la tâche à effectuer. L'agent apprend donc à effectuer la tâche en maximisant le signal de récompense. C'est la méthode d'apprentissage automatique privilégiée pour les problèmes de contrôle car elle permet de développer des politiques de contrôle flexibles qui peuvent s'adapter à des environnements complexes et incertains. La figure 1.1 le fonctionnement général de l'apprentissage par renforcement.

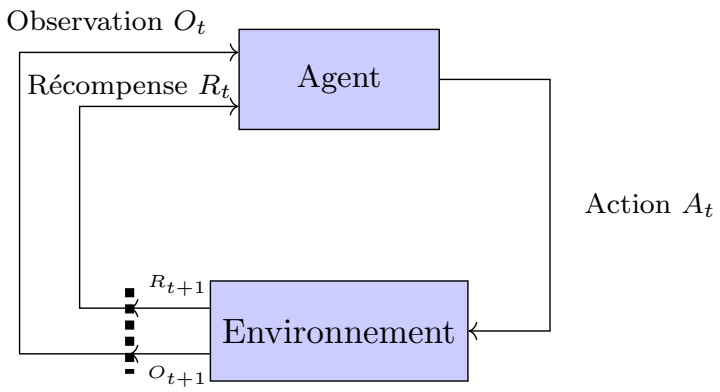

FIGURE 1.1 Apprentissage par renforcement

1.2 Éléments de la problématique

Le contrôle de dirigeable représente un défi en raison des dynamiques complexes et de la sensibilité des dirigeables aux perturbations externes, une conséquence directe de sa surface importante. Parallèlement au regain d'intérêt pour les dirigeables, les entreprises du secteur cherchent à implémenter des systèmes de contrôle basés sur l'apprentissage automatique, plus spécifiquement l'apprentissage par renforcement.

Toutefois, l'apprentissage par renforcement fait face à de nombreux défis lors de son application sur les systèmes réels. En effet, les politiques de contrôles apprises par renforcement n'ont généralement pas de garanties de sécurité. Pour les compagnies développant des dirigeables, l'application sur les systèmes réels est souvent impossible pour deux principales raisons. Premièrement, les systèmes sont généralement encore en phase de développement. Deuxièmement, le déploiement de politiques de contrôles apprises par renforcement n'étant pas sécuritaire, il présente des risques et pourrait engendrer des incidents extrêmement coûteux.

Cette recherche a donc pour objectif de développer une plateforme de test complète comprenant un robot dirigeable à faible coût, un simulateur adapté et une infrastructure permettant la création, l'entraînement et le déploiement de politiques d'apprentissage par renforcement sur le système. Afin de faciliter son utilisation, le dirigeable conçu doit être adapté aux espaces de laboratoires disponibles pour l'équipe.

1.2.1 Conception d'un dirigeable pour utilisations dans des espaces fermés

Bien que l'utilisation d'un dirigeable dans des espaces fermés permette d'éliminer certaines variables comme les changements de volume ou de température, cette approche introduit des contraintes spécifiques.

Premièrement, la taille du robot est un facteur important à considérer. Le dirigeable doit être suffisamment petit pour pouvoir manœuvrer dans un espace restreint, tout en conservant une flottabilité suffisante pour pouvoir maintenir son avantage énergétique par rapport aux drones conventionnels.

Ensuite, un avantage des dirigeables est qu'ils peuvent soulever d'importantes charges utiles. Toutefois, un petit dirigeable ne permet pas une grosse charge utile, et son fonctionnement est donc particulièrement sensible à la masse de ses composantes. Le dirigeable doit donc fonctionner avec un nombre limité d'actionneurs et de capteurs, tout en permettant un contrôle adéquat.

Finalement, comme il fonctionne dans des espaces restreints, le dirigeable doit pouvoir à la fois voler en mode stationnaire et en vitesse de croisière, malgré les limitations du nombre d'actionneurs imposés par les contraintes de masse.

1.2.2 Modélisation aérodynamique

La modélisation d'un dirigeable doit prendre en compte les effets aérodynamiques qui influencent considérablement son mouvement. Un modèle fidèle à la réalité est essentiel pour développer des politiques d'apprentissage par renforcement transférables à la réalité.

Déterminer les coefficients aérodynamiques est un domaine de recherche en soi. Les méthodes traditionnelles permettant d'obtenir ces coefficients, soit les essais en soufflerie ou la simulation par mécanique des fluides numérique (CFD), sont généralement chronophages, coûteuses en ressources ou demandent une expertise spécifique. De plus, les modèles existants se limitent généralement à de petits angles d'attaques et ne modélisent pas adéquatement les mouvements d'un dirigeable en vol stationnaire.

1.2.3 Limitation des simulateurs

Il existe très peu de simulateurs en logiciel libre permettant de simuler le mouvement de dirigeables, et ceux qui existent se limitent souvent à un mouvement en deux dimensions. Cette réalité contraste avec la disponibilité de nombreux modèles pour les drones plus traditionnels. Ainsi, les équipes de recherche souhaitant appliquer leurs algorithmes d'apprentissage

par renforcement aux dirigeables sont donc souvent obligées de développer leurs propres modèles et outils.

Les plateformes de simulation physique comme MuJoCo offrent un cadre intéressant pour le développement d'algorithmes d'apprentissage par renforcement grâce à leur compatibilité avec de nombreux outils et environnements. Cependant, elles n'incluent pas de modèles de dynamique des fluides appropriés pour simuler fidèlement le comportement des dirigeables. Il n'existe donc pas de solution facile et accessible pour l'entraînement de politiques de contrôle de dirigeables.

1.2.4 Déploiement d'algorithmes d'apprentissage par renforcement

Pour maximiser l'utilité de la plateforme, celle-ci doit faciliter le développement, l'entraînement et la validation de politiques d'apprentissage par renforcement. L'infrastructure doit donc permettre l'expérimentation et la comparaison de différentes méthodes, en plus d'assurer une transition entre la simulation et la réalité. Pour ce faire, différents outils et bibliothèques existent déjà, et il sera donc essentiel que la plateforme puisse tirer profit des outils déjà existants.

1.2.5 Apprentissage par renforcement sur des systèmes réels

Tous les simulateurs présentent ce que l'on appelle un *reality gap*, un écart entre la simulation et la réalité. Cet écart constitue un obstacle majeur à l'application de l'apprentissage par renforcement entraîné en simulation sur des systèmes réels.

En raison des contraintes, notamment en matière de sécurité, très peu d'études sur l'apprentissage par renforcement appliqué aux dirigeables testent leurs méthodes sur des systèmes réels : la majorité des travaux restent confinés à la simulation.

1.3 Objectifs de recherche

Le travail présenté dans ce mémoire a les objectifs suivants :

1. Concevoir et construire un robot de type dirigeable pouvant être utilisé à l'intérieur d'un laboratoire afin de pouvoir tester des algorithmes d'apprentissage par renforcement.
2. Développer un modèle mathématique pour ce robot de type dirigeable.
3. Créer un simulateur à partir du modèle mathématique permettant de tester des algorithmes d'apprentissage par renforcement sur ce robot.

4. Développer des politiques d'apprentissage par renforcement en simulation pour ce robot.

1.4 Plan du mémoire

La suite de ce mémoire est divisée en quatre chapitres. Le chapitre 2 présente une revue de la littérature sur les dirigeables, l'apprentissage par renforcement et les simulateurs. Le chapitre 3 présente le design mécanique du robot de type dirigeable, le modèle mathématique du robot et le simulateur développé. Le chapitre 4 présente les politiques d'apprentissage par renforcement développées et les résultats obtenus. Enfin, le chapitre 5 conclut le mémoire, souligne les limitations du travail de recherche et présente des pistes pour des travaux futurs.

CHAPITRE 2 REVUE DE LITTÉRATURE

Ce chapitre présente un aperçu de la recherche dans les différents thèmes essentiels à ce mémoire. Il commence par un aperçu des dirigeables exitants, présente ensuite les enjeux de la modélisation dynamique des dirigeables, un aperçu de la simulation pour apprentissage par renforcement et présente finalement les travaux sur l'apprentissage par renforcement appliqué aux dirigeables.

2.1 Dirigeables de petite et moyenne taille

Dans cette revue de littérature, l'accent sera mis sur les dirigeables sans équipage (*unmanned*) de petite et moyenne taille, typiquement de l'ordre de quelques mètres jusqu'à une vingtaine de mètres. En effet, les dirigeables de grande taille utilisés pour le transport de marchandises ou de personnes, comme le LCA60T de Flying Whales, ont des systèmes complexes et inadaptés à la conception de dirigeables intérieurs. Le projet présenté exige un nombre limité d'actionneurs afin de respecter les contraintes de poids.

La mission AURORA (*Autonomous Unmanned Remote Monitoring Robotic Airship*) est une référence dans la recherche moderne sur les dirigeables autonomes. Le dirigeable utilisé pour les différentes étapes de cette mission a évolué avec le temps, mais la base des systèmes d'actionnement utilisés est restée la même. Le dirigeable initialement utilisé, le AS800, mesure 9 mètres de long par 2,25 mètres de diamètre. Il est équipé de deux propulseurs orientables de chaque côté de la gondole, avec un angle d'inclinaison allant initialement de -30° à +90° [8], puis étendu de -30° à +120° [9]. Les deux principaux propulseurs offrent une poussée vectorielle permettant à la fois le déplacement en mode croisière et la compensation du poids à basse vitesse. Il dispose de quatre gouvernes arrières disposées en forme de « X », chacune ayant une déflexion de -25° à +25° permettant des commandes combinées de profondeur, de direction et de roulis. Cette forme permet à la fois des commandes équivalentes aux gouvernes de profondeur et de direction de la forme classique en croix, et des commandes équivalentes à un aileron avec la rotation des quatre gouvernes dans la même direction [9]. Un troisième propulseur, facultatif, se situe à la proue du dirigeable pour assurer un contrôle de précision à basse vitesse. Un second dirigeable a ensuite été créé par l'équipe de AURORA, très semblable au AS800, mais mesurant 10,5 mètres de long et ayant un diamètre de 3 mètres. Sa principale innovation résidait dans l'ajout d'une possible commande différentielle des moteurs principaux [3, 10].

Plusieurs dirigeables adoptent un système d'actionnement similaire à ceux du projet AURORA, avec quelques variations. Par exemple, les dirigeables de [11, 12], conçus pour une utilisation en intérieur, sont plus petits (∼2–3 m) et disposent de deux propulseurs principaux inclinables à 180°. Ils possèdent un empennage, mais pas de gouvernes, et intègrent un moteur supplémentaire à l'arrière. Les dirigeable de [13] et du projet DIVA [14] ont une taille et un système de propulsion semblables à ceux d'AURORA, mais leurs deux propulseurs principaux sont différentiables. Les quatre gouvernes arrières de [13] sont disposées en croix, alors que celles de [14] sont en forme de « X ».

Le projet InSAC pour l'acquisition de données environnementale a utilisé deux dirigeables souples. Le NOAMAY [15], possède des gouvernes en « X » et quatre propulseurs orientables montés sur une structure attachée sous l'enveloppe. Deux propulseurs sont situés à l'avant du centre de gravité du dirigeable et deux à l'arrière. L'enveloppe est identique à celle utilisée pour le projet AURORA. Plus récemment, le projet a utilisé le Naomini [4]. Il est équipé de six propulseurs orientables fixés directement sur l'enveloppe. Chaque propulseur peut être incliné sur 360°. Il possède également trois gouvernes arrières. Sa conception permet d'utiliser deux, quatre ou six propulseurs selon les besoins.

Disposer les moteurs en les fixant de chaque côté de l'enveloppe est aussi une solution utilisée pour le dirigeable ALTAV Quanser MkII [16]. Ce modèle de 4,5 mètres dépourvu d'empennage possède quatre propulseurs orientables fixés sur l'enveloppe, deux de chaque côté. Ces propulseurs peuvent produire une poussée vers le haut, vers l'avant et vers l'arrière. Le dirigeable créé par [17] quant à lui possède quatre moteurs à orientation fixe placés de chaque côté de son enveloppe. Conçu pour une utilisation en intérieur, il mesure 2 mètres de long et ne possède pas d'empennage.

Les dirigeables classiques de petite et moyenne taille sont donc majoritairement souples et ont généralement un nombre limité de moteurs, souvent entre deux et quatre. Ils sont le plus souvent sous-actionnés et ne permettent presque jamais de contrôle latéral.

D'autres dirigeables existent avec des formes non conventionnelles [18, 19], mais ne sont pas détaillés ici, car leur aérodynamique est différente des dirigeables traditionnels.

2.2 Modélisation dynamiques des dirigeables

La modélisation dynamique des dirigeables est nettement plus complexe que celle des avions. En effet, la modélisation des dirigeables doit intégrer la flottabilité, les effets d'inertie de l'air environnant (masse ajoutée), ainsi que les forces aérodynamiques dont le calcul est encore un important sujet de recherche. En plus de dynamiques plus complexes, les dirigeables

ne bénéficient pas de modèles aussi bien testés et largement validés par des décennies de données expérimentales, entre autres en raison du déclin de leur utilisation après les années 1930 [20]. Le principal défi à la modélisation des dirigeables réside dans la détermination des coefficients aérodynamiques. Différentes approches sont proposées dans la littérature. La plus populaire, principalement car la seule accessible à l'âge d'or des dirigeables, est la détermination des coefficients en soufflerie [5]. Les résultats en soufflerie peuvent ensuite être extrapolés pour créer des modèles plus complets de dirigeables de dimensions similaires [10,21]. Un autre approche est d'utiliser des formules semi-empiriques ou analytiques, dérivées de la mécanique des fluides appliquée à des corps allongés, comme celles proposées par [22,23]. Plus récemment, des travaux explorent l'utilisation de la mécanique des fluides numérique (ou Computational Fluid Dynamics, CFD) pour la détermination des effets aérodynamiques sur les dirigeables [24]. La CFD a l'avantage de pouvoir aussi être utilisée pour optimiser la forme des dirigeables et analyser leur stabilité [25, 26].

Un point commun entre les différentes méthodes de détermination des effets aérodynamiques est le manque de modèle à angle d'attaque élevé [20]. Cette modélisation est pourtant essentielle pour les manoeuvres à basse vitesse et le vol stationnaire.

2.3 Simulateurs pour apprentissage par renforcement

À partir des modèles de dirigeables développés, certains travaux ont créé des simulateurs pour tester les algorithmes de contrôle. Par exemple, les projets AURORA, DIVA, DRONI et INSAC [10, 14, 15, 27] ont développé et amélioré un simulateur de dirigeable en Simulink/MATLAB. Ce simulateur n'est toutefois pas disponible en logiciel libre. Price et al. [13] ont quant à eux créé un simulateur dans Gazebo. Ce simulateur en logiciel libre a été réutilisé dans différents travaux testant des algorithmes d'apprentissage par renforcement [28,29]. C'est un des seuls simulateurs de dirigeable en logiciel libre. Il existe donc très peu de simulateurs accessibles pour les dirigeables.

L'utilisation d'un moteur physique permet de faciliter la simulation robotique. En effet, un moteur physique permet de simuler la dynamique des objets en prenant en compte les forces et les moments qui agissent sur eux. Un des simulateurs physiques les plus populaires pour tester des algorithmes d'apprentissage par renforcement en robotique est MuJoCo (*Multi-Joint dynamics with Contact*) [30]. En plus d'obtenir des performances similaires aux autres moteurs physiques disponibles [31], MuJoCo facilite grandement l'étude du contrôle par apprentissage par renforcement et son implémentation sur les modèles compatibles, car il permet l'utilisation facile des bibliothèques et *benchmarks* populaires comme Stable Baselines3 [32] et Gymnasium [33]. Développer un modèle adéquat de dirigeable sur MuJoCo faciliterait

donc grandement la recherche en apprentissage par renforcement sur les dirigeables.

Toutefois, MuJoCo est un simulateur fait pour les robots à corps rigide et, comme la majorité des simulateurs physiques [34], n'est pas adapté pour la simulation des interactions avec les fluides nécessaire pour les dirigeables. Il est donc difficile d'utiliser MuJoCo pour simuler des dirigeables, car il ne prend pas en compte la dynamique complexe des fluides. Cependant, à l'aide de la bibliothèque `dm_control` [35], il est possible d'ajouter des forces et des moments sur les objets, ce qui permet de simuler l'aérodynamique d'un dirigeable.

2.4 Apprentissage par renforcement appliqué aux dirigeables

L'apprentissage par renforcement a démontré d'excellentes performances pour la résolution de problèmes complexes, notamment le contrôle de robots ayant des dynamiques non linéaires. Bien que la majorité des avancées dans le domaine aient été obtenues en simulation [36], des résultats ont aussi été obtenus sur des systèmes réels, notamment pour le contrôle de drones [37, 38] et de robots quadrupèdes [39].

Le développement moderne des dirigeables est entre autres difficile en raison du défi que représente le contrôle de ces véhicules qui sont généralement sous-actionnés, en plus d'avoir des dynamiques non linéaires, couplées et sous-amorties [4]. Les méthodes de contrôle non linéaire classiques manquent souvent de robustesse, car elles reposent fortement sur la précision du modèle dynamique du véhicule utilisé, modèle qui est lui-même complexe à établir. Ainsi, l'utilisation de l'apprentissage par renforcement présente une option pour répondre aux défis posés par le contrôle des dirigeables.

Les premiers pas de la recherche sur l'apprentissage par renforcement appliquée aux dirigeables se concentraient sur des modèles basés sur les processus gaussiens et étaient appliqués à des tâches de petite dimension [40], [41]. Les dernières avancées utilisent plutôt des algorithmes d'apprentissage par renforcement profond (*Deep Reinforcement Learning*, DRL), qui permettent de s'attaquer à des problèmes de haute dimension comme le contrôle d'un dirigeable dans un environnement 3D. Par exemple, pour suivre une trajectoire dans un environnement 3D, [42] propose un contrôleur basé sur PPO qui intègre les saturations des actionneurs, tandis que [43] utilise un algorithme de Q-learning associé à un réseau de neurones pour accélérer l'entraînement du modèle.

Des approches hybrides [28, 29] utilisent un contrôleur plus traditionnel pour la stabilité (un PID et un contrôleur H_∞, respectivement) qui est combiné avec un agent utilisant un algorithme de DRL (*Proximal Policy Optimization*, PPO) qui permet d'optimiser la performance. Les deux méthodes ont démontré de meilleurs résultats pour la poursuite d'une trajectoire

que la simple utilisation du contrôleur traditionnel.

La combinaison de l'apprentissage par renforcement et de l'apprentissage par démonstration a également été explorée dans [12].

CHAPITRE 3 DÉTAILS DE LA SOLUTION

Cette section décrit le développement de RAFALE (*Robotic Airship for Flight And Learning Experiments*), un robot de type dirigeable pour utilisation à l'intérieur et le simulateur l'accompagnant.

3.1 Le prototype physique

Le prototype robotique RAFALE a été conçu et construit afin de servir de plateforme de test pour l'apprentissage par renforcement. La principale considération dans la conception de RAFALE était la nécessité que celui-ci puisse être utilisé dans des espaces restreints afin de permettre des tests à l'intérieur. Ainsi, il doit être petit et doit pouvoir rester en vol stationnaire. Un système d'empennage et de gouvernes est présent afin d'assurer une certaine stabilité et de permettre le développement d'algorithmes de contrôle plus avancés. Ces gouvernes sont essentielles pour obtenir des résultats généralisables à de plus gros dirigeables.

Le plus grand défi dans la conception de RAFALE est de réduire le poids du système tout en maintenant une charge utile suffisante pour supporter les différents composants électroniques et capteurs. La conception du système a également pris en compte la facilité d'assemblage et de démontage, afin de permettre un accès facile aux composants internes et de réduire le stress sur l'enveloppe. Finalement, l'objectif de RAFALE est aussi d'offrir une plateforme de test à coût réduit pour le laboratoire. Ainsi, la conception du système a été faite en tenant compte de la disponibilité des composants et de leur coût.

La conception électronique du système s'inspire fortement du travail de [44], et a bénéficié des conseils du Laboratoire INIT de École de technologie supérieure.

3.1.1 L'enveloppe

La sélection de l'enveloppe est à la base de toutes les autres décisions de conception du dirigeable. Cinq critères doivent être pris en compte lors de la sélection de l'enveloppe : le matériel, la forme, la taille, le volume et le ratio longueur/diamètre. Le matériel permet de déterminer la structure du dirigeable, à savoir s'il est rigide ou souple. Pour des petits robots dirigeables, la structure est généralement souple, comme c'est le cas de tous les modèles présentés en 2.1. La forme de l'enveloppe est importante, car elle influence la traînée aérodynamique. Traditionnellement, les dirigeables ont la forme d'un corps profilé axisymétrique, permettant ainsi de réduire la force de traînée. La taille de l'enveloppe détermine où le di-

rigeable peut être utilisé, tandis que le volume détermine la flottabilité et la charge utile. Dans notre cas cela signifie qu'on doit faire un compromis entre le fait que le dirigeable doit être assez petit pour assurer une manoeuvrabilité en laboratoire, et avoir une flottabilité permettant de porter les composants et capteurs nécessaires. Enfin, le ratio longueur/diamètre est directement relié aux effets aérodynamiques. Un ratio plus grand signifie une traînée plus faible, mais un ratio plus petit permet d'augmenter la charge utile pour une longueur donnée.

L'enveloppe sélectionnée est l'enveloppe SB-181-200 de Windreiter. C'est une enveloppe d'un matériau multicoupe en aluminium, nylon et polyéthylène. Elle est gonflée par l'arrière et se referme avec une attache en plastique. C'est une enveloppe souple avec une forme traditionnelle ellipsoïdale. Elle a une longueur de 2 mètres et le rapport entre sa longueur et son diamètre est de 4. Elle a un volume de 259,18 litres qui lui permet de soulever environ 180 grammes lorsqu'elle est remplie d'hélium. L'enveloppe a été choisie principalement pour son ratio, soit le même que le AS800 utilisé dans le projet AURORA [10] et que le YEZ-2A utilisé dans [5].

3.1.2 Le système de commande

Le système de commande, comprenant le système de propulsion et le système de gouverne, est utilisé différement selon la vitesse du système. À basse vitesse, les gouvernes ne peuvent pas être utilisées : le système dépend alors entièrement du système de propulsion. À haute vitesse, les gouvernes peuvent être utilisées pour le contrôle des moments, et on peut donc diminuer l'utilisation de la différentiabilité des moteurs. Aucun élément du système de contrôle ne permet les déplacements latéraux, et aucun système n'est prévu pour le contrôle de tangage à basse vitesse. Le dirigeable est donc sous-actionné.

Le système de propulsion

Le système de propulsion est composé de deux moteurs de levée et deux moteurs de propulsion. Les moteurs de levée sont utilisés pour le vol stationnaire, le contrôle d'altitude et le contrôle de roulis. Les moteurs de propulsions sont utilisés pour la translation ainsi que pour le contrôle du lacet à basse vitesse. Cette configuration de moteurs permet de combiner les avantages des moteurs de levée et de propulsion, tout en évitant le poids supplémentaire d'un système de servo qui aurait permis la vectorisation des directions des moteurs. Afin que les deux moteurs de levée puissent soulever le dirigeable en limitant les moments non désirés, ils doivent être placés sous le centre de gravité. En plaçant les moteurs sur une structure fixée sous la l'enveloppe plutôt qu'en les fixant directement sur l'enveloppe, on réduit le stress causé à l'enveloppe par le poids des moteurs en plus de faciliter l'assemblage et le démontage.

L'empennage

Comme le contrôle du roulis est assuré par les moteurs de levée, il n'est pas nécessaire d'avoir 4 surfaces de contrôle. Toutefois, il est essentiel de pouvoir contrôler le tangage puisqu'aucun moteur ne permet ce contrôle. Une gouverne de direction pourrait aussi venir remplacer le contrôle par les moteurs de propulsion à haute vitesse, puisque l'utilisation de gouvernes permet la réduction de la consommation d'énergie. Ainsi, l'empennage créé est en forme de croix et composé de deux gouvernes, une de profondeur et une de direction. Le système final comprend donc quatre empennages fixes, deux gouvernes et deux servos pour assurer la rotation. Les gouvernes sont limitées à des mouvements de [-45°,+45°]. Leur conception a été inspirée par le design de la version 4 du silent-runner développé par Windreiter [45].

3.1.3 La gondole

Comme l'enveloppe n'est pas parfaitement étanche, des fuites d'hélium sont à prévoir sur des périodes prolongées. Ainsi, pour éviter le stress sur l'enveloppe et les possibles déchirures lorsque celle-ci se dégonfle, la gondole doit pouvoir facilement être retirée et remise sur l'enveloppe. Pour ce faire, la gondole est attachée à l'enveloppe à l'aide de velcro. Cela permet également d'ajuster sa position pour s'assurer que le centre de gravité se trouve bien sous le centre de volume. La gondole est également conçue pour être facilement démontable afin de permettre un accès facile aux composants internes. Elle contient les moteurs de levée et de propulsion ainsi que les composants électroniques. Comme elle contient les moteurs de levée, elle doit se trouver sous le centre de gravité pour éviter de créer un moment de tangage. La gondole est également conçue pour être aussi légère que possible, tout en étant suffisamment rigide pour supporter les moteurs et les composants électroniques. Elle est fabriquée à partir d'une surface imprimée en PLA et de tiges de carbone et de bambou. Les moteurs de propulsion sont placés juste à l'avant de la gondole, ce qui permet de réduire le moment de tangage causé par la poussée des moteurs, et de maintenir une gondole de grosseur limitée pour en réduire le poids.

3.1.4 Le centre de masse

Dans les modèles classiques de dirigeables, on présume que le centre de masse est sous le centre de volume, ce qui permet de limiter les moments créés par la force de portance et la force de gravité. Nous avons déjà établi que pour assurer un moment vertical sans création de moment de tangage, les moteurs de levée doivent se trouver sous le centre de gravité, et donc sous le centre de volume. Ajoutons que la gondole avec les moteurs et le reste des électroniques

est en un seul bloc, et qu'il y a un système d'empennage à l'arrière. Ces conditions impliquent qu'on doit avoir une masse supplémentaire à l'avant pour compenser la masse du système d'empennage. Ainsi, la batterie, qui possède une masse importante comparativement aux autres composants, est placée à l'avant du dirigeable. Il y a également la possibilité d'ajouter du lest à l'avant de l'enveloppe pour assurer un positionnement du centre de masse idéal.

3.1.5 Le système de communication

Pour pouvoir opérer le système en mode manuel à l'aide d'une radiocommande, le système de communication est composé d'un récepteur de radiocommande ELRS (ExpressLRS) qui permet de communiquer avec la radiocommande Commando 8.

Le système possède aussi un module WiFi ESP8266 qui permet au dirigeable de communiquer par wifi avec une station de base. On peut envoyer des commandes à partir de la station de base ainsi que recevoir les signaux des capteurs.

3.1.6 La conception électronique

Le système électronique est basé sur l'utilisation d'un contrôleur de vol, qui a pour mission de faciliter l'implémentation de contrôle automatique. Le contrôleur est un *PixRacer Pro*, qui est généralement utilisé avec le pilote automatique PX4. Le schéma présenté à la figure 3.1 présente la connexion entre les différentes composants électroniques. En plus des éléments mentionnés plus haut, on retrouve un régulateur de tension électronique (ou Battery Eliminator Circuit (BEC)) qui permet de réduire la tension de la batterie pour alimenter le contrôleur de vol. Il y a également un *Electronic Speed Control* (ESC), un type de variateur électronique de vitesse, qui permet de contrôler la vitesse des moteurs à partir des signaux envoyés par le contrôleur de vol.

3.1.7 Les capteurs

Les capteurs sont essentiels pour le contrôle du dirigeable. Ils fournissent des informations cruciales sur son attitude, sa vitesse et, dans certains cas, sa position, ce qui permet d'assurer un contrôle précis et stable du système.

Le PixRacer Pro est équipé de plusieurs centrales inertielles (IMU) qui combinent gyroscope, accéléromètre et magnétomètre, permettant notamment d'estimer l'orientation du dirigeable en vol. Le système est également équipé d'un GPS et d'un baromètre, qui ne peuvent toutefois pas être utilisés dans le cadre d'une utilisation intérieure.

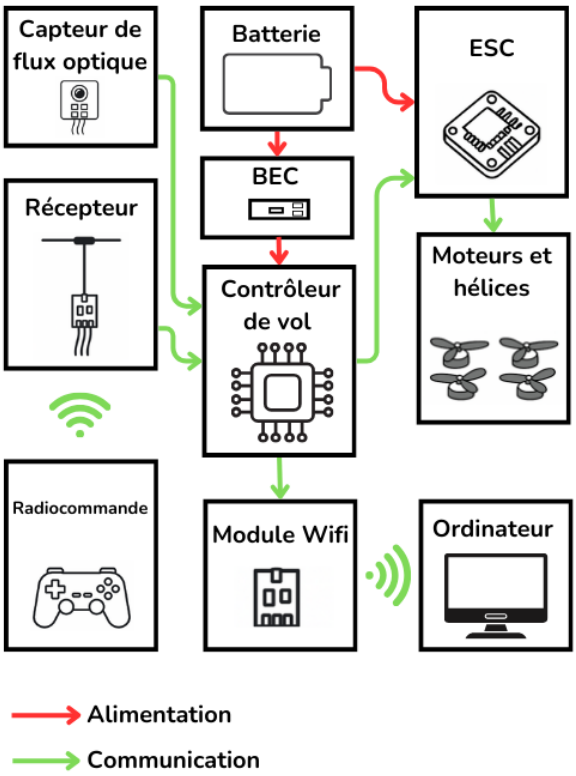

FIGURE 3.1 Électronique

Un capteur de flux optique a été ajouté afin de pouvoir mesurer la vitesse du dirigeable par rapport au sol et de déterminer sa distance avec le sol. Ce capteur est nécessaire pour le vol stationnaire, car il permet de maintenir une altitude constante. Il fonctionne en projetant un motif lumineux sur le sol et en mesurant son déplacement à l'aide d'une caméra. En analysant les variations du motif, le capteur peut estimer la vitesse du dirigeable dans les directions X et Y. Le sol doit être texturé pour identifier adéquatement la vitesse de déplacement. Le capteur de flux optique utilisé est le MTF-01, qui voit jusqu'à 8 mètres de distance avec une précision de 4 centimètres.

3.1.8 Le prototype final

La figure 3.2 présente le modèle 3D final du dirigeable, et la figure 3.3 présente une photo du prototype physique. Des pattes ont été ajoutées à la gondole pour permettre au dirigeable de se poser sur le sol sans endommager les hélices, mais elle ne se trouvent pas sur le modèle 3D.

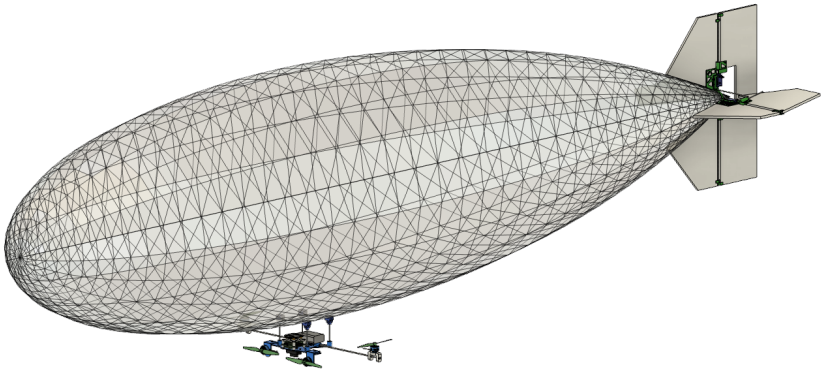

FIGURE 3.2 Design 3D de RAFALE

FIGURE 3.3 Dirigeable physique

3.2 Le modèle mathématique

Le modèle mathématique développé pour RAFALE est inspiré par les travaux de [5] et [9]. L'ensemble des variables utilisées sont définies après leur première utilisation, mais le lectorat peut toujours se référer à l'Annexe A. Le modèle peut être décrit par l'équation suivante :

$$M\frac{dx}{dt} = F_{\mathrm{d}} + F_{\mathrm{a}} + F_{\mathrm{p}} + G \tag{3.1}$$

où

- M est la matrice de masse, contenant à la fois l'inertie réelle et l'inertie ajoutée ou virtuelle.
- x contient les vitesses linéaires et angulaires et se nomme le vecteur d'état. Il s'agit d'un vecteur de 6 éléments soit x = $[\mathrm{u\ v\ w\ p\ q\ r}]^T$.
- F_{d} est le vecteur dynamique contenant les forces et moments causés par les effets de Coriolis et centrifuges.
- F_{a} est le vecteur des forces et moments aérodynamiques.
- F_{p} est le vecteur des forces et moments de propulsion.
- G est le vecteur des forces et moments dus à la flottabilité et la gravité.

Les hypothèses suivantes ont été posées pour simplifier le modèle :

- Le dirigeable est un corps rigide (on ignore les effets aéroélastiques).
- Le centre de gravité est directement en dessous du centre de volume, qui coïncide avec le centre de flottabilité.
- Le dirigeable est symmétrique par rapport au plan XZ.
- La masse est constante.
- La densité de l'air et celle de l'hélium restent constantes, une approximation valable puisque le dirigeable est utilisé à l'intérieur.

3.2.1 Cadres référentiels

Trois cadres référentiels seront utilisés dans les équations, soit le référentiel local, qui est rattaché au dirigeable, le référentiel terrestre qui suivra ici la convention Nord-Ouest-Haut, et le référentiel du vent, repère dans lequel le vent est fixe. Les référentiels terrestre et local sont représenté dans l'image suivante. L'équation 3.1 est dans le référentiel local.

La figure 3.4 présente les axes de référence du dirigeable. L'origine du dirigeable est placée au centre de volume.

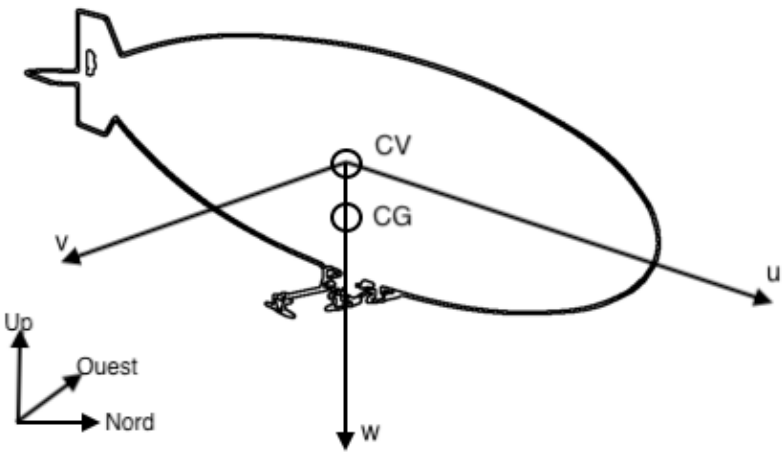

FIGURE 3.4 Axes de référence du dirigeable

3.2.2 Matrice de masses

$$M = \begin{bmatrix} m_x & 0 & 0 & 0 & ma_z - X_{\dot{q}} & 0 \\ 0 & m_y & 0 & -ma_z - Y_{\dot{p}} & 0 & ma_x - Y_{\dot{r}} \\ 0 & 0 & m_z & 0 & -ma_x - Z_{\dot{q}} & 0 \\ 0 & -ma_z - L_{\dot{v}} & 0 & I_x - L_{\dot{p}} & 0 & -J_{xz} \\ ma_z - M_{\dot{u}} & 0 & -ma_x - M_{\dot{w}} & 0 & I_y - M_{\dot{q}} & 0 \\ 0 & ma_x - N_{\dot{v}} & 0 & -J_{xz} & 0 & I_z - N_{\dot{r}} \end{bmatrix}$$

où :

- m est la masse du robot dirigeable.
- $m_x = m - X_{\dot{u}}$
- $m_y = m - Y_{\dot{v}}$
- $m_z = m - Z_{\dot{w}}$
- $X_{\dot{u}}$, $Y_{\dot{v}}$ et $Z_{\dot{w}}$ sont les termes de masse virtuelle pour les axes X, Y et Z.
- a_x, a_y et a_z sont les coordonnées du centre de gravité par rapport au centre de volume.
- I_x, I_y et I_z sont les termes d'inertie par rapport aux axes OX, OY et OZ.
- $L_{\dot{p}}$, $M_{\dot{q}}$ et $N_{\dot{r}}$ sont les termes d'inertie virtuelle par rapport aux axes OX, OY et OZ.
- $J_x = I_x - L_{\dot{p}}$
- $J_y = I_y - M_{\dot{q}}$

— $J_z = I_z - N_{\dot{r}}$
— $J_{xz} = I_{xz} + N_{\dot{p}} = I_{xz} + L_{\dot{r}}$, où I_{xz} est le produit d'inertie par rapport à l'axe OY.
— I_{xy} et I_{yz} sont nuls car le centre de gravité et le centre de volume sont sur la même ligne.

3.2.3 Forces dynamiques

Ce vecteur contient les forces et moments causés par les effets de Coriolis et centrifuges. Il peut être représenté ainsi :

$$F_d = [F_1 \quad F_2 \quad F_3 \quad F_4 \quad F_5 \quad F_6]^T$$

où :
— $F_1 = -m_z wq + m_y rv + m[a_x(q^2 + r^2) - a_z rp]$
— $F_2 = -m_x ur + m_z pw + m[-a_x pq - a_z rq]$
— $F_3 = -m_y vp + m_x qu + m[-a_x rp + a_z(q^2 + p^2)]$
— $F_4 = -(J_z - J_y)rq + J_{xz}pq + ma_z(ur - pw)$
— $F_5 = -(J_x - J_z)pr + J_{xz}(r^2 - p^2) + m[a_x(vp - qu) + a_z(wq - rv)]$
— $F_6 = -(J_y - J_x)qp - J_{xz}qr + m[-a_x(ur - pw)]$

3.2.4 Forces de propulsion

Le vecteur des forces de propulsion est donné par :

$$F_p = [X_{prop} \quad Y_{prop} \quad Z_{prop} \quad L_{prop} \quad M_{prop} \quad N_{prop}]^T$$

Selon la configuration des moteurs de notre modèle,

— $X_{prop} = T_{ds} + T_{dp}$, poussée du moteur le long de l'axe OX
— $Y_{prop} = 0$, poussée du moteur le long de l'axe OY
— $Z_{prop} = -(T_{ls} + T_{lp})$, poussée du moteur le long de l'axe OZ
— $L_{prop} = (T_{lp} - T_{ls})d_{yl}$, moment de poussée du moteur le long de l'axe OX (roulis)
— $M_{prop} = (T_{ds} - T_{dp})d_{zd} - (T_{lp} + T_{ls})d_{xl}$, moment de poussée du moteur le long de l'axe OY (tangage)
— $N_{prop} = (T_{dp} - T_{ds})d_{yd}$, moment de poussée du moteur le long de l'axe OZ (lacet)
— T_{ds} est la poussée du moteur de propulsion tribord
— T_{dp} est la poussée du moteur de propulsion bâbord
— T_{ls} est la poussée du moteur de levage tribord

— T_{lp} est la poussée du moteur de levage bâbord
— d_x est la distance horizontale du centre de volume à l'hélice le long de l'axe principal du dirigeable
— d_y est la distance horizontale du centre de volume à l'hélice perpendiculaire à l'axe principal du dirigeable
— d_z est la distance verticale du centre de volume à l'hélice

Les forces de propulsion sont influencées par la force de l'air, mais pour simplifier le modèle, on les considère comme indépendantes [3].

3.2.5 Forces aérodynamiques

Le modèle aérodynamique du dirigeable vient d'une adaptation de [9] des données obtenues en soufflerie pour le dirigeable YEZ-2A de Westinghouse [46]. Il est possible de s'inspirer de ces modèles puisque, bien que la forme de l'enveloppe soit un peu différente, le rapport entre la longueur et le diamètre de l'enveloppe de RAFALE est de 4, tout comme le YEZ-2A et le AS800 utilisé pour AURORA.

Le modèle aérodynamique présenté inclut :

1. Les forces et moments aérodynamiques dus au déplacement du dirigeable dans l'air.
2. Les facteurs d'amortissement des forces et des moments dynamiques.

Il exclut les effets de la masse ajoutée, qui sont plutôt inclus directement dans la matrice de masse.

Forces et moments aérodynamiques

Lorsqu'on fait des mesures en soufflerie pour des dirigeables de grande taille, on utilise un modèle réduit pour mesurer les forces et moments subis par le système selon la vitesse du vent. On divise ensuite ces valeurs par des forces et des moments de référence qui permettent d'obtenir des coefficients non dimensionnels s'appliquant à tous les dirigeables ayant une forme similaire, peu importe la taille. C'est ce qui nous permet d'utiliser les résultats obtenus dans des travaux précédents pour notre modèle. Les forces de traînée, latérale et de portance ainsi que les moments aérodynamiques sont donc calculés dans le référentiel du vent relatif. La vitesse relative de l'air dans le référentiel du corps est $x_a = [u_a \ v_a \ w_a \ p_a \ q_a \ r_a]$ et tous les forces et moments aérodynamiques dépendent de la vitesse vraie V_t, soit la norme de la

vitesse relative du vent.

$$V_t = \sqrt{u_a^2 + v_a^2 + w_a^2}$$

Les coefficients mesurés en soufflerie dépendent de l'angle d'incidence (α), de l'angle de dérapage (β) et des angles de déflexion des gouvernes de profondeur (d_e) et de direction (d_r).

L'angle d'incidence correspond à :

$$\alpha = arctan(w_a/u_a)$$

Et l'angle de dérapage est

$$\beta = arcsin(v_a/V_t)$$

Les forces aérodynamiques calculées dans le référentiel du vent équivalent à :

$$F_a^w = \begin{bmatrix} A_X \\ A_Y \\ A_Z \\ A_L \\ A_M \\ A_N \end{bmatrix} = \begin{bmatrix} C_d F_{\text{ref}} \\ C_Y F_{\text{ref}} \\ C_l F_{\text{ref}} \\ C_L M_{\text{ref}} \\ C_M M_{\text{ref}} \\ C_N M_{\text{ref}} \end{bmatrix}$$

A_X, A_Y et A_Z sont les forces de traînée, latérale et de portance, A_L, A_M et A_N sont les moments de roulis, tangage et lacet, et C_D, C_Y, C_L, C_l, C_m et C_n sont les coefficients correspondants.

La force de référence est :

$$F_{\text{ref}} = \frac{1}{2}\rho V_t^2 S_{ref}$$

Et le moment de référence est

$$M_{\text{ref}} = \frac{1}{2}\rho V_e^2 S_{ref} L_{ref}$$

Où $S_{ref} = \text{Vol}^{2/3}$ est la surface de référence, Vol est le volume du dirigeable et ρ est la densité de l'air. Le moment de référence est exprimé en fonction de la vitesse équivalente V_e plutôt que la vitesse vraie V_t. Cette vitesse équivalente permet de mieux considérer les moments

aérodynamiques lorsque la vitesse angulaire est importante.

$$V_e = \sqrt{V_t^2 + \frac{(p_a^2 + q_a^2 + r_a^2)L_{\text{ref}}^2}{4}}$$

L_{ref} est une longueur de référence qui conventionnellement est définie comme $\text{Vol}^{1/3}$.

Les coefficients utilisés proviennent des tables de coefficients de [46] et des travaux de [21] permettant d'extrapoler les tables.

Tableau 3.1 Coefficients utilisés pour calculer les forces aérodynamiques

Coefficient	**Définition**
C_{d_0}	Coefficient de traînée à $\alpha = 0$ et $\beta = 0$
C_{d_i}	Coefficient de traînée induite (traînée causée par la présence de α et β)
C_{Y_b}	Coefficient de force latérale en fonction de β
$C_{Y_{dr}}$	Coefficient de force latérale en fonction de la déflexion de la gouverne de direction
C_{l_0}	Coefficient de portance à $\alpha = 0$
C_{l_a}	Coefficient de portance en fonction de α
$C_{l_{de}}$	Coefficient de portance en fonction de la déflexion de la gouverne de profondeur
C_{M_0}	Coefficient de tangage à $\alpha = 0$
C_{M_a}	Coefficient de tangage en fonction de α
$C_{M_{ab}}$	Coefficient de tangage en fonction de β dû à α (effet de couplage croisé)
$C_{M_{ba}}$	Coefficient de tangage en fonction de α dû à β (effet de couplage croisé)
C_{M_b}	Coefficient de tangage en fonction de β
$C_{M_{de}}$	Coefficient de tangage en fonction de la déflexion de la gouverne de profondeur
C_{N_b}	Coefficient de lacet en fonction de β
$C_{N_{dr}}$	Coefficient de lacet en fonction de la déflexion de la gouverne de direction
C_{L_b}	Coefficient de roulis en fonction de β

C_{l_0} et C_{M_0} sont nuls puisqu'on estime qu'aucune force de portance ou tangage n'est présente à un angle d'incidence de 0. À partir de ces coefficients on peut ensuite obtenir les coefficients aérodynamiques recherchés. On utilise les coefficients présentés dans le tableau 3.1 pour calculer les coefficients utilisés dans la définition de F_a^w.

$$
\begin{aligned}
C_d &= C_{d_0} + C_{d_i}(C_{l_a}^2 s_1(\alpha) + C_{Y_b}^2 s_2(\beta)) \\
C_Y &= C_{Y_b} s_3(\beta) + C_{Y_{d_r}} d_r |\cos(\alpha)\cos(\beta)| \\
C_l &= C_{l_0}(1 - s_2(\beta)) + C_{l_a} s_4(\alpha)(1 + |s_3(\beta)|) + C_{l_{d_e}} d_e \cos(\alpha)\cos(\beta) \\
C_L &= C_{L_b} s_5(\beta) s_6(\alpha) \\
C_M &= C_{M_0} + (C_{M_a} + C_{M_{ab}} |\sin(\beta)|) s_7(\alpha) + (C_{M_b} + C_{M_{ba}} \sin(\alpha)) |s_3(\beta)| + C_{M_{d_e}} d_e \cos(\alpha)\cos(\beta) \\
C_N &= C_{N_b} s_8(\beta) s_6(\alpha) + C_{N_{dr}} d_r \cos(\alpha)\cos(\beta)
\end{aligned}
$$

Où s_1 ... s_8 sont des courbes permettant de mettre en lien les résultats obtenus en soufflerie. Ces équations permettent aussi d'étendre les résultats obtenus de [46] et [21], à des angles d'incidence plus élevés.

Facteurs d'amortissement des forces et des moments dynamiques

Les coefficients dynamiques utilisés pour calculer les termes d'amortissement sont les dérivées des coefficients aérodynamiques selon des taux de giration non dimensionnels q^* et r^*, qui sont exprimés en fonction des vitesses angulaires q et r dans le référentiel du vent relatif.

$$
\begin{aligned}
q^* &= \frac{q_a L_{\text{ref}}}{V_t} \\
r^* &= \frac{r_a L_{\text{ref}}}{V_t}
\end{aligned}
$$

On utilise les coefficient dynamiques C_{Yr}, C_{nr}, C_{Zq} et C_{mq}. Ainsi la modification des forces et moments par le facteur d'amortissement correspond à :

$$
F_{amortissement} = \begin{bmatrix} 0 \\ C_{Yr} F_{\text{ref}} r^* \\ C_{Zq} F_{\text{ref}} q^* \\ 0 \\ C_{mq} M_{\text{ref}} q^* \\ C_{nr} M_{\text{ref}} r^* \end{bmatrix}
$$

Vecteur de forces aérodynamique

Les forces aérodynamiques ayant été calculées dans le référentiel du vent, elles doivent être multipliées par une matrice de transformation S_1 pour pouvoir être reportées dans le référentiel local et ainsi être additionnées dans l'équation 3.1.

$$F_a = S_1(F_a^w + F_{amortissement})$$

3.2.6 Forces de flottabilité et gravité

Le vecteur G représente la différence entre la force de gravité et la force de flottabilité.

$$G = \begin{bmatrix} S_2(W - B) \\ OC \times S_2 W \end{bmatrix}$$

Où W est le vecteur de force de gravité, B est le vecteur de force de flottabilité, OC est le vecteur de distance entre le centre de volume (où est appliqué B) et le centre de gravité (où est appliqué W). S_2 est la matrice de transformation permettant de transposer les résultats du référentiel terrestre au référentiel local.

3.3 Le simulateur

Le simulateur utilisé pour RAFALE est le simulateur MuJoCo [30]. Le modèle 3D présenté dans la figure 3.2 est séparé en grandes sections qui sont chacunes importées sous forme de maillage 3D. Un fichier XML permet de mettre en relation ces maillages et de déterminer les jointures et les actionneur du système. Comme MuJoCo est un moteur physique, une fois le modèle bien défini dans le XML, il s'occupe de calculer les forces dynamiques, les forces dues à la gravité et les forces de propulsion. Pour ajouter la force de flottabilité, il existe un paramètre `gravcomp` qui permet d'ajouter une force au centre de volume correspondant à une fraction du poids de l'objet. En calculant la force de flottabilité de l'enveloppe et le poids de l'enveloppe, on peut donc trouver le paramètre `gravcomp` permettant de simuler la force de flottabilité. Comme expliqué dans 2.3, MuJoCo n'a pas de modèle adéquat de mécanique des fluides permettant de calculer les forces aérodynamqiues sur le dirigeable. Celles-ci sont donc calculées dans un code python puis appliquées au modèle à l'aide de `dm_control`. La figure 3.5 permet de visualiser l'intégration des forces aérodynamiques au modèle MuJoCo.

Pour les calculs de flottabilité et d'aérodynamique, on considère la densité de l'hélium et

de l'air à 20°C et à pression normale. On assume que la température et la pression restent constantes, ce qui est raisonnable pour une utilisation intérieure. Elles sont respectivement $\rho_{\text{hélium}} = 0.1634\text{g/L}$ et $\rho_{\text{air}} = 1.225\text{g/L}$. Les valeurs de référence utilisées pour les calculs aérodynamiques sont $\text{Vol} = 0.2592\text{m}^3$, $S_{\text{ref}} = 0.4065\text{m}$ et $L_{\text{ref}} = 0.6376\text{m}$.

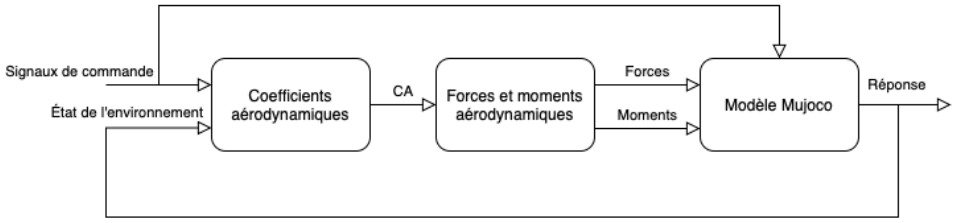

FIGURE 3.5 Intégration de l'aérodynamique avec MuJoCo

Afin d'évaluer la performance du simulateur on utilise la mesure du facteur temps réel (RTF), défini par :

$$\text{RTF} = \frac{\text{Temps simulé}}{\text{Temps d'exécution}}$$

Le RTF a été estimé en moyennant le RTF sur 1000 épisodes de 4000 pas de temps chacun. Il est d'environ 12,04 pour notre simulateur, ce qui signifie que 12 secondes de déplacement du dirigeable peuvent être simulées en 1 seconde.

CHAPITRE 4 RÉSULTATS THÉORIQUES ET EXPÉRIMENTAUX

Afin d'évaluer la capacité de notre nouveau système à apprendre des politiques de contrôle, il a été entraîné à réaliser deux tâches. La première consiste à atteindre une altitude désirée puis y rester, et la deuxième demande au système de se déplacer à une vitesse et une altitude constante. Ces deux tâches permettent de bien visualiser l'effet des algorithmes d'apprentissage par renforcement sur notre système, ce qui permettra éventuellement de développer des algorithmes plus complets de suivi de trajectoires. Afin de réaliser ces tâches, deux environnements ont été créés à l'aide de Gymnasium [33].

4.1 Les environnements

Deux environnements ont été créés pour les deux tâches à accomplir. Les deux environnements partagent leur état initial, leurs conditions d'arrêts et la logique derrière la définition de l'espace d'actions. Il est à noter que les vitesses utilisées dans les deux environnements sont dans le référentiel terrestre.

État initial

L'état initial pour l'ensemble des expériences est une position de [0, 0, 3] avec des angles de tangage, lacet et roulis à 0.

Conditions d'arrêt

Quatre conditions d'arrêt ont été définies pour notre système. L'épisode se termine :

1. lorsqu'il atteint le pas de temps maximal défini pour la tâche ($\max_{\text{pas}}$) ;
2. lorsque l'angle de tangage est supérieur à 45° ;
3. lorsque l'angle de roulis est supérieur à 45° ;
4. lorsque l'altitude est inférieure ou égale à un mètre.

Les conditions 2 et 3 permettent de maintenir le système dans un état sécuritaire en prévenant des angles de roulis et de tangage trop importants, au-delà desquels le contrôle est plus difficile, voire impossible. La dernière condition est présente puisque les problèmes abordés ne sont pas des problèmes de décollage et d'atterrissage.

Lorsqu'un épisode se termine avec une des conditions 2-4, l'agent reçoit une pénalité de $p_{\text{échec}}$.

Espace d'actions

Pour les deux tâches sélectionnées, l'espace d'actions choisi est un espace discret, comme dans [47], ce qui a deux avantages. Premièrement, l'espace d'actions est beaucoup plus petit et donc on limite l'exploration. Ensuite, cela permet d'assurer un passage progressif entre les différentes commandes envoyées aux actionneurs, et donc d'éviter les variations brusques. L'espace d'actions est adapté selon la tâche.

4.1.1 Vol stationnaire à altitude spécifiée

L'objectif de cette tâche est d'apprendre au dirigeable à faire du vol stationnaire à une certaine altitude. Le dirigeable se déplace donc dans un premier temps à la position demandée, puis doit y rester. Pour l'entraînement, l'altitude est choisie aléatoirement dans une plage d'altitudes prédéfinie.

Espace d'actions

Puisque l'objectif de cette tâche est le vol stationnaire et le contrôle d'altitude, la commande des gouvernes n'est pas nécessaire, on ne permet donc que le contrôle des moteurs. Les actions sont celles présentées dans le tableau 4.1 :

Tableau 4.1 Table de correspondance des actions pour la tâche de vol stationnaire

n	**Action**	δ[$\mathbf{T}_{ls}$, $\mathbf{T}_{lp}$, $\mathbf{T}_{lp}$, $\mathbf{T}_{lp}$, **dr, de**]
1	Neutre	[0, 0, 0, 0, 0, 0]
2	Accélérer (z)	[0.05, 0.05, 0, 0, 0, 0]
3	Ralentir (z)	[-0.05, -0.05, 0, 0, 0, 0]
4	Accélérer (x)	[0, 0., 0.05, 0.05, 0, 0]
5	Ralentir (x)	[0, 0, -0.05, -0.05, 0, 0]
6	Gauche (moteur)	[0, 0, -0.05, 0.05, 0.0, 0.0]
7	Droite (moteur)	[0, 0, 0.05, -0.05, 0, 0]

Récompense

La récompense pour chaque pas de temps contient 4 parties :

`recompense_survie`, qui donne une récompense fixe (r_{survie}) à chaque pas de temps où le dirigeable est encore en vie, essentiel pour s'assurer de récompenser le système pour chaque moment où il n'atteint pas une condition d'arrêt.

`punition_position`, qui soustrait à la récompense une valeur proportionnelle à la distance entre le dirigeable et la position désirée (Δ_{pos}).

`bonus_zone_cible`, qui additionne une valeur de récompense lorsque le dirigeable se trouve près de la position désirée. Cette valeur encourage le dirigeable à rester en vol stationnaire après avoir atteint la position désirée. La fonction permettant de représenter si la fonction se trouve dans la zone cible est :

$$1_{\text{zone cible}} = \begin{cases} 1 & \text{si } \Delta_{\text{pos}} < \epsilon \\ 0 & \text{sinon} \end{cases}$$

où ϵ est un paramètre à définir.

`punition_orientation`, qui soustrait à la récompense une valeur proportionnelle à la différence entre les angles de tangage, de roulis et de lacet désirés et les angles réels ($\Delta\theta$).

La récompense totale est donc :

$$r_{(t)}^{\text{altitude}} = r_{\text{survie}} - w_{\text{position}}\Delta_{pos} - w_{\text{bonus zone}}1_{\text{zone cible}} - w_{\text{orientation}}\Delta\theta$$

où w_{position}, $w_{\text{bonus zone}}$, $w_{\text{orientation}}$ sont les poids attribués à chaque partie de la fonction de récompense.

Espace d'observation

L'espace d'observation utilisée par l'algorithme d'apprentissage contient les éléments suivants :

- Différence entre la position désirée et la position réelle en x.
- Différence entre la position désirée et la position réelle en y.
- Différence entre la position désirée et la position réelle en z.
- Vitesse en x.
- Vitesse en y.
- Vitesse en z.
- Différence entre l'angle désiré et l'angle réel de lacet.
- Différence entre l'angle désiré et l'angle réel de tangage.
- Différence entre l'angle désiré et l'angle réel de roulis.
- Vitesse angulaire lacet.
- Vitesse angulaire tangage.
- Vitesse angulaire roulis.

4.1.2 Vol de croisière à vitesse spécifiée

Cette tâche permet d'évaluer la capacité d'un algorithme d'apprentissage par renforcement à voler en mode croisière, et ainsi à utiliser les gouvernes à l'arrière. L'objectif de la tâche est d'apprendre au dirigeable à se déplacer en vol de croisière à une vitesse en x demandée (référentiel terrestre), tout en restant à l'altitude de départ.

Espace d'actions

Contrairement à la tâche de vol stationnaire, cette tâche nécessite l'utilisation des gouvernes, précisément la gouverne de profondeur qui est essentielle ici pour compenser le tangage créé par les moteurs permettant d'avancer. Les actions sont celles présentées dans le tableau 4.2 :

TABLEAU 4.2 Table de correspondance des actions pour la tâche de vol de croisière

n	**Action**	δ[$\mathbf{T}_{ls}$, $\mathbf{T}_{lp}$, $\mathbf{T}_{lp}$, $\mathbf{T}_{lp}$, **dr, de**]
1	Neutre	[0, 0, 0, 0, 0, 0]
2	Accélérer (z)	[0.05, 0.05, 0, 0, 0, 0]
3	Ralentir (z)	[-0.05, -0.05, 0, 0, 0, 0]
4	Accélérer (x)	[0, 0., 0.05, 0.05, 0, 0]
5	Ralentir (x)	[0, 0, -0.05, -0.05, 0, 0]
6	Gauche (moteur)	[0, 0, -0.05, 0.05, 0.0, 0.0]
7	Droite (moteur)	[0, 0, 0.05, -0.05, 0, 0]
8	Gauche (gouverne)	[0, 0, 0, 0, 0.05, 0]
9	Droite (gouverne)	[0, 0, 0, 0, -0.05, 0]
10	Haut (gouverne)	[0, 0, 0, 0, 0, 0.05]
11	Bas (gouverne)	[0, 0, 0, 0, 0, -0.05]

Récompense

La récompense pour chaque pas contient quatre parties :

`récompense_survie`, qui donne une récompense fixe (r_{survie}) à chaque pas de temps où le dirigeable est encore en vie, essentiel pour s'assurer de récompenser le système pour chaque moment où il n'atteint pas une condition d'arrêt.

`punition_vitesse`, qui soustrait à la récompense une valeur proportionnelle à la différence entre la vitesse désirée du dirigeable et la vitesse réelle ($\Delta\mathbf{v}$).

`punition_altitude`, qui soustrait à la récompense une valeur proportionnelle à la différence entre l'altitude désirée du dirigeable et l'altitude réelle (Δz).

`punition_orientation`, qui soustrait à la récompense une valeur proportionnelle à la différence entre les angles de tangage, de roulis et de lacet désirés et les angles réels ($\Delta\theta$).

La récompense totale est donc :

$$r_{(t)}^{\text{vitesse}} = r_{\text{survie}} - w_{\text{vitesse}}\Delta\mathbf{v} - w_{\text{altitude}}\Delta z - w_{\text{orientation}}\Delta\theta$$

où w_{vitesse}, w_{altitude}, $w_{\text{orientation}}$ sont les poids attribués à chaque partie de la fonction de récompense.

Espace d'observation

L'espace d'observation utilisée par l'algorithme d'apprentissage contient les éléments suivants :

- Position en x.
- Position en y.
- Position en z.
- Différence entre la vitesse désirée et la vitesse réelle en x.
- Différence entre la vitesse désirée et la vitesse réelle en y.
- Différence entre la vitesse désirée et la vitesse réelle en z.
- Différence entre l'angle désiré et l'angle réel de lacet.
- Différence entre l'angle désiré et l'angle réel de tangage.
- Différence entre l'angle désiré et l'angle réel de roulis.
- Vitesse angulaire lacet.
- Vitesse angulaire tangage.
- Vitesse angulaire roulis.

4.2 Résultats

Des modèles d'apprentissage par renforcement ont été entraînés sur les deux environnements décrits précédemment. Les modèles entraînés ont été créés à l'aide de la librairie `Stable Baseline3`, qui offre une façon rapide et facile d'entraîner des algorithmes sur différents modèles. L'approche utilisée pour les deux environnements est un peu différente. Pour le premier environnement, on a comparé des modèles entraînés avec PPO, DQN et A2C, et on a ensuite comparé la performance de plusieurs modèles entraînés avec PPO. Pour le deuxième environnement, on ne fait que comparer des modèles créés avec PPO, et on visualise sa performance dans différentes situations.

Les algorithmes PPO, DQN et A2C ont été sélectionnés en raison de leur capacité à accepter les espaces d'actions discrets. L'algorithme Proximal Policy Optimization, ou PPO, [48] est un algorithme populaire en raison de sa stabilité et de son efficacité. C'est un algorithme de DRL populaire pour application sur les dirigeables, comme on peut voir dans la section 2.4.

Pour chaque entraînement N_{env} copies de l'environnement sont créées puis exécutées en parallèle pour un nombre total de N_{pas} chacun. Cette méthode permet de stabiliser l'apprentissage.

4.2.1 Vol stationnaire à altitude spécifiée

Pour tester l'apprentissage par renforcement sur la tâche de vol stationnaire, trois algorithmes d'apprentissage par renforcement ont été testés, permettant ainsi de tirer profit de la facilité d'appliquer différents algorithmes avec `Stable Baseline3`. Les 3 algorithmes sélectionnés sont PPO, A2C et DQN.

Les paramètres utilisés pour l'entraînement sont :

Tableau 4.3 Paramètres de l'environnement pour la tâche de vol stationnaire

Paramètre	**Valeur**
N_{env}	5
N_{pas}	1000000
r_{survie}	1
$w_{position}$	0.5
$w_{bonus\ zone}$	0.5
ϵ	0.05
$w_{orientation}$	$1/\pi$
$p_{échec}$	-100
Plage d'altitudes	[2, 4]
max_{pas}	2000

Les résultats de l'entraînement sont présentés dans la figure 4.1.

On peut déjà observer que DQN performera moins bien que les autres algorithmes sur la tâche considérée.

Au cours de l'entraînement, un modèle est sauvegardé tous les 1 million de pas de temps. À partir de ces modèles intermédiaires, la performance de chaque algorithme est évaluée à différents stades d'entraînement, sur 100 épisodes où la cible est échantillonnée aléatoirement entre 2 et 4 mètres d'altitude. Afin d'évaluer si le modèle est capable de généraliser la tâche à accomplir, on évalue sa performance sur 6000 pas de temps plutôt que les 2000 utilisés pour l'entraînement. On trace ensuite le retour moyen pour chaque modèle ainsi que la longueur

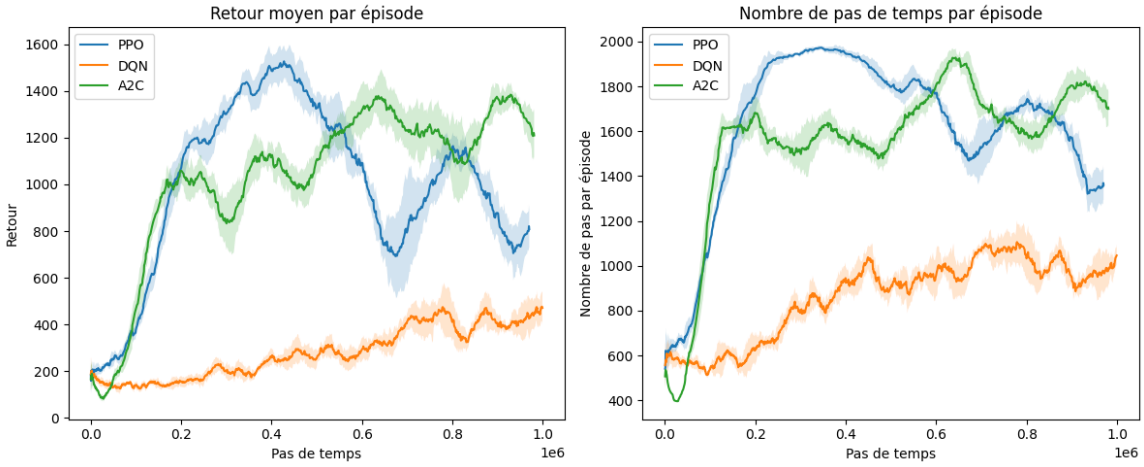

FIGURE 4.1 Courbes d'entraînement pour la tâche de vol stationnaire à altitude donnée

moyenne de chaque épisode, et on obtient la figure 4.2.

On voit sur la figure que le modèle ayant la meilleure performance est le modèle PPO après 200000 pas d'entraînement.

Trois autres modèles ont été entraînés avec PPO, mais cette fois-ci avec 10 environnements exécutés en parallèle, pour un total de 10 millions de pas de temps. L'évaluation de tous les modèles créés pour PPO sur 100 épisodes, pour un pas de temps maximal de 6000 par épisode et pour des valeurs aléatoires d'altitude entre 2 et 4 est présenté à la figure 4.3.

On observe également dans cette image que le modèle le plus performant est le modèle PPO entraîné avec la *seed* = 0 sur 200000 pas de temps. La figure 4.4 permet de visualiser la performance du modèle. Celui-ci est évalué sur huit valeurs d'altitude, dont trois en dehors de la plage d'altitudes utilisée pendant l'entraînement.

Trois constats peuvent être tirés de ces graphiques. Premièrement, pour les altitudes situées dans la plage d'entraînement, le modèle obtient de bons résultats et parvient même à performer sur des épisodes trois fois plus longs que ceux utilisés lors de l'entraînement. Deuxièmement, bien que le modèle montre une bonne performance pour une altitude cible de 4.5m, il ne parvient pas à se généraliser aux altitudes cibles de 5m ou 1.5m. Les courbes indiquent qu'il atteint dans ces deux cas une condition d'arrêt en moins de 1000 pas de temps. Troisièmement, on voit que le modèle a plus de difficulté à se stabiliser lorsque l'altitude cible se trouve sous sa position de départ : des importants moments de tangage apparaissent dans ces situations, contrairement aux cas où l'altitude cible est supérieure à la position initiale.

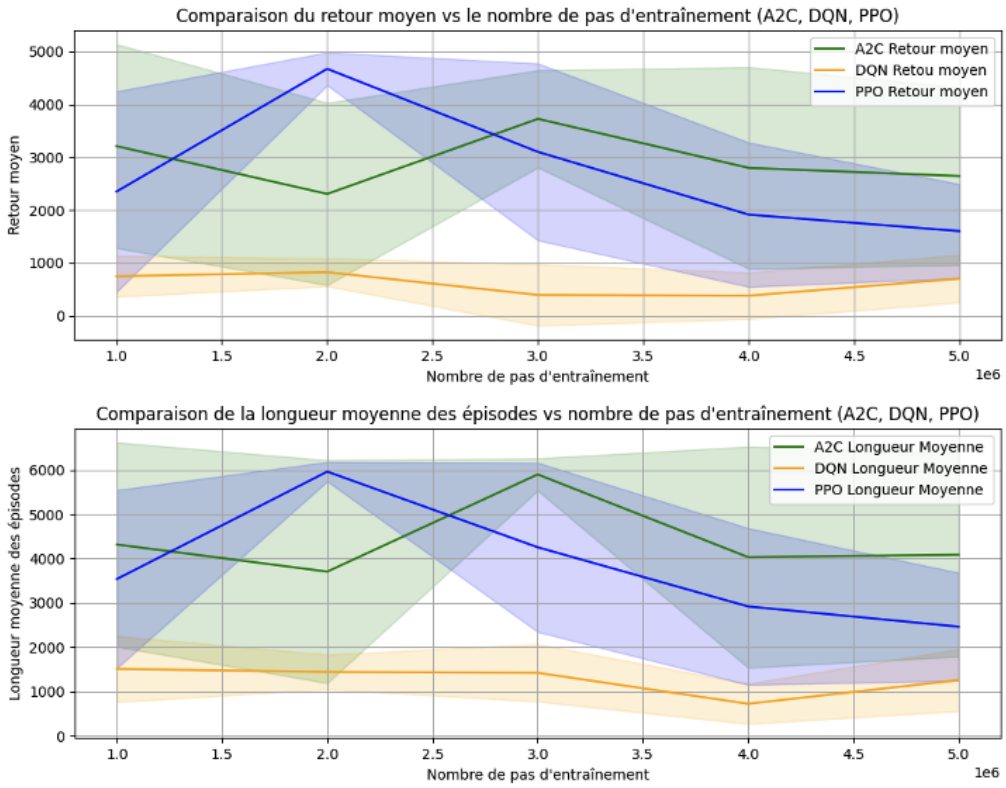

FIGURE 4.2 Évaluation des modèles à différents stades d'entraînement (cibles entre 2 et 4 m)

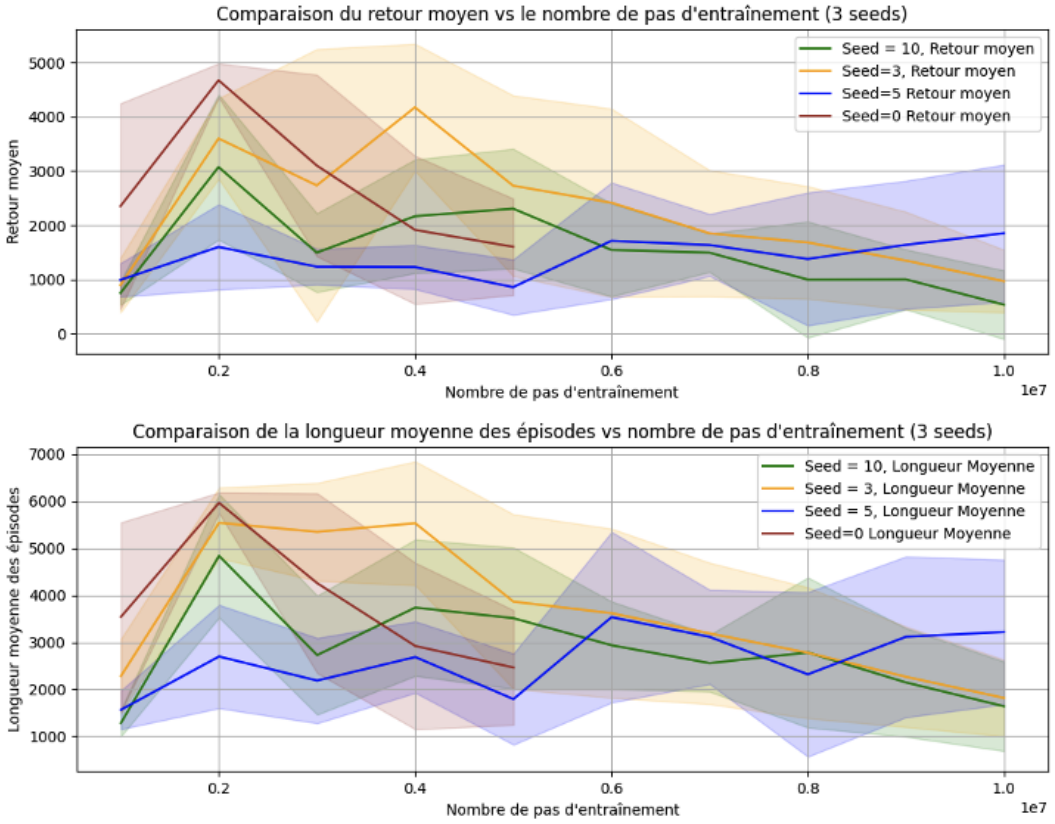

FIGURE 4.3 Évaluation des modèles entraînés avec PPO à différents stades d'entraînement (cibles entre 2 et 4 m)

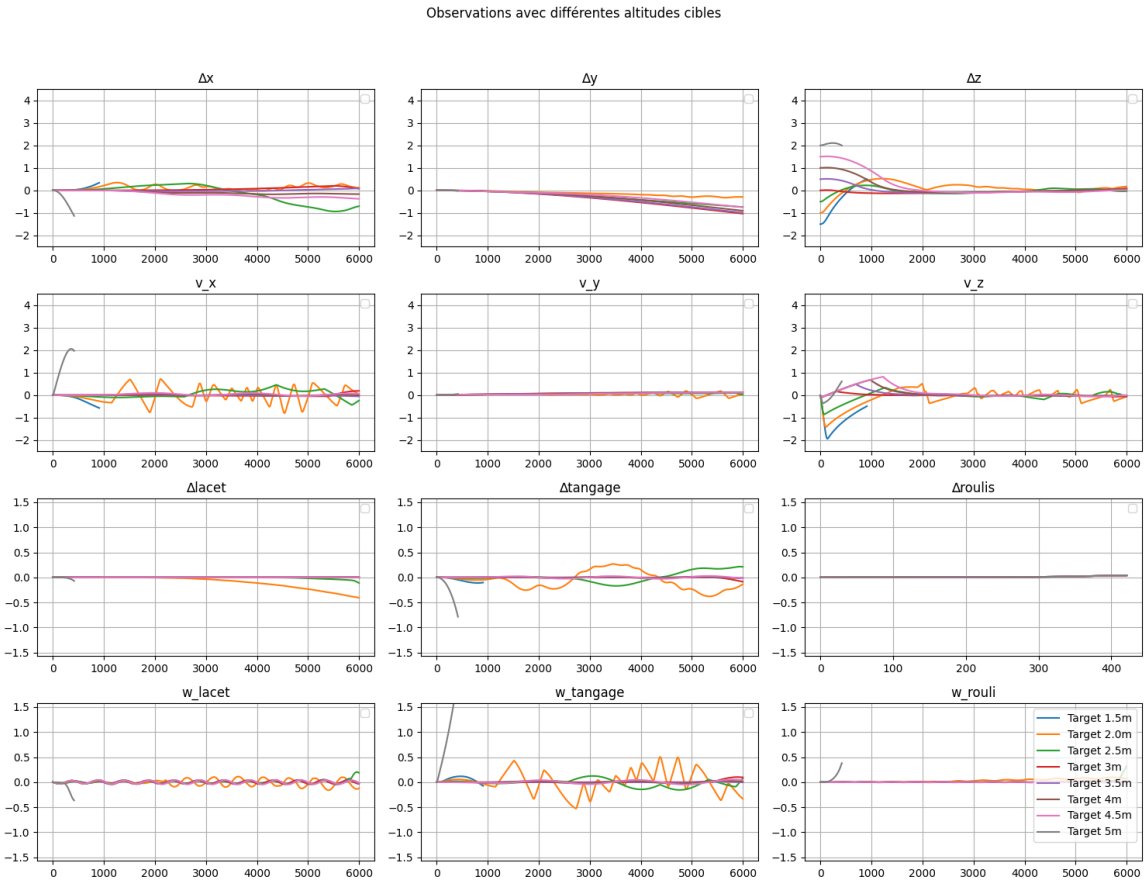

FIGURE 4.4 Observations selon le pas de temps pour différentes altitudes

4.2.2 Vol de croisière à vitesse spécifiée

Pour cette tâche, un algorithme PPO a été entraîné sur 3 *seeds* différents.

Les paramètres utilisés pour l'entraînement sont présentés aux tableaux 4.4.

TABLEAU 4.4 Paramètres de l'environnement pour la tâche de vol de croisière

Paramètre	**Valeur**
N_{env}	10
N_{pas}	1000000
r_{survie}	1
$w_{vitesse}$	0.3
$w_{altitude}$	0.5
$w_{orientation}$	$1/\pi$
$p_{échec}$	-100
Plage de vitesses	[0.1, 1.5]
max_{pas}	2000

Les modèles entraînés ont été testés sur 100 épisodes où la vitesse cible était échantillonnée aléatoirement entre 0.1 et 1.5 m/s. Comme les trois modèles présentaient des performances similaires, un seul a été sélectionné aléatoirement. Les résultats de l'entraînement de ce modèle sont présentés à la figure 4.5.

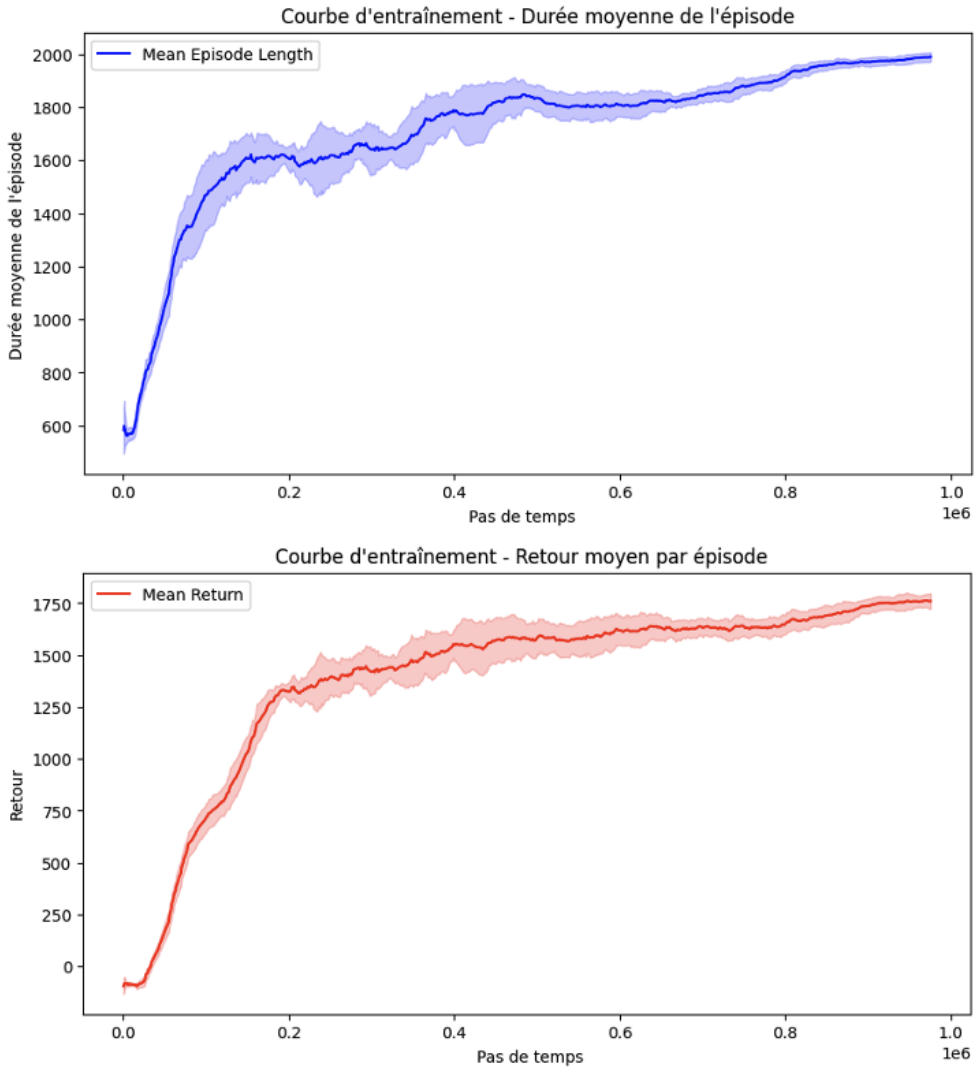

FIGURE 4.5 Résultats de l'entraînement pour la tâche de vol en mode croisière

Le modèle obtenu a été testé pour trois vitesses cibles différentes, et les résultats sont présentés dans les figures 4.6, 4.7 et 4.8. L'examen de ces trois figures nous permet de constater que le système réussit à maintenir la vitesse et l'altitude demandées pour les trois vitesses explorées. Le nombre de pas maximal par épisode pour l'entraînement était de 2000, mais l'évaluation est faite sur des épisodes de 4000 pas pour vérifier si les résultats se généralisent. Pour les petites vitesses (0.1 m/s et 0.5m/s) on obtient de bons résultats avec des vitesses en x et en z somme toute constantes et une certaine compensation des moments de tangage. Toutefois, à plus haute vitesse (1 m/s), les moments créés sont trop grands, en plus d'être exacerbés par la plus forte présence des effets aérodynamiques à grande vitesse, et le système ne réussit pas à se rendre à 4000 pas, et termine abruptement à 2676 pas en raison d'une violation à la condition sur l'angle maximal de tangage. Un entraînement sur des épisodes plus longs serait donc nécessaire pour apprendre au système à compenser ces importants moments.

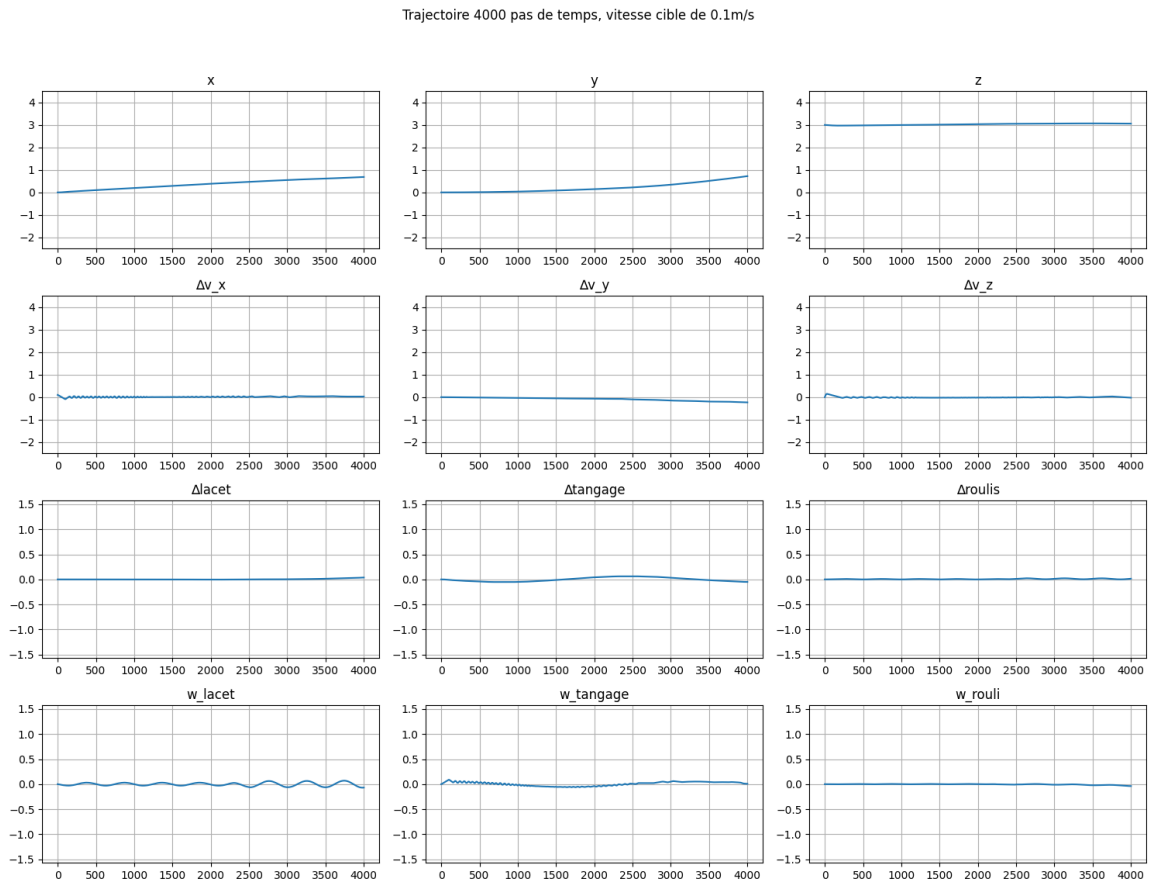

FIGURE 4.6 Visualisation des observations générées par le modèle entraîné : 4000 pas de temps et vitesse cible de 0.1 m/s

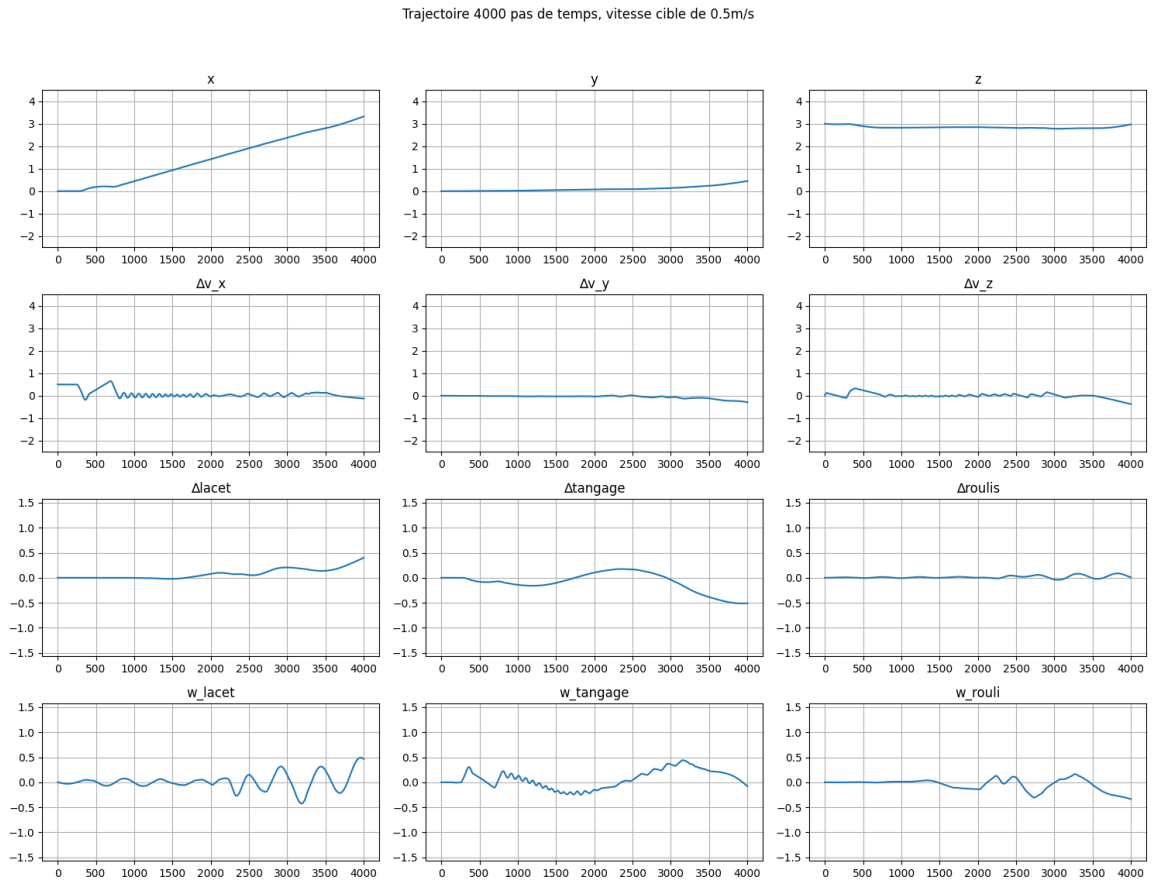

FIGURE 4.7 Visualisation des observations générées par le modèle entraîné : 4000 pas de temps et vitesse cible de 0.5 m/s

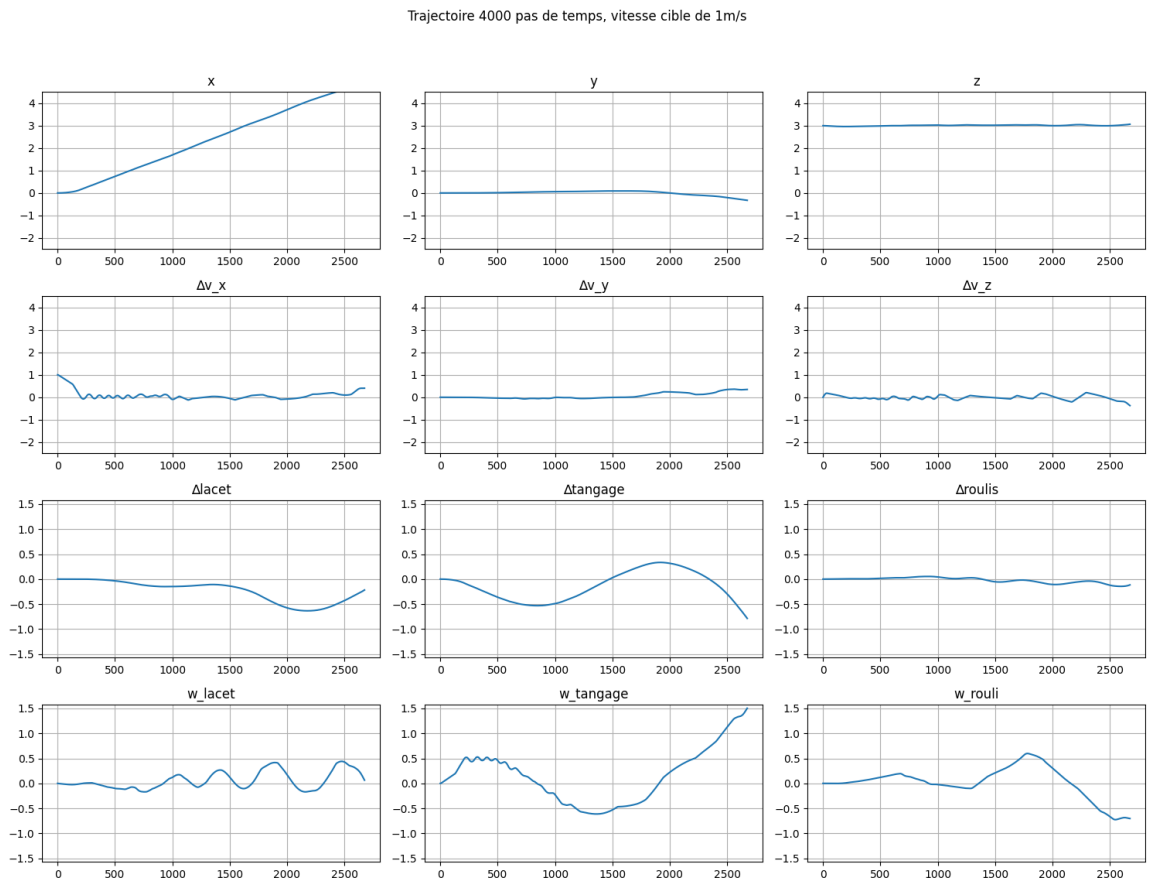

FIGURE 4.8 Visualisation des observations générées par le modèle entraîné : 4000 pas de temps et vitesse cible de 1 m/s

Une des raisons pour laquelle on entraîne notre système à faire cette tâche est de valider que l'apprentissage par renforcement peut utiliser adéquatement les gouvernes à l'arrière et comprendre leur utilité en vol croisière. La figure 4.9 montre que la gouverne de profondeur est bien utilisée pour compenser le moment de tangage.

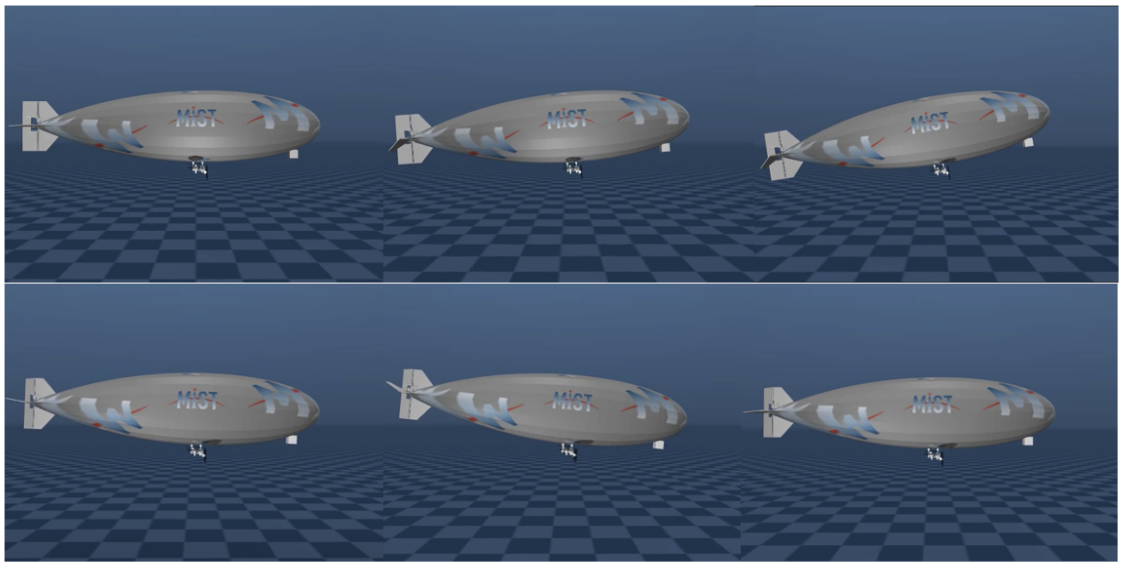

FIGURE 4.9 Mouvements de la gouverne de profondeur du dirigeable selon l'angle de tangage

4.3 Conclusion

Les expériences ont démontré qu'il est possible d'entraîner des modèles de contrôle par apprentissage par renforcement sur notre système. On constate en premier lieu que PPO donne de bons résultats pour l'entraînement. Une analyse plus approfondie des hyperparamètres de l'algorithme serait pertinente pour des tâches plus complexes. On observe également que le contrôle de l'altitude s'avère plus difficile lorsque le dirigeable doit descendre. De plus, le système rencontre des difficultés à fonctionner à des vitesses supérieures à 1 m/s, en raison du moment de tangage induit par les moteurs de propulsion. Il est donc nécessaire de limiter la vitesse à une plage plus basse. Enfin, les deux algorithmes évalués permettent une généralisation satisfaisante en fonction du nombre de pas de temps, mais montrent des limites lorsqu'il s'agit d'atteindre des cibles situées en dehors des plages d'entraînement.

CHAPITRE 5 CONCLUSION

Les travaux présentés dans cette recherche constituent le premier pas vers la recherche sur le contrôle des dirigeables pour le laboratoire. Elle a été créée comme outil pour les travaux futurs. On y présente la conception d'un robot dirigeable pour utilisation intérieure et un simulateur incluant les effets aérodynamiques. Des algorithmes d'apprentissage par renforcement sont ensuite testés sur deux tâches différentes. Ce robot sera utilisé dans des travaux futurs pour appliquer des algorithmes d'apprentissage par renforcement sur un dirigeable dans le monde réel.

5.1 Synthèse des travaux

Le résultat des travaux est le robot dirigeable RAFALE présenté dans le chapitre 3 et le simulateur l'accompagnant. C'est un dirigeable de petite taille pour utilisation à l'intérieur qui peut à la fois voler en mode croisière et en mode stationnaire. Il possède quatre moteurs et deux gouvernes permettant de le contrôler. Un simulateur a été développé à l'aide de MuJoCo afin de permettre d'entraîner des politiques d'apprentissage par renforcement pour le contrôle du système. Un modèle aérodynamique a été développé et intégré avec le moteur physique de MuJoCo pour simuler adéquatement les mouvements du robot dirigeable.

5.2 Limitations de la solution proposée

Le simulateur du dirigeable, bien que basé sur des données de dirigeables semblables, n'a pas été validé avec le système réel. L'acquisition de données de vol du prototype permettrait de valider adéquatement le simulateur. De plus, différentes hypothèses ont été utilisées pour la modélisation du dirigeable, comme la rigidité du dirigeable. Ces hypothèses, bien que courantes dans le milieu, simplifient des dynamiques qui sont en réalité plus complexes et le modèle peut donc différer du mouvement du dirigeable dans le monde réel. De plus, si les résultats au chapitre 4 montrent que l'apprentissage par renforcement est une avenue pour le contrôle du système, les algorithmes ont seulement été testés en simulation, et appliquer ces modèles au système réel n'est pas trivial.

5.3 Améliorations futures

Les travaux présentés dans ce mémoire sont la base des travaux qui seront menés dans le cadre du projet de recherche ENVIA Québec. La prochaine étape sera certainement de tester sur le robot dirigeable les algorithmes entraînés, et peut-être même d'entraîner des algorithmes d'apprentissage par renforcement directement sur le système réel. En effet, comme le dirigeable est fait d'une enveloppe souple, le risque d'accidents graves est grandement réduit, ce qui fait de ce prototype une plateforme de choix pour apprendre directement dans le monde réel. L'étape suivante serait de développer des algorithmes d'apprentissage par renforcement permettant d'obtenir des garanties de sécurité sur le vol du dirigeable. Une des avenues explorées par le projet est l'utilisation de *Control Barrier Functions* (CBF), qui permet d'offrir une certaine garantie sur les actions du système [49]. Pour se faire, un module de simulation de vent pourrait être intégré au simulateur. Également, comme mentionné dans les limitations, le simulateur pourrait prochainement être validé à l'aide de données de vol qui permettraient d'affiner le modèle.

RÉFÉRENCES

[1] United Nations Environment Programme, “Emissions Gap Report 2024 : No more hot air ... please ! With a massive gap between rhetoric and reality, countries draft new climate commitments,” 2024.

[2] B. E. Prentice, “Intermodal Competition : Cargo Airships versus Long-Haul Trucking for Perishable Commodities,” *Journal of Transportation Technologies*, vol. 14, n°. 02, p. 195–211, 2024.

[3] A. Moutinho *et al.*, “Airship robust path-tracking : A tutorial on airship modelling and gain-scheduling control design,” *Control Engineering Practice*, vol. 50, p. 22–36, mai 2016.

[4] J. Azinheira *et al.*, “Hexa-Propeller Airship for Environmental Surveillance and Monitoring in Amazon Rainforest,” vol. 11, n°. 4, p. 249, 2024. [En ligne]. Disponible : https://www.mdpi.com/2226-4310/11/4/249

[5] S. B. V. Gomes et J. G. Ramos, “Airship dynamic modeling for autonomous operation,” 1998.

[6] L. M. Nicolai, G. Carichner et L. M. Nicolai, *Fundamentals of Aircraft and Airship Design*, ser. AIAA Educational Series. Reston, VA : American Institute of Aeronautics and Astronautics, 2010.

[7] H. Lamb, “The Inertia Coefficients of an Ellipsoid Moving in Fluid,” *Advisory Committee for Aeronautics, Reports and Memoranda, No. 623*, oct. 1918.

[8] E. C. de Paiva, S. B. V. Gomes et M. Bergerman, “A control system development environment for AURORA’s semi-autonomous robotic airship.”

[9] J. R. Azinheira *et al.*, “Lateral/directional control for an autonomous, unmanned airship,” vol. 73, n°. 5, 2001.

[10] E. C. de Paiva *et al.*, “Project AURORA : Infrastructure and flight control experiments for a robotic airship,” *Journal of Field Robotics*, vol. 23, n°. 3/4, p. 201–222, 2006.

[11] A. Rottmann *et al.*, “Towards an Experimental Autonomous Blimp Platform.”

[12] O. Daskiran, B. Huff et A. Dogan, “Low Speed Airship Control using Reinforcement Learning and Expert Demonstrations.”

[13] E. Price *et al.*, “Simulation and Control of Deformable Autonomous Airships in Turbulent Wind,” vol. 412, p. 608–626. [En ligne]. Disponible : http://arxiv.org/abs/2012.15684

[14] M. Andronie, “Nova Science Publishers, Inc.”

[15] A. S. Marton *et al.*, “Filtering and Estimation of State and Wind Disturbances Aiming Airship Control and Guidance.”

[16] T. Liesk, M. Nahon et B. Boulet, “Design and experimental validation of a controller suite for an autonomous, finless airship,” dans *2012 American Control Conference (ACC)*. IEEE, p. 2491–2496. [En ligne]. Disponible : http://ieeexplore.ieee.org/document/6315218/

[17] G. Aman, “Indoor Blimp Control.”

[18] P. González *et al.*, “Developing a Low-Cost Autonomous Indoor Blimp,” vol. 3, n°. 1, 2008.

[19] J. E. S. Gordoniz, N. Reeves et D. St-Onge, “Modular Foldable Airship Concept for Subterranean Exploration,” 2021.

[20] Y. Li, M. Nahon et I. Sharf, “Airship dynamics modeling : A literature review,” *Progress in Aerospace Sciences*, vol. 47, n°. 3, p. 217–239, avr. 2011.

[21] J. L. Mendonça Junior *et al.*, “Airship Aerodynamic Coefficients Estimation Based on Computational Method for Preliminary Design,” dans *AIAA Aviation 2019 Forum*. Dallas, Texas : American Institute of Aeronautics and Astronautics, juin 2019.

[22] S. P. Jones et J. D. DeLaurier, “Aerodynamic estimation techniques for aerostats and airships,” *J. AIRCRAFT*, 1983.

[23] G. A. Khoury, édit., *Airship Technology*, 2e éd., ser. Cambridge Aerospace Series. Cambridge ; New York : Cambridge University Press, 2012, n°. 10.

[24] T. Lutz *et al.*, “Aerodynamic Investigations on Inclined Airship Bodies.”

[25] A. Dumas *et al.*, “CFD Analysis and Optimization of a Variable Shape Airship,” 2012.

[26] M. Carrión *et al.*, “Study of Hybrid Air Vehicle Stability Using Computational Fluid Dynamics.”

[27] A. Marton *et al.*, “Hybrid model-based and data-driven wind velocity estimator for an autonomous robotic airship,” *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 42, 03 2020.

[28] Y. T. Liu *et al.*, “Deep Residual Reinforcement Learning based Autonomous Blimp Control.”

[29] Y. Zuo, Y. T. Liu et A. Ahmad, “Autonomous Blimp Control via H-infinity Robust Deep Residual Reinforcement Learning,” mars 2023.

[30] E. Todorov, T. Erez et Y. Tassa, “Mujoco : A physics engine for model-based control,” dans *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012, p. 5026–5033.

[31] T. Erez, Y. Tassa et E. Todorov, “Simulation tools for model-based robotics : Comparison of Bullet, Havok, MuJoCo, ODE and PhysX.”

[32] A. Raffin *et al.*, “Stable-baselines3 : Reliable reinforcement learning implementations,” *Journal of Machine Learning Research*, vol. 22, n°. 268, p. 1–8, 2021. [En ligne]. Disponible : http://jmlr.org/papers/v22/20-1364.html

[33] M. Towers *et al.*, “Gymnasium : A Standard Interface for Reinforcement Learning Environments,” nov. 2024.

[34] J. Collins *et al.*, “A Review of Physics Simulators for Robotic Applications,” vol. 9, 2021.

[35] S. Tunyasuvunakool *et al.*, “Dm_control : Software and tasks for continuous control,” 2020.

[36] C. Tang *et al.*, “Deep Reinforcement Learning for Robotics : A Survey of Real-World Successes,” *Annual Review of Control, Robotics, and Autonomous Systems*, nov. 2024.

[37] E. Kaufmann *et al.*, “Champion-level drone racing using deep reinforcement learning,” *Nature*, vol. 620, n°. 7976, p. 982–987, août 2023.

[38] J. Eschmann, D. Albani et G. Loianno, “Learning to Fly in Seconds,” *IEEE Robotics and Automation Letters*, vol. 9, n°. 7, p. 6336–6343, juill. 2024.

[39] G. B. Margolis et P. Agrawal, “Walk These Ways : Tuning Robot Control for Generalization with Multiplicity of Behavior.”

[40] J. Ko *et al.*, “Gaussian Processes and Reinforcement Learning for Identification and Control of an Autonomous Blimp.”

[41] A. Rottmann *et al.*, “Autonomous Blimp Control Using Model-Free Reinforcement Learning in a Continuous State and Action Space.”

[42] H. Gou *et al.*, “Path Following Control for Underactuated Airships with Magnitude and Rate Saturation,” 2020.

[43] C. Nie, Z. Zheng et M. Zhu, “Three-Dimensional Path-Following Control of a Robotic Airship with Reinforcement Learning,” *International Journal of Aerospace Engineering*.

[44] C. Louis, T. Ilyass et S.-O. David, “CAVERNAUTE : A design and manufacturing pipeline of a rigid but foldable indoor airship aerial system for cave exploration,” sept. 2024.

[45] Windreiter Team, “Main page : The silent_runner,” 2020. [En ligne]. Disponible : http://silent-runner.net/

[46] S. B. V. Gomes, “An Investigation of the Flight of Airships with Application to the YEZ-2A,” Thèse de doctorat, Cranfield Institute of Technology, oct. 1990.

[47] Y. T. Liu *et al.*, “Autonomous Blimp Control using Deep Reinforcement Learning,” sept. 2021.

[48] J. Schulman *et al.*, “Proximal Policy Optimization Algorithms,” août 2017.

[49] A. D. Ames *et al.*, “Control Barrier Functions : Theory and Applications,” mars 2019.

ANNEXE A DÉFINITIONS DES VARIABLES DU MODÈLE DYNAMIQUE DU DIRIGEABLE

Tableau A.1 Tableau des définitions des variables utilisées dans le modèle dynamique du dirigeable.

Variable	**Définition**
M	Matrice de masse, contenant l'inertie réelle et virtuelle.
$x = [u, v, w, p, q, r]$	Vecteur d'état : vitesses linéaires et angulaires dans le référentiel du corps.
x_a	Vecteur d'état : vitesses linéaires et angulaires dans le référentiel du vent relatif.
F_{d}	Vecteur dynamique des forces et moments dus aux effets de Coriolis et centrifuges.
F_{a}	Vecteur des forces et moments aérodynamiques.
F_{p}	Vecteur des forces et moments de propulsion.
G	Vecteur des forces et moments dus à la flottabilité et la gravité.
m	Masse du robot dirigeable.
$X_{\dot{u}}$, $Y_{\dot{v}}$, $Y_{\dot{v}}$	Termes de masse virtuelle.
m_x, m_y, m_z	Masses en selon l'axe contenant la masse du système et la masse vituelle.
OC	distance entre le centre de gravité et le centre de volume.
a_x, a_y, a_z	Coordonnées du centre de gravité par rapport au centre de volume.
I_x, I_y, I_z	Termes d'inertie par rapport aux axes OX, OY et OZ.
$L_{\dot{p}}$, $M_{\dot{q}}$, $N_{\dot{r}}$	Termes d'inertie virtuelle par rapport aux axes OX, OY, et OZ.
J_x, J_y, J_z	Termes d'inertie combinant l'intertie du système et l'inertie virtuelle.
T_{ds}, T_{dp}, T_{ls}, T_{lp}	Poussées des moteurs de propulsion et de levage à bâbord et tribord.
d_x, d_y, d_z	Distances entre le centre de volume et l'hélice concerné
V_t	Vitesse vraie.
V_e	Vitesse équivalente.
α	Angle d'incidence.
β	Angle de dérapage
C_D, C_Y, C_L, C_l, C_m, C_n	Coefficients aérodynamiques (traînée, portance, force latérale, moments).
A_X, A_Y, A_Z, A_L, A_M, A_N	Forces aérodynamiques (traînée, portance, force latérale, moments).
F_{ref}, M_{ref}	Force et moment de référence
S_{ref}, L_{ref}	Surface et longueur de référence, respectivement Volume$^{2/3}$ et Volume$^{1/3}$
ρ	Densité de l'air.
q^*, r^*	Taux de giration non dimensionnels.
C_{Yr}, C_{nr}, C_{Zq} et C_{mq}	Coefficient dynamiques pour amortissement.
S_1, S_2	Matrices de transformation.