

# Apprentissage de dynamiques latentes : stabilisation par structure hamiltonienne

F. Lioni<sup>1</sup>, N. Gutowski<sup>2</sup>, S. Aubin<sup>3</sup>, P. Martinet<sup>1</sup>

<sup>1</sup> Université de Côte d'Azur, INRIA

<sup>2</sup> Université d'Angers, LERIA

<sup>3</sup> Direction Générale de l'Armement - Techniques Terrestres

## Résumé

*L'apprentissage de modèles physiques à partir de données est un problème de longue date en identification de systèmes et en apprentissage automatique. Garantir une généralisation fiable sur le domaine d'intérêt est crucial pour le déploiement sécurisé des modèles appris dans les applications de robotique et d'ingénierie. Ce défi est accentué dans les contextes impliquant des horizons de prédiction à long terme et des variables d'état partiellement observées, où les représentations latentes non contraintes ont tendance à accumuler des erreurs au fil du temps. Dans ce contexte, nous proposons une architecture neuronale hamiltonienne pour l'apprentissage de systèmes dynamiques partiellement observés. En exploitant un biais inductif de conservation de l'énergie, notre approche combine un encodeur de contexte avec une formulation hamiltonienne, permettant l'inférence de l'état latent à partir d'observations partielles tout en préservant la structure physique sous-jacente. Le modèle est entraîné sur des fenêtres d'observation courtes, bruitées et partielles, puis évalué sur des prédictions à long terme afin de mesurer la stabilité et la robustesse de sa généralisation. Les expériences menées sur des variantes non amorties et conservatives des systèmes de Duffing, du pendule, de Hénon-Heiles et de Kepler démontrent une amélioration significative de la stabilité à long terme et de la précision de prédiction par rapport aux modèles de référence non structurés.*

## 1 Introduction

L'apprentissage de modèles dynamiques à partir de données est au cœur de la modélisation, du contrôle pour la robotique et l'ingénierie [10]. Pour les applications critiques en matière de sécurité, une précision sur de courtes échelles de temps ne suffit pas, les modèles doivent fournir des prédictions fiables à long terme, là où de petites erreurs de modélisation peuvent s'accumuler, entraînant une dérive catastrophique. Cette sensibilité a déplacé l'attention des régresseurs traditionnels en temps discret vers des formulations en temps continu [9] pour leurs propriétés de robustesse accrues. En assimilant les mises à jour résiduelles à des schémas de discrétisation d'équations différentielles, les Neural ODEs se sont imposées comme un cadre robuste pour l'apprentissage de champs de vecteurs paramétrés [1].

Bien que ces modèles offrent une meilleure cohérence temporelle, ils restent fondamentalement de nature "boîte noire", car ils n'imposent pas explicitement d'invariants physiques ou de quantités conservées.

Pour combler cette lacune, des architectures informées par la physique telles que les Réseaux de Neurones Hamiltoniens (HNNs) [4] et les Réseaux de Neurones Lagrangiens (LNNs) [11] ont été proposées. En intégrant des biais inductifs basés sur l'énergie directement dans l'architecture du réseau, ces modèles préservent la structure géométrique sous-jacente du flot, garantissant la stabilité à long terme et la conservation de l'énergie. Cependant, leur utilité pratique est souvent limitée par une exigence stricte : ils supposent généralement l'accès à l'état canonique complet du système (par exemple, les positions et les moments). Dans les contextes d'ingénierie réels, les capteurs ne fournissent que des observations partielles, rendant ces approches structurées inapplicables sans une méthode de reconstruction des dimensions manquantes de l'espace des phases.

Dans le régime d'observation partielle, l'état du système doit être inféré par des représentations latentes. Pourtant, lorsque ces espaces latents ne sont pas contraints, le modèle est libre d'apprendre des transitions mathématiquement pratiques mais physiquement impossibles. Ce manque d'ancrage structurel provoque une dérive de la dynamique latente, menant à des trajectoires qui violent des lois fondamentales, comme la conservation de l'énergie [8]. Dans les tâches à long horizon, des petites incohérences dans le processus d'inférence latente s'amplifient avec le temps, entraînant des comportements instables, dissipatifs ou physiquement peu plausibles qui compromettent la fiabilité du modèle appris.

Pour remédier à ces limitations, nous proposons un cadre qui impose une structure hamiltonienne dans un espace latent appris. Plutôt que de traiter l'état latent comme un vecteur de caractéristiques arbitraire, nous le contraignons à évoluer comme un ensemble de coordonnées structurées régies par un flot hamiltonien, fournissant une représentation de la dynamique sous-jacente respectueuse de la géométrie symplectique. Nos principales contributions sont les suivantes :

- Nous proposons un cadre de modélisation dynamique latente qui impose une structure hamiltonienne dans un espace latent appris, utilisant des contraintes géométriques comme biais inductif pour

favoriser des flots conservatifs et une stabilité à long terme sous observabilité partielle.

- Nous introduisons un encodeur basé sur le contexte qui infère les conditions initiales latentes à partir de fenêtres d’observation courtes et bruitées, permettant une reconstruction efficace de l’espace des phases sans accès aux états canoniques complets.
- Nous fournissons une analyse théorique et empirique (App. A) caractérisant comment la longueur du contexte, le bruit et la discrétisation induisent conjointement une incertitude d’initialisation irréductible, et comment les projections symplectiques préservant la structure atténuent l’amplification de cette incertitude au fil du temps.
- À travers des expériences approfondies sur des systèmes dynamiques non linéaires, nous démontrons que la structure hamiltonienne réduit systématiquement l’accumulation d’erreurs à long terme et la variabilité d’un entraînement à l’autre par rapport aux modèles de référence Neural ODE latents non contraints.

## 1.1 Travaux antérieurs

La prédiction précise à long terme demeure un défi persistant pour les modèles de systèmes dynamiques appris à partir de données. Même lorsque les erreurs à un pas de temps sont faibles, la répétition des prédictions peut amplifier les incohérences numériques et de modélisation, entraînant des trajectoires instables et des violations d’invariants physiques. Les travaux antérieurs ont abordé ce problème selon trois axes complémentaires (i) les formulations en temps continu qui améliorent la cohérence temporelle, (ii) les biais inductifs préservant la structure pour encoder la physique connue, et (iii) les mécanismes d’inférence d’états latents pour les systèmes partiellement observés. Nous passons en revue ces axes et mettons en évidence une lacune à leur intersection : garantir la stabilité des prédictions à long terme sous observabilité partielle, tout en préservant la structure hamiltonienne.

**Neural ODEs et prédictions longues.** Les méthodes basées sur les *Neural Ordinary Differential Equations* (Neural ODEs) [1] apprennent des dynamiques en temps continu en paramétrant un champ de vecteurs, lequel est intégré numériquement pour correspondre aux trajectoires observées. Au-delà de l’objectif d’entraînement standard à un pas, les méthodes d’entraînement multi-pas montrent une robustesse accrue au bruit et aux erreurs cumulatives lorsque les modèles sont optimisés sur des trajectoires complètes [9]. L’entraînement multi-pas favorise l’apprentissage de dynamiques stables par rapport à la prédiction étape par étape. Cependant, les Neural ODEs restent largement non contraintes : le champ de vecteurs appris n’est pas tenu de strictement satisfaire aux lois de conservation ou aux propriétés géométriques du système sous-jacent. Par conséquent, les projections à long terme peuvent encore dériver, en particulier sous observabilité partielle ou en présence de bruit d’observation, là où l’état inféré peut évoluer de manière physiquement incohérente.

**Dynamiques préservant la structure.** Une approche complémentaire consiste à incorporer des biais inductifs qui préservent la structure physique connue. Les *Hamiltonian Neural Networks* (HNN) [4] et les *Lagrangian Neural Networks* (LNN) [11, 2] paramètrent une fonction d’énergie (ou d’action) dont ils dérivent la dynamique via les équations de Hamilton ou d’Euler-Lagrange. Ce formalisme favorise le respect des invariants physiques et garantit une meilleure stabilité à long terme. D’autres approches s’appuient sur des fonctions de Lyapunov [12] pour contraindre la stabilité, visant ainsi à contrôler la dérive de l’énergie et à préserver la cohérence des prédictions multi-pas sur le long terme. En pratique, cependant, beaucoup de ces méthodes supposent l’accès aux coordonnées canoniques (par ex.,  $(q, p)$ ) ou à des mesures d’état suffisamment riches, ce qui limite leur applicabilité directe aux systèmes réels où les capteurs ne fournissent que des observations partielles.

**Apprentissage sous observabilité partielle.** Lorsque l’état complet n’est pas observé, les modèles introduisent typiquement un état latent inféré à partir de séquences d’observations à l’aide d’encodeurs ou de modèles récurrents [5, 7]. Dans ce contexte, les Neural ODEs latents [13] étendent l’apprentissage en temps continu à ce cadre en couplant un réseau d’inférence avec un modèle de transition latent, permettant l’apprentissage de la dynamique à partir de mesures irrégulières et incomplètes. Cependant, les transitions latentes sont non contraintes, ainsi, rien n’empêche les dynamiques inférées de violer les lois de conservation, et de petites erreurs d’inférence peuvent s’accumuler en une dérive substantielle à long horizon. Les approches préservant la structure sous observabilité partielle reposent souvent sur l’apprentissage de plongements latents à haute dimension, transformant les observations brutes vers des formes canoniques sans supposer l’accès préalable aux coordonnées physiques du système [14], mais elles ne traitent pas directement la stabilité à long terme des projections latentes lorsque l’inférence est effectuée à partir de fenêtres courtes et bruitées.

Notre travail se situe à cette intersection : nous imposons une structure hamiltonienne directement au sein des dynamiques latentes inférées. Cette approche est couplée à une reconstruction de l’espace des phases exploitant des plongements à retard (*delay embeddings*) appris via un encodeur de type *Gated Recurrent Unit* (GRU).

## 2 Méthode

Afin d’aborder le problème de l’apprentissage de dynamiques stables à long terme à partir d’observations partielles et bruitées, notre approche repose sur trois piliers : (i) la reconstruction de l’espace des phases via un encodeur de contexte, (ii) une formulation hamiltonienne des dynamiques latentes, et (iii) une intégration numérique préservant la structure et robuste au bruit d’observation.

**Reconstruction de l’espace des phases.** Nous considérons un système déterministe où l’état complet  $x(t) \in \mathbb{R}^{d_x}$  est régi par  $\dot{x}(t) = F(x(t))$ . Nous n’avons accès qu’à une observation partielle  $y(t) \in \mathbb{R}^{d_y}$ , souvent corrompue par du

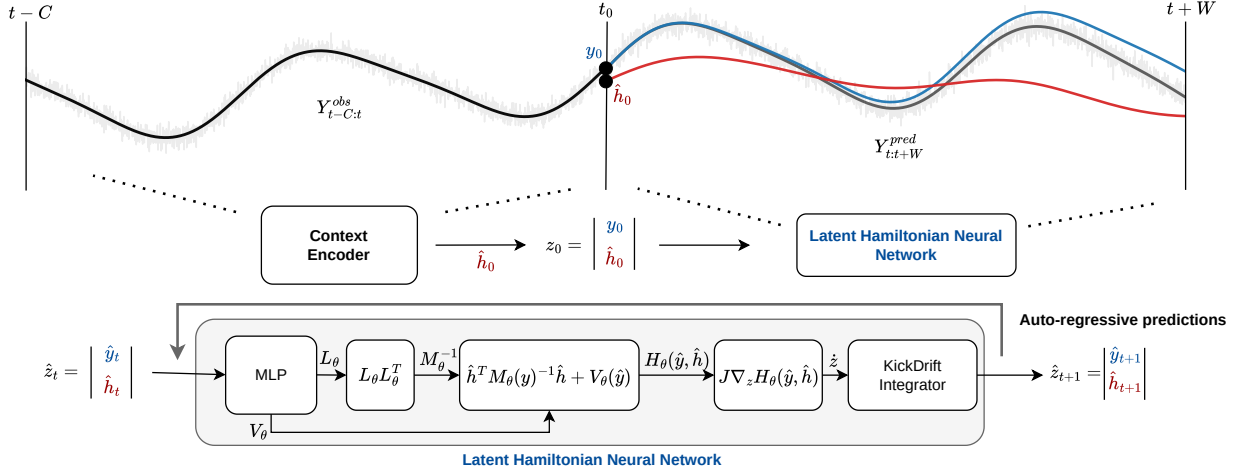


FIGURE 1 – Aperçu du cadre de prévision Hamiltonien latent proposé. Une fenêtre d’observations passées de longueur  $C$  est encodée en un état latent initial  $h_0$  par un encodeur de contexte. Le Réseau de Neurons Hamiltonien latent modélise la dynamique latente en temps continu, laquelle est intégrée numériquement à l’aide d’un intégrateur symplectique de type « kick–drift » de premier ordre pour produire les trajectoires latentes futures sur un horizon  $W$ . Les prédictions observables  $\hat{y}_{t:t+W}$  sont obtenues via un processus de décodage auto-régressif dans l’espace des observations, tandis que les variables latentes évoluent sous une structure hamiltonienne explicite qui capture les propriétés dynamiques sous-jacentes du système.

bruit :  $y^{\text{obs}}(t) = y(t) + \varepsilon(t)$ . Nous utilisons un encodeur de contexte  $E_\phi$ , implémenté sous forme de réseau de neurones récurrent de type GRU, pour projeter un historique de  $C$  observations  $Y_{t-C:t}^{\text{obs}}$  vers une prédiction d’état latent initial  $\hat{h}_0 \in \mathbb{R}^{d_h}$ . Nous formons ensuite l’état augmenté  $z_0 = [y_0, \hat{h}_0]^\top$  en concaténant la prédiction latente initiale avec une observation initiale. Dans nos expériences, nous imposons  $d_h = d_y$  afin de faciliter le couplage symplectique.

**Dynamique Hamiltonienne latente riemannienne.** Pour garantir la stabilité à long terme, nous contraignons l’évolution de  $z_t$  à un flot hamiltonien. Nous paramétrons une fonction hamiltonienne latente  $H_\theta(y, h)$  à l’aide d’un réseau de neurones structuré :

$$H_\theta(y, h) = \frac{1}{2} h^\top M_\theta(y)^{-1} h + V_\theta(y), \quad (1)$$

où  $V_\theta$  représente l’énergie potentielle et  $M_\theta(y)$  est une matrice de masse apprise dépendant de la position. Cette formulation riemannienne permet au modèle d’apprendre des propriétés inertielles dépendantes de la position et généralise les énergies cinétiques euclidiennes classiques. Pour garantir que  $M_\theta(y)$  soit toujours définie positive symétrique (SPD), nous paramétrons son inverse via un facteur de Cholesky  $L_\theta(y)$  :

$$M_\theta(y)^{-1} = L_\theta(y)L_\theta(y)^\top, \quad (2)$$

où  $L_\theta$  est une matrice triangulaire inférieure avec une diagonale strictement positive imposée par une transformation softplus [3]. Cette formulation riemannienne permet au modèle d’apprendre des structures cinétiques complexes et non linéaires dans l’espace latent. Le champ de vecteurs est ensuite dérivé du crochet de Poisson :

$$\dot{z} = J\nabla_z H_\theta(z), \quad J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}. \quad (3)$$

**Intégration préservant la structure.** Nous utilisons un intégrateur symplectique de premier ordre (kick–drift / Euler symplectique) [6] pour projeter la dynamique hamiltonienne apprise. Pour un pas de temps  $\Delta t$ , la mise à jour est :

$$h_{k+1} = h_k - \Delta t \nabla_y H_\theta(y_k, h_k) \quad (4)$$

$$y_{k+1} = y_k + \Delta t \nabla_h H_\theta(y_k, h_{k+1}) \quad (5)$$

Ce choix préserve la forme symplectique au niveau discret, ce qui empêche la dissipation artificielle ou l’injection d’énergie pouvant survenir avec des solveurs non symplectiques, et réduit considérablement la dérive à long terme. Il est également efficace sur le plan computationnel : chaque étape ne nécessite que deux évaluations de gradient et s’avère moins coûteuse que les schémas symplectiques de second ordre courants (ex. Störmer–Verlet/leapfrog), ce qui le rend bien adapté à l’entraînement et aux projections longues.

**Objectif d’entraînement par fenêtre.** Le modèle  $(\theta, \phi)$  est entraîné de bout en bout pour minimiser l’erreur quadratique moyenne sur une fenêtre de prédiction  $W$  :

$$\mathcal{L}(\theta, \phi) = \mathbb{E} \left[ \frac{1}{W} \sum_{k=0}^{W-1} \|\hat{y}_{t+k} - y_{t+k}^{\text{obs}}\|_2^2 \right], \quad (6)$$

où  $\hat{y}$  est la prédiction de l’observation prédite. L’entraînement sur des fenêtres bruitées  $y^{\text{obs}}$  force l’encodeur à extraire une condition initiale latente robuste, tandis que la contrainte hamiltonienne garantit que la prédiction suivante reste physiquement plausible.

**Robustesse avec état latent initial bruité.** Nous fournissons une analyse théorique (Annexe A) expliquant comment l’observabilité partielle, la longueur du contexte et la discrétisation numérique déterminent conjointement la robustesse

à long terme. Notre idée clé est que l’erreur de prédiction à long horizon est dominée par l’amplification de l’incertitude résiduelle dans l’état latent initial inféré, plutôt que par l’erreur locale de prédiction à un pas.

Sous observabilité partielle, le moment latent initial  $\hat{h}_0$  doit être inféré à partir d’une fenêtre de contexte finie et bruitée  $Y_{t-C:t}^{\text{obs}}$ . L’encodeur GRU agit comme un filtre temporel : augmenter la longueur du contexte  $C$  permet d’agréger des mesures redondantes et d’améliorer le rapport signal/bruit de  $\hat{h}_0$ . Cependant, comme les prédictions sont ancrées à la coordonnée initiale observée  $y_0^{\text{obs}}$ , le bruit de mesure induit une contribution irréductible à la covariance initiale  $\Sigma_0 = \text{Cov}([y_0^{\text{obs}}, \hat{h}_0]^\top)$ . Par conséquent, augmenter  $C$  ne réduit la variance de  $\hat{h}_0$  que jusqu’à un plancher d’incertitude induit par le bruit qui ne peut être éliminé par la seule inférence latente.

Une linéarisation au premier ordre de la dynamique latente montre que cette incertitude résiduelle est propagée et amplifiée au cours du temps par l’application répétée de la fonction de transition. En conséquence, même de modestes différences dans la qualité de l’initialisation peuvent entraîner des écarts de performance substantiels à long terme. Crucialement, le choix de la discrétisation numérique affecte directement cette amplification. Alors que les schémas non symplectiques peuvent introduire des artefacts dissipatifs ou expansifs fallacieux qui amplifient artificiellement l’incertitude, les prédictions symplectiques préservant la structure maintiennent la rigidité géométrique du flot hamiltonien et limitent l’amplification numérique.

Ensemble, cette analyse explique les compromis empiriques observés entre la longueur du contexte  $C$ , l’horizon de prédiction  $W$  et la stabilité à long terme, et motive la conception conjointe de la structure hamiltonienne latente et de la discrétisation symplectique.

### 3 Cadre expérimental

Nous focalisons l’évaluation de notre méthode sur ses capacités de prédiction à long horizon, dans un contexte d’observation partielle corrompu par du bruit, pour tester la robustesse aux erreurs cumulatives. Dans ce contexte, les modèles doivent inférer des moments non observés à partir de courtes fenêtres de configurations bruitées et produire des prédictions stables sur de longs horizons. Nous rapportons les performances en utilisant la RMSE à long horizon sur les données de test sur 3 graines (seeds) et analysons l’impact de la longueur du contexte  $C$  et de l’horizon de prédiction  $W$ .

**Données, observabilité partielle et bruit.** Les expériences sont menées sur quatre systèmes dynamiques non linéaires : Duffing, le Pendule, Hénon–Heiles (noté *Henon*), et Kepler. Duffing et le Pendule sont des systèmes 2D avec des états  $x = (q, \dot{q})$  et  $x = (\theta, \omega)$ , respectivement, tandis que Hénon–Heiles et Kepler sont des systèmes hamiltoniens 4D avec des coordonnées canoniques  $x = (q_1, q_2, p_1, p_2)$ . Toutes les trajectoires sont générées par simulation numérique avec un pas de temps fixe  $\Delta t = 0.02$  sur  $T = 500$  pas de temps. Pour chaque système, nous générons 256 trajectoires d’en-

traînement, 64 de validation et 64 de test, chacune à partir de conditions initiales aléatoires. Nous considérons à la fois des observations propres et des observations bruitées, dans le cadre bruité, un bruit gaussien i.i.d. d’écart-type  $\sigma = 0.02$  est ajouté aux variables observées pour tous les systèmes. Le problème d’apprentissage est partiellement observé : le modèle ne reçoit que les variables de configuration (par ex.,  $q$  pour Duffing/Pendule,  $(q_1, q_2)$  pour Hénon–Heiles et Kepler), tandis que les variables de moment ne sont pas observées et doivent être inférées. Les équations dynamiques complètes et les distributions des conditions initiales utilisées pour la génération des données sont fournies en App. C.

**Pourquoi observer uniquement  $q$  ?** Nous choisissons un modèle d’observation portant uniquement sur la position (configuration seule) car il constitue le régime d’inférence le plus difficile et fournit donc un test de résistance significatif pour la robustesse à long horizon. Récupérer les positions à partir des vitesses/moments est une opération de lissage (intégration), alors que récupérer les vitesses/moments à partir des positions nécessite une différenciation, laquelle est mal conditionnée et amplifie fortement le bruit de mesure. Ainsi, observer  $q$  force le modèle à effectuer une estimation de moment implicite et robuste au bruit à partir de fenêtres courtes, plutôt que de s’appuyer sur une application inverse plus facile. Ce cadre met en évidence si une méthode peut reconstruire de manière stable les composantes manquantes de l’espace des phases et maintenir des prédictions physiquement plausibles sous observabilité partielle.

**Entraînement et évaluation à long horizon.** L’entraînement est effectué sur des fenêtres glissantes. Pour chaque trajectoire, nous extrayons des échantillons composés d’un contexte passé  $y_{t-C:t}$  de longueur  $C$  et d’une fenêtre de prédiction future  $y_{t:t+W}$  de longueur  $W$ , en utilisant un pas (stride) de 10. Lors de l’évaluation, les prédictions du modèle sont obtenues de manière auto-régressive sur un long horizon de  $L = 200$  pas, en se conditionnant sur le contexte observé et en mesurant la RMSE par rapport à l’état réel (ground-truth) non bruité.

Nous comparons trois méthodes : (i) **Latent NODE (E)**, un champ latent non contraint intégré avec Euler explicite, (ii) **Latent HNN (E)**, un modèle à paramétrage hamiltonien utilisant Euler explicite ; et (iii) **Latent HNN (S)**, utilisant un schéma symplectique (kick–drift). Bien que les réseaux puissent compenser partiellement les erreurs de discrétisation, les solveurs non symplectiques échouent à apprendre les lois de conservation car leur Inverse Modified Differential Equations (IMDE) n’est pas hamiltonienne [15]. Les états initiaux sont inférés via un encodeur GRU. Entraînés pendant 150 époques avec Adam ( $10^{-4}$ ) et un écrêtage de gradient (gradient clipping) de 1.0, les meilleurs points de sauvegarde (checkpoints) minimisent la RMSE de validation à long horizon ( $L = 200$ ).

**Sur la complexité des solveurs et les gradients d’entraînement.** Les solveurs d’ordre supérieur offrent des avantages limités car l’entraînement des Neural ODEs approxime une IMDE spécifique au solveur plutôt que l’ODE réelle [15]. Les réseaux à haute capacité absorbent efficacement les er-

reurs de discrétisation, mais cette compensation manque de cohérence physique pour les systèmes hamiltoniens, à moins que l’intégrateur ne soit symplectique. Par souci d’efficacité, nous utilisons une discrétisation à pas fixe et différencions à travers le graphe de calcul discret, en privilégiant la stabilité et la vitesse d’entraînement plutôt que l’ordre formel de discrétisation.

## 4 Résultats

Dans cette section, nous évaluons les performances de notre méthode par rapport aux modèles de référence non contraints. Notre analyse se concentre sur trois aspects clés : la stabilité des trajectoires à long terme, la conservation de l’énergie et la robustesse de la reconstruction de l’espace des phases sous observabilité partielle.

**Résumé des performances à long horizon.** Le Tableau 1 rapporte la RMSE à long horizon pour une longueur de projection (rollout) fixe  $L = 200$  pour quatre systèmes sous observations propres et bruitées ( $\sigma = 0,02$ ). Pour chaque modèle, le couple  $(C, W)$  est sélectionné en minimisant la RMSE moyenne de validation à long horizon sur plusieurs graines (seeds), et nous rapportons la RMSE de test correspondante (moyenne  $\pm$  std). Dans l’ensemble, imposer un paramétrage hamiltonien améliore la stabilité à long terme par rapport à la méthode Latent NODE. L’utilisation d’un intégrateur symplectique renforce encore cet effet en empêchant la dérive induite par l’intégration. Le bénéfice est plus marqué sur les systèmes de Kepler et de Hénon–Heiles, où le modèle HNN (S) permet des réductions d’erreurs importantes et une plus faible variabilité entre les graines dans le cadre bruité, indiquant une meilleure robustesse aux erreurs cumulatives sous observabilité partielle. Sur le Pendule, le modèle HNN (E) est moins performant que Latent NODE, soulignant qu’un modèle hamiltonien, sans une discrétisation préservant cette structure, peut être préjudiciable, le passage au modèle HNN (S) rétablit des gains constants, confirmant l’importance de la cohérence entre le modèle et l’intégrateur.

La Figure 2 apporte une interprétation complémentaire en visualisant l’évolution de la RMSE (erreur quadratique

moyenne) de test sur des prédictions à long terme (ou rollouts) pour chaque système. Le système du pendule met en évidence l’importance de la cohérence de l’intégration numérique et de la contrainte symplectique dans la dynamique latente. Bien que le paramétrage hamiltonien puisse en principe améliorer la stabilité, le modèle HNN (E) ne surpasse pas systématiquement le modèle de référence non structuré. En revanche, le modèle HNN (S) atteint une erreur plus faible et une variance réduite sur une large gamme de tailles de fenêtres, ce qui indique que la contrainte hamiltonienne seule est insuffisante à moins que le schéma d’intégration numérique ne préserve la structure géométrique sous-jacente. Ce comportement s’étend de manière cohérente à des dynamiques plus complexes, soulignant davantage le rôle de contraintes structurelles fortes dans la stabilisation des représentations latentes.

Sur le système de Hénon–Heiles, les modèles latents non structurés présentent des trajectoires incohérentes selon les graines aléatoires, ce qui entraîne une variance élevée et une croissance rapide de l’erreur à mesure que l’horizon de prédiction augmente. Les modèles basés sur le formalisme hamiltonien imposent une contrainte forte sur la dynamique latente qui régularise son évolution, aboutissant à des trajectoires plus cohérentes, une variance réduite et une meilleure précision à long horizon. La variante symplectique renforce encore cet effet en préservant la structure géométrique apprise lors de l’intégration numérique.

Une tendance similaire est observée sur le système de Kepler. Le modèle de référence Latent NODE montre à nouveau des prédictions instables et très variables selon les graines, avec une accumulation d’erreurs sévère à mesure que la longueur du contexte augmente. Les modèles hamiltoniens atténuent considérablement ces effets en imposant une dynamique latente structurée, et le couplage de cette structure avec une intégration symplectique produit les prédictions à long horizon les plus stables et les plus robustes.

Ensemble, ces résultats confirment que des contraintes géométriques fortes, combinées à une intégration numérique cohérente avec la structure, sont essentielles pour réduire la variabilité entre les exécutions et maintenir la précision sur des horizons étendus sous observabilité partielle.

Model	Duffing				Henon			
	Clean		Noisy		Clean		Noisy	
	RMSE ↓	Gain ↑	RMSE ↓	Gain ↑	RMSE ↓	Gain ↑	RMSE ↓	Gain ↑
Latent NODE	0.022 ± 0.000	–	0.035 ± 0.002	–	0.038 ± 0.005	–	0.112 ± 0.034	–
Latent HNN (E)	0.007 ± 0.000	+214	0.029 ± 0.003	+21	0.037 ± 0.006	+3	0.057 ± 0.004	+96
Latent HNN (S)	0.004 ± 0.000	+450	0.029 ± 0.003	+21	0.038 ± 0.003	+0	0.052 ± 0.000	+115
Model	Pendulum				Kepler			
	Clean		Noisy		Clean		Noisy	
	RMSE ↓	Gain ↑	RMSE ↓	Gain ↑	RMSE ↓	Gain ↑	RMSE ↓	Gain ↑
Latent NODE	0.009 ± 0.000	–	0.023 ± 0.001	–	0.308 ± 0.000	–	0.349 ± 0.032	–
Latent HNN (E)	0.043 ± 0.000	–79	0.048 ± 0.002	–52	0.033 ± 0.000	+835	0.052 ± 0.000	+575
Latent HNN (S)	0.003 ± 0.000	+200	0.016 ± 0.001	+44	0.029 ± 0.000	+957	0.052 ± 0.013	+568

TABLE 1 – RMSE de test (moyenne  $\pm$  écart-type sur les graines aléatoires) pour des observations propres et bruitées ( $\sigma = 0,02$ ); pour chaque modèle, le couple  $(C, W)$  est sélectionné sur l’ensemble de validation. Le gain est relatif au modèle Latent NODE (plus il est élevé, mieux c’est). Les résultats détaillés et les balayages complets de  $(C, W)$  sont rapportés en Annexe E.

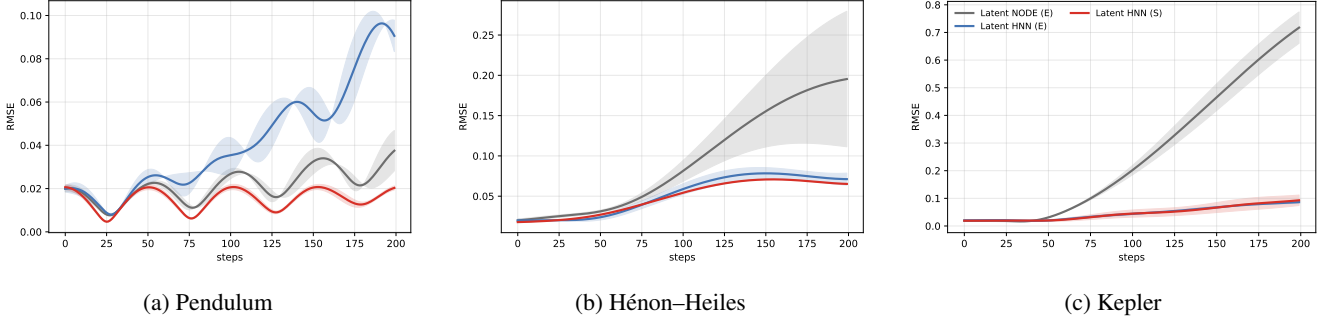


FIGURE 2 – RMSE à long horizon ( $L = 200$ ) sous observations bruitées en fonction des paramètres de fenêtre. Le paramètre complémentaire est sélectionné sur l’ensemble de validation : le meilleur  $C$  pour chaque  $W$  et le meilleur  $W$  pour chaque  $C$ . Voir l’Annexe E pour les détails expérimentaux complets.

**Taille du contexte et horizon d’entraînement.** La Figure 3 rapporte une analyse de sensibilité sur le système Hénon-Heiles sous des observations bruitées, mettant en lumière les rôles complémentaires de la longueur de contexte  $C$  et de l’horizon de déploiement  $W$  dans la robustesse à long horizon. À mesure que  $W$  augmente, la RMSE à long horizon et sa variance se stabilisent progressivement pour tous les modèles. Ce comportement est cohérent avec la décomposition biais-variance des projections à long terme [9] : faire la moyenne de la perte sur des trajectoires plus longues réduit la variance de l’objectif d’entraînement jusqu’à atteindre un régime asymptotique pour les systèmes dynamiques non expansifs. Dans notre configuration, cette condition est satisfaite car la dynamique hamiltonienne, combinée à l’intégration symplectique, préserve le volume de l’espace des phases et évite l’amplification artificielle de l’incertitude (voir App. A). Néanmoins, un plancher de variance subsiste même pour de grandes valeurs de  $W$ , reflétant l’incertitude irréductible induite par la coordonnée initiale observée et bruitée  $y_0^{obs}$ . Puisque les prédictions sont ancrées à cette observation, le bruit de mesure induit une contribution non nulle à la covariance initiale qui ne peut être éliminée en étendant l’horizon de projection (App. A). Au-delà de ce point, augmenter  $W$  ne réduit plus la variance de la prédic-

tion.

Une fois que le signal de supervision est stabilisé par un  $W$  suffisamment grand, l’augmentation de la longueur du contexte  $C$  améliore la robustesse en permettant à l’encodeur de mieux exploiter ce signal d’entraînement. Des fenêtres de contexte plus longues renforcent le filtrage temporel effectué par l’encodeur, menant à une inférence plus cohérente du moment latent initial  $\hat{h}_0$  à travers les différentes graines aléatoires. Empiriquement, cet effet se manifeste principalement par une réduction de la variance de la RMSE à long horizon plutôt que par une diminution supplémentaire de sa moyenne, indiquant une meilleure robustesse de l’inférence latente sous l’effet du bruit. Dans l’ensemble, ces résultats démontrent que la robustesse à long horizon provient de l’interaction entre des prédictions symplectiques stabilisées sur  $W$  et un filtrage latent efficace sur  $C$ , tous deux opérant sous un seuil d’incertitude irréductible fixé par l’observation initiale bruitée.

**Géométrie de l’espace des phases et jauge d’énergie.** L’hamiltonien appris définit une jauge d’énergie latente, c’est-à-dire qu’il est défini à une échelle et un décalage près. Dans ce contexte, la Figure 4 fournit des diagnostics géométriques complémentaires à travers les trajectoires de l’espace des phases latent. Nous soulignons que l’Hamiltonien appris

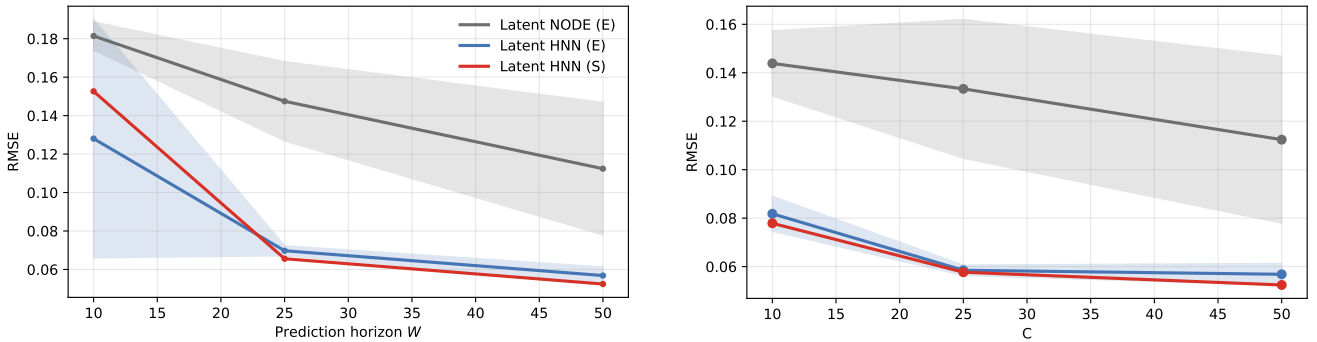


FIGURE 3 – Chaque graphique présente la RMSE de test obtenue en sélectionnant le paramètre de fenêtre complémentaire optimal (meilleur  $C$  pour chaque  $W$  à gauche, et meilleur  $W$  pour chaque  $C$  à droite) sur le système de Hénon-Heiles sous observations bruitées.

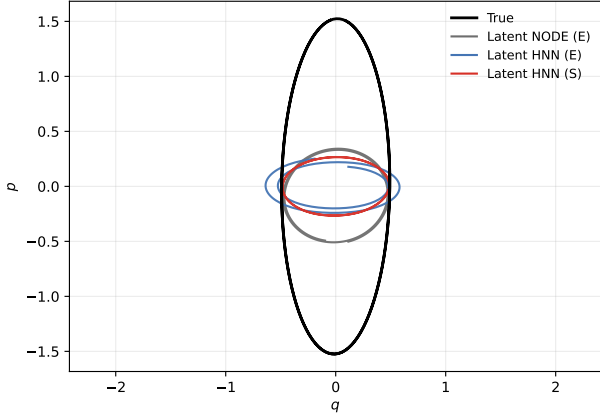


FIGURE 4 – Trajectoires dans l’espace des phases latent pour le système du pendule sous observations bruitées (meilleur  $(C, W)$ ). Le modèle HNN (S) symplectique préserve une structure orbitale cohérente, contrairement aux variantes non symplectiques. L’hamiltonien appris définit une jauge d’énergie latente : l’échelle absolue des orbites n’a pas de signification physique.

$H_\theta$  ne doit pas être interprété comme la véritable énergie physique. Il est défini à un choix de jauge près et sert de coordonnée d’énergie latente plutôt que de quantité physique absolue. Par conséquent, la taille et l’échelle des orbites dans l’espace des phases ne sont pas significatives en soi d’un point de vue physique.

Malgré cette liberté de jauge, les trajectoires dans l’espace des phases révèlent clairement l’effet des contraintes symplectiques. Le modèle HNN (S) symplectique produit des orbites cohérentes et fermées qui préservent la géométrie qualitative de la dynamique, alors que la NODE non contrainte et le modèle Hamiltonien couplé à une discrétisation non symplectique présentent une contraction ou une expansion artificielle des trajectoires. Différents choix de jauge (par exemple, induits par la matrice de masse apprise) peuvent redimensionner les orbites, mais ils n’affectent pas la préservation de la structure symplectique sous-jacente (voir l’Annexe B pour plus de détails).

Cette cohérence géométrique se traduit directement par une réduction de la dérive hamiltonienne. Bien que la dérive d’énergie ne soit pas utilisée comme objectif d’optimisation, elle constitue un diagnostic utile de la cohérence numérique et structurelle : les modèles qui violent la géométrie symplectique présentent une injection ou une dissipation systématique d’énergie, alors que le modèle HNN symplectique maintient une dérive faible et bornée sur de longs horizons.

**Dérive énergétique et robustesse.** La Figure 5 complète l’analyse dans l’espace des phases en présentant la dérive relative de l’Hamiltonien appris  $H_\theta$  sur de longs horizons de prédiction sous observations bruitées, ainsi que la variabilité entre les différentes graines aléatoires. Comme  $H_\theta$  définit une jauge d’énergie latente plutôt que l’énergie physique réelle, la dérive rapportée doit être interprétée comme un diagnostic de cohérence interne et non comme une mesure

de l’erreur d’énergie physique. Sur l’ensemble des systèmes, la contrainte hamiltonien réduit considérablement la variabilité d’une exécution à l’autre, produisant des bandes de confiance plus étroites et une meilleure robustesse face aux erreurs de prédiction cumulatives. Cet effet est particulièrement marqué sur Hénon–Heiles et Kepler, où les prédictions non contraintes ou non symplectiques présentent une dérive croissant rapidement et une forte dispersion selon les graines, tandis que les modèles hamiltoniens produisent un comportement nettement plus cohérent. Le couplage de la structure hamiltonienne avec une discrétisation symplectique borne davantage la dérive sur tout l’horizon, indiquant une réduction de l’injection ou de la dissipation d’énergie induite par le solveur.

Notamment, l’hamiltonien appris peut présenter une dérive plus faible que l’énergie évaluée sur les trajectoires de référence. Ceci n’est pas contradictoire : cette dernière reflète les effets de discrétisation et la métrique de dérive relative choisie, tandis que le premier correspond à un flot latent structurellement contraint qui impose une évolution conservative par construction. Dans l’ensemble, ces diagnostics renforcent l’interprétation géométrique : la structure hamiltonienne limite l’espace des dynamiques latentes admissibles, et l’intégration symplectique préserve cette structure au cours du temps, garantissant ainsi des prédictions à long horizon plus stables et fiables sous observabilité partielle.

**Ablation : Matrice de masse identité vs apprise.** En suivant l’interprétation géométrique ci-dessus, nous isolons maintenant le rôle de la métrique latente induite par l’énergie cinétique. Fixer la matrice de masse à l’identité est un choix naturel et physiquement cohérent pour les systèmes considérés, dont les vrais hamiltoniens possèdent une énergie cinétique séparable avec une masse constante. Cependant, dans notre formulation latente, les coordonnées  $(y, h)$  sont apprises et ne coïncident pas nécessairement avec les variables physiques canoniques. En conséquence, même lorsque le système sous-jacent admet une matrice de masse constante, la géométrie latente effective induite par l’observabilité partielle et les représentations apprises est généralement non euclidienne.

Empiriquement, contraindre  $M_\theta(y) = I$  dégrade significativement les performances à long horizon sur le système de Hénon–Heiles. Comme le montre le Tableau 2, la variante avec une matrice de masse apprise dépendant de l’état atteint une RMSE à long horizon ( $L = 200$ ) nettement plus faible sous observations partielles. Bien que les deux mo-

Mass matrix	RMSE ( $L = 200$ )
Learned $M_\theta(y)$	$0.0420 \pm 0.0025$
Identity $M = I$	$0.0597 \pm 0.0076$

TABLE 2 – Ablation de la matrice de masse latente sur le système de Hénon–Heiles avec observations partielles ( $C = W = 50$ ). Fixer la masse à l’identité dégrade la précision à long horizon malgré la préservation de la structure hamiltonienne.

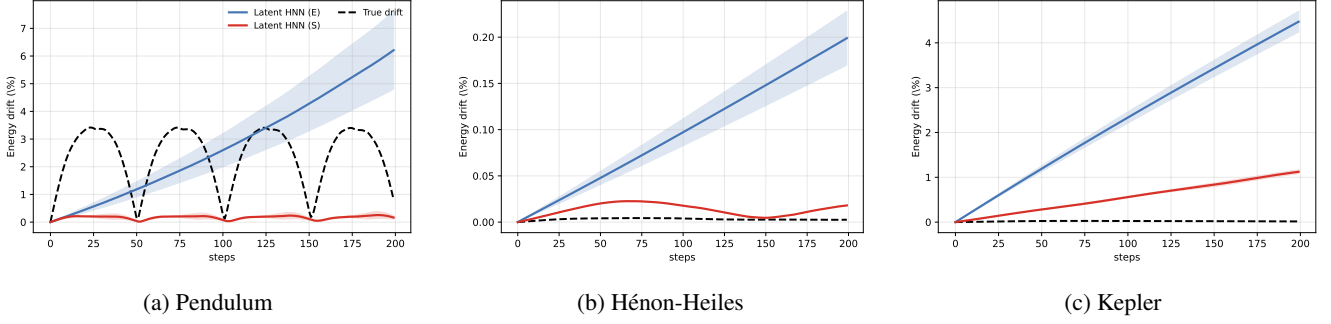


FIGURE 5 – Dérive hamiltonienne relative au cours du temps, exprimée en pourcentage de l’énergie initiale. Les lignes pleines représentent la moyenne sur les graines (seeds) et les zones ombrées indiquent l’écart-type  $\pm 1$ .

dèles restent hamiltoniens et conservent l’énergie en temps continu, fixer la masse limite la puissance expressive de l’hamiltonien latent et restreint la flexibilité du terme cinétique. Géométriquement, cela correspond à imposer une métrique latente globalement euclidienne, ce qui est insuffisant pour capturer la structure courbe de l’espace des phases observée dans les visualisations précédentes.

Cette restriction impacte également l’inférence de l’état initial : en imposant une relation rigide entre les moments latents et la dynamique observée, elle réduit la capacité de l’encodeur à inférer de manière robuste  $\hat{h}_0$  à partir d’observations bruitées et partielles. L’apprentissage d’une matrice de masse dépendant de l’état pallie ces limitations en adaptant la métrique latente à la géométrie de l’espace des phases inféré, ce qui se traduit par une meilleure robustesse et une précision accrue à long horizon.

## 5 Limitations

Notre étude se concentre sur des systèmes hamiltoniens simulés de faible dimension, qui sont bien adaptés pour isoler l’effet de la structure hamiltonienne et de la discrétisation symplectique sous observabilité partielle. Bien que les résultats soulignent systématiquement des gains en termes de robustesse et de stabilité à long horizon, l’extension de cette approche à des systèmes de plus haute dimension et à des modalités d’observation plus complexes (par exemple, des images ou des schémas de détection irréguliers) reste une direction ouverte. Le modèle dynamique proposé cible principalement des régimes approximativement conservatifs et ne capture pas explicitement la dissipation ou les forçages externes. L’extension du cadre à des contextes non conservatifs ou contrôlés, par exemple via des formulations port-hamiltoniennes ou des approches connexes, constitue une avenue naturelle pour de futurs travaux.

Enfin, sous observabilité partielle, l’état latent initial est inféré à partir d’une fenêtre de contexte finie, ce qui induit un plancher d’incertitude irréductible en présence de bruit et peut limiter la précision à long horizon dans des régimes difficiles. Bien que nos résultats montrent que les contraintes géométriques atténuent l’amplification d’erreur, des objectifs d’inférence plus structurés (par exemple, des encodeurs basés sur le filtrage ou le lissage) pourraient encore améliorer la robustesse.

## 6 Conclusion

Nous avons introduit un cadre de modélisation hamiltonien latent pour l’apprentissage de systèmes dynamiques sous observabilité partielle, dans lequel un paramétrage hamiltonien structuré est combiné à une inférence d’état basée sur le contexte et à une intégration temporelle symplectique.

Plutôt que de viser la récupération exacte de l’énergie physique, l’approche proposée exploite la structure hamiltonienne comme un biais inductif géométrique qui contraint l’espace des flots latents admissibles.

À travers des expériences approfondies sur des systèmes non linéaires canoniques, nous avons montré que ce biais géométrique apporte des bénéfices concrets pour la prédiction à long horizon. En particulier, le paramétrage hamiltonien réduit systématiquement la variance entre les initialisations aléatoires, tandis que les projections symplectiques atténuent la dérive induite par le solveur et la distorsion artificielle de l’espace des phases.

Ensemble, ces effets se traduisent par des prédictions plus stables et plus fiables, en particulier dans les régimes où les erreurs cumulatives et l’observabilité partielle prédominent.

Notre analyse souligne que l’hamiltonien appris doit être interprété comme une jauge d’énergie latente : il définit un système de coordonnées conservatif dans lequel la dynamique évolue de manière cohérente, même lorsque l’énergie physique n’est pas directement observable ou est affectée par la discrétisation et le bruit.

De ce point de vue, la dérive énergétique sert de diagnostic de cohérence interne et de robustesse numérique, plutôt que de mesure de fidélité physique.

Dans l’ensemble, ce travail positionne la structure hamiltonienne et l’intégration symplectique comme des outils complémentaires pour contrôler l’amplification des erreurs à long terme dans les modèles dynamiques latents.

En imposant explicitement des contraintes géométriques à la fois au niveau du modèle et du solveur, notre approche offre une voie rigoureuse et pratique vers une prédiction robuste à long horizon sous observabilité partielle, et pose les bases de l’extension de l’apprentissage préservant la structure à des systèmes dynamiques plus complexes, de plus grande dimension et pilotés par les données.

## Références

- [1] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, 2018.
- [2] Miles D. Cranmer, Sam Greydanus, Stephan Hoyer, Peter W. Battaglia, David N. Spergel, and Shirley Ho. Lagrangian neural networks. *CoRR*, abs/2003.04630, 2020.
- [3] Charles Dugas, Yoshua Bengio, François Bélisle, Claude Nadeau, and René Garcia. Incorporating second-order functional knowledge for better option pricing. In *Advances in Neural Information Processing Systems*, pages 451–457, 2000.
- [4] Samuel Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian neural networks. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, 2019.
- [5] Danijar Hafner, Timothy P. Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. In *International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 2555–2565, 2019.
- [6] Ernst Hairer, Christian Lubich, and Gerhard Wanner. *Geometric Numerical Integration*, volume 31 of *Springer Series in Computational Mathematics*. Springer, 2006.
- [7] Rahul G. Krishnan, Uri Shalit, and David Sontag. Structured inference networks for nonlinear state space models. In *AAAI Conference on Artificial Intelligence*, pages 2101–2109, 2017.
- [8] Kookjin Lee and Kevin T. Carlberg. Deep conservation : A latent-dynamics model for exact satisfaction of physical conservation laws. In *AAAI Conference on Artificial Intelligence*, 2019.
- [9] Fabien Lionti, Nicolas Gutowski, Sébastien Aubin, and Philippe Martinet. Bias-variance analysis of multi-step loss functions for dynamical system identification. In *Proceedings of the International Joint Conference on Neural Networks*, 2025.
- [10] Lennart Ljung. *System Identification : Theory for the User*. Prentice Hall PTR, 1999.
- [11] Michael Lutter, Christian Ritter, and Jan Peters. Deep lagrangian networks : Using physics as model prior for deep learning. *CoRR*, abs/1907.04490, 2019.
- [12] Spencer M. Richards, Felix Berkenkamp, and Andreas Krause. The lyapunov neural network : Adaptive stability certification for safe learning of dynamic systems. *CoRR*, abs/1808.00924, 2018.
- [13] Yulia Rubanova, Ricky T. Q. Chen, and David Duvenaud. *Latent ODEs for Irregularly-Sampled Time Series*. Curran Associates, 2019.
- [14] Peter Toth, Danilo Jimenez Rezende, Andrew Jaegle, Sébastien Racanière, Aleksandar Botev, and Irina Higgins. Hamiltonian generative networks. *CoRR*, abs/1909.13789, 2019.
- [15] Aiqing Zhu, Pengzhan Jin, Beibei Zhu, and Yifa Tang. On numerical integration in neural ordinary differential equations. In *International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 27527–27547, 2022.

## A Inférence de l'état initial : l'allongement du contexte améliore le rapport signal/bruit jusqu'à un plancher de variance

Nous fournissons des détails mathématiques supplémentaires étayant la discussion sur la robustesse en §2. La quantité clé gouvernant la robustesse à long horizon sous observabilité partielle est la covariance de l'état latent initialisé,  $\Sigma_0 = \text{Cov}(z_0)$ . Nous expliquons comment la longueur du contexte  $C$  influence  $\Sigma_0$ , pourquoi un plancher de variance irréductible subsiste, et comment des rollouts préservant la structure sur un horizon  $W$  contrôlent la propagation de cette incertitude.

**Configuration (observations partielles et état initial filtré).** Nous considérons un état latent augmenté  $z_t = [y_t, h_t]^\top$  ainsi que des observations bruitées

$$y_t^{\text{obs}} = y_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, R), \quad (7)$$

avec une dynamique latente déterministe  $z_{t+1} = f_\theta(z_t)$  induite par un champ de vecteurs hamiltonien (intégré numériquement comme dans (4)–(5)). Un encodeur de contexte  $E_\phi$  (GRU) projette une fenêtre d'observations  $Y_{t-C:t}^{\text{obs}}$  vers une estimation  $\hat{h}_0 = E_\phi(Y_{t-C:t}^{\text{obs}})$ , et nous initialisons le rollout de prédictions autorégressives par

$$z_0 = [y_0^{\text{obs}}, \hat{h}_0]^\top. \quad (8)$$

D'après (8), la covariance initiale se décompose en une incertitude portant sur le moment latent inféré  $\hat{h}_0$  et une incertitude associée à la coordonnée initiale observée  $y_0^{\text{obs}}$ . Ces deux contributions jouent des rôles fondamentalement distincts dans la robustesse à long horizon.

**Propagation sur  $W$  et rôle de la "rigidité" symplectique.** Dans [9], la fonction objectif associé à un rollout de prédictions autorégressives à  $T$  pas admet la décomposition biais-variance suivante :

$$L(\theta) = \text{Bias}_T^2 + \text{Var}_T \quad (9)$$

avec

$$\text{Var}_T = \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\text{Tr}(\Sigma_t)], \quad \Sigma_{t+1} = J_f(\mu_t) \Sigma_t J_f(\mu_t)^\top, \quad (10)$$

sous une linéarisation au premier ordre autour de la trajectoire moyenne  $\mu_t$ . En déroulant (10), on obtient  $\Sigma_t = A_t \Sigma_0 A_t^\top$  avec  $A_t = \prod_{k=0}^{t-1} J_f(\mu_k)$ , ce qui montre que  $\text{Var}_T$  hérite d'une contribution non nulle provenant de la covariance initiale  $\Sigma_0$ .

Dans le régime hamiltonien considéré, la dynamique en temps continu préserve la structure géométrique et demeure typiquement non contractante et non expansive. Ainsi, l'incertitude initiale représentée par  $\Sigma_0$  persiste à long terme plutôt que de s'atténuer. L'emploi d'un schéma d'intégration symplectique impose une rigidité numérique aux projections sur  $W$  pas : la préservation discrète de la forme symplectique empêche l'apparition d'effets dissipatifs ou instables artificiels, fréquents avec des solveurs non symplectiques. Par suite, accroître  $W$  stabilise l'objectif d'entraînement et améliore le rapport signal/bruit, en effectuant une moyenne sur des trajectoires plus longues et géométriquement consistantes.

**Effet de la longueur du contexte  $C$  sur la qualité d'inférence de  $\hat{h}_0$ .** Nous considérons l'estimation du moment latent initial  $\hat{h}_0$  à partir d'une séquence d'observations bruitées  $Y_{0:C}^{\text{obs}}$  régie par (7). Soit  $\mu_t(h_0)$  la prédiction déterministe de  $y_t$  conditionnelle à  $h_0$ , obtenue par propagation de la dynamique latente. La log-vraisemblance associée à la fenêtre s'écrit

$$\mathcal{L}(h_0) = \frac{1}{2} \sum_{t=0}^C \|y_t^{\text{obs}} - \mu_t(h_0)\|_{R^{-1}}^2 + \text{const}. \quad (11)$$

En notant  $J_t = \partial \mu_t(h_0) / \partial h_0$ , une approximation de type Gauss-Newton, équivalente à l'information de Fisher, donne

$$\mathcal{I}_C(h_0) \approx \sum_{t=0}^C J_t^\top R^{-1} J_t, \quad (12)$$

d'où découle la borne de Cramér-Rao

$$\text{Cov}(\hat{h}_0) \succeq \mathcal{I}_C(h_0)^{-1}. \quad (13)$$

Étant donné que l'augmentation de  $C$  ajoute des termes définis positifs (semi-définis positifs) dans (12), la matrice d'information  $\mathcal{I}_C$  est non décroissante en  $C$ . Lorsqu'elle est de rang plein, la borne (13) décroît donc avec  $C$ . Ce résultat formalise le rôle de l'encodeur de contexte comme filtre temporel : des fenêtres plus longues fournissent des mesures redondantes qui améliorent le rapport signal/bruit de  $\hat{h}_0$ . Cependant, l'efficacité de ce filtrage dépend de manière cruciale de la qualité du signal de supervision utilisé durant l'entraînement. Lorsque l'horizon de rollout  $W$  est suffisamment grand et que l'intégration est symplectique, l'objectif d'entraînement moyenne les erreurs de prédiction le long de trajectoires longues et géométriquement cohérentes, produisant ainsi un signal d'apprentissage à fort rapport signal/bruit. Dans ce régime, l'encodeur peut pleinement exploiter sa capacité de filtrage à mesure que  $C$  augmente. À l'inverse, pour de faibles valeurs de  $W$  ou en présence de rollouts non symplectiques, le signal de supervision devient lui-même bruité et biaisé par des artefacts numériques, ce qui limite le gain apporté par l'augmentation de  $C$ , malgré la croissance théorique de l'information de Fisher.

**Plancher de variance induit par l’ancrage observationnel.** Les projections sont initialisées à partir de la coordonnée observée  $y_0^{\text{obs}}$  comme défini en (8). Ainsi, indépendamment de la qualité d’inférence de  $\hat{h}_0$ , la covariance  $\text{Cov}(y_0^{\text{obs}}) = R$  impose

$$\Sigma_0 \succeq \begin{bmatrix} R & 0 \\ 0 & 0 \end{bmatrix}, \quad (14)$$

ce qui définit une borne inférieure irréductible sur l’incertitude initiale.

En outre, en régime d’observabilité partielle, la matrice d’information de Fisher (12) peut être singulière, empêchant l’identification complète de certaines directions latentes à partir d’un contexte de longueur finie. Il en résulte que l’allongement du contexte  $C$  réduit l’incertitude sur  $\hat{h}_0$  sans pouvoir annuler l’incertitude globale  $\Sigma_0$ .

**Interaction entre la longueur du contexte  $C$  et l’horizon de rollout  $W$ .** L’analyse précédente met en évidence que les bénéfices associés à l’augmentation de la longueur du contexte  $C$  et de l’horizon de rollout  $W$  sont intrinsèquement couplés. La longueur du contexte  $C$  régit l’efficacité statistique de l’inférence de l’état initial en agissant comme un filtre temporel, tandis que l’horizon  $W$  détermine le rapport signal/bruit de l’objectif d’entraînement lui-même. Lorsque  $W$  est suffisamment grand et que les rollouts sont produits par un intégrateur préservant la structure (symplectique), le signal de supervision devient stable et de faible variance, permettant à l’encodeur d’exploiter pleinement des fenêtres de contexte plus longues. Dans ce régime, accroître  $C$  réduit principalement la variance du moment latent inféré  $\hat{h}_0$ , ce qui conduit à des prédictions à long horizon plus cohérentes d’une graine aléatoire à l’autre.

À l’inverse, lorsque  $W$  est trop court ou lorsque les rollouts sont affectés par des artefacts numériques, le signal d’entraînement devient bruité et biaisé, et la capacité de filtrage de l’encodeur sature prématurément : l’augmentation de  $C$  produit alors des gains décroissants malgré l’accroissement théorique de l’information. Enfin, indépendamment du choix de  $(C, W)$ , un plancher de variance irréductible subsiste sous observabilité partielle, en raison de l’ancrage dans l’observation initiale bruitée  $y_0^{\text{obs}}$ . Ce mécanisme explique les compromis empiriques observés dans nos expériences et motive la sélection conjointe de  $(C, W)$ , ainsi que l’intégration symplectique, comme un problème de co-conception pour un apprentissage robuste à long horizon.

## B Sur la liberté de jauge de l’Hamiltonien appris

Dans ce travail, l’Hamiltonien  $H_\theta$  est appris dans un système de coordonnées latent et ne doit pas être interprété comme l’énergie physique réelle du système sous-jacent. Il définit plutôt une fonction d’énergie latente, identifiable uniquement à une transformation de jauge près.

**Liberté de jauge et non-identifiabilité de l’Hamiltonien appris.** Un même flot hamiltonien peut être représenté par différentes fonctions d’énergie, liées par des transformations de coordonnées. Les transformations affines  $H_\theta \mapsto aH_\theta + b$  ( $a > 0$ ) préservent la structure du champ hamiltonien à une dilatation temporelle près, tandis que des transformations dépendantes de l’état peuvent être interprétées comme des modifications de la métrique induite par l’énergie cinétique. Il en résulte que ni l’échelle absolue ni l’offset de  $H_\theta$  ne sont identifiables de manière unique à partir des seules observations.

**Implications pour les visualisations dans l’espace des phases.** Dans la mesure où le terme cinétique  $h^\top M_\theta(y)^{-1}h$  induit une métrique latente sur l’espace des phases, différents choix de jauge (par exemple via des matrices de masse apprises distinctes) entraînent des redimensionnements géométriques différents des trajectoires. Cela explique pourquoi la taille et la forme des orbites dans l’espace des phases peuvent varier d’un modèle à l’autre, même lorsque la structure qualitative de la dynamique est préservée. Il est essentiel de noter que ces redimensionnements n’affectent ni la symplecticité ni l’organisation topologique des trajectoires, qui constituent les objets principaux de notre analyse.

**Rôle de la dérive de l’Hamiltonien appris.** Compte tenu de la liberté de jauge, le niveau absolu de  $H_\theta$  ne possède pas de signification physique directe. La dérive relative de  $H_\theta$  est donc interprétée comme un indicateur de cohérence interne plutôt que comme une mesure d’erreur d’énergie physique. Une dérive croissante traduit typiquement des incohérences entre la dynamique apprise et sa discrétisation (intégration non symplectique, distorsions géométriques), alors qu’une dérive faible et bornée indique une préservation effective de la structure. Cette lecture rejoint l’analyse biais–variance de l’Annexe A, dans laquelle les violations structurelles induisent une amplification numérique de l’incertitude.

## C Systèmes dynamiques et distributions des conditions initiales

Dans cet appendice, nous détaillons les équations dynamiques employées pour la génération des trajectoires, ainsi que les distributions de conditions initiales (CI) propres à chaque système. Les simulations sont réalisées avec un pas de temps constant  $\Delta t = 0.02$  et une durée totale de  $T = 500$  pas. Sauf mention explicite, les distributions de CI sont identiques pour les phases d’entraînement, de validation et d’évaluation.

**Oscillateur de Duffing non amorti (2D).** Le système étudié est l’oscillateur de Duffing conservatif, défini par  $x = (q, v) \in \mathbb{R}^2$  et régi par

$$\dot{q} = v, \quad \dot{v} = -\frac{k}{m}q - \frac{\alpha}{m}q^3. \quad (15)$$

Les paramètres sont fixés à  $m = 1$ ,  $k = 1$  et  $\alpha = 1$ . Les conditions initiales  $(q_0, v_0)$  sont tirées indépendamment selon

$$(q_0, v_0) \sim \mathcal{U}([-\beta, \beta]^2), \quad \beta = 0.5. \quad (16)$$

**Pendule simple conservatif (2D).** Le système étudié est un pendule sans dissipation, défini par  $x = (\theta, \omega) \in \mathbb{R}^2$  et régi par

$$\dot{\theta} = \omega, \quad \dot{\omega} = -\frac{g}{\ell} \sin(\theta). \quad (17)$$

Les paramètres physiques sont fixés à  $g = 9.81$  et  $\ell = 1$ . Les conditions initiales  $(\theta_0, \omega_0)$  sont tirées indépendamment selon

$$(\theta_0, \omega_0) \sim \mathcal{U}([-\beta, \beta]^2), \quad \beta = 0.5. \quad (18)$$

**Problème de Kepler (4D).** Nous considérons le problème de Kepler planaire, défini par les coordonnées canoniques  $x = (q_1, q_2, p_1, p_2) \in \mathbb{R}^4$  et l'Hamiltonien

$$H(q, p) = \frac{1}{2} \|p\|_2^2 - \frac{\mu}{\|q\|_2}, \quad q = (q_1, q_2), \quad p = (p_1, p_2), \quad (19)$$

qui induit la dynamique

$$\dot{q} = p, \quad \dot{p} = -\mu \frac{q}{\|q\|_2^3}. \quad (20)$$

Nous utilisons  $\mu = 1$ . Les conditions initiales sont générées en échantillonnant un rayon  $r \sim \mathcal{U}[r_{\min}, r_{\max}]$  avec  $(r_{\min}, r_{\max}) = (0.7, 1.3)$ , ainsi qu'un angle  $\varphi \sim \mathcal{U}[0, 2\pi]$ , puis en posant

$$q_0 = (r \cos \varphi, r \sin \varphi), \quad p_0 = v_{\text{circ}} (-\sin \varphi, \cos \varphi), \quad v_{\text{circ}} = \sqrt{\mu/r}. \quad (21)$$

Cette procédure génère des orbites quasi elliptiques avec une vitesse initiale tangentielle.

**Système de Hénon–Heiles (4D).** Nous considérons le système hamiltonien de Hénon–Heiles, défini par les coordonnées canoniques  $x = (q_1, q_2, p_1, p_2) \in \mathbb{R}^4$  et l'Hamiltonien

$$H(q, p) = \frac{1}{2}(p_1^2 + p_2^2) + V(q_1, q_2), \quad V(q_1, q_2) = \frac{1}{2}(q_1^2 + q_2^2) + \left(q_1^2 q_2 - \frac{1}{3} q_2^3\right). \quad (22)$$

Les équations du mouvement induites sont

$$\dot{q}_1 = p_1, \quad \dot{q}_2 = p_2, \quad \dot{p}_1 = -\frac{\partial V}{\partial q_1} = -(q_1 + 2q_1 q_2), \quad \dot{p}_2 = -\frac{\partial V}{\partial q_2} = -(q_2 + q_1^2 - q_2^2). \quad (23)$$

Les conditions initiales sont échantillonnées à une énergie cible fixée  $E$  (nous utilisons  $E = 0.05$ ). Nous tirons d'abord  $q_0 \sim \mathcal{U}([-\beta, \beta]^2)$  avec  $\beta = 0.5$ , et nous acceptons les échantillons vérifiant  $V(q_0) < E$ . Pour chaque  $q_0$  accepté, nous tirons un angle aléatoire  $\psi \sim \mathcal{U}[0, 2\pi]$  et ajustons la norme du moment de façon à correspondre à l'énergie cinétique résiduelle :

$$\|p_0\|_2 = \sqrt{2(E - V(q_0))}, \quad p_0 = \|p_0\|_2 (\cos \psi, \sin \psi). \quad (24)$$

## D Architecture du modèle et réglages d'hyperparamètres

Cet appendice résume les architectures et les hyperparamètres par défaut utilisés pour l'ensemble des modèles présentés en §2 et §3.

**Configuration commune (dimensions observées et latentes).** Soit  $y_t \in \mathbb{R}^{d_y}$  les variables observées (configurations) et  $h_t \in \mathbb{R}^{d_h}$  l'augmentation latente (interprétée comme des moments dans les modèles hamiltoniens latents). Dans le cadre partiellement observé, nous utilisons  $d_h = d_y$  afin de permettre un appariement symplectique canonique. Dans le cadre pleinement observé, nous posons  $d_h = 0$  et  $y_t$  contient l'état canonique complet (par exemple  $[q, p]$ ).

**Encodeur de contexte (GRU).** L'inférence de l'état latent initial repose, pour tous les modèles, sur un encodeur commun  $E_\phi$  :

- **Backbone** : un réseau récurrent de type GRU, recevant des entrées de dimension  $d_y$ , avec une dimension cachée  $H_{\text{enc}} = 64$  et  $L_{\text{enc}} = 2$  couches empilées.
- **Readout** : l'état caché final est transformé par une projection linéaire vers  $\mathbb{R}^{d_h}$  afin d'obtenir l'estimation du moment latent initial  $\hat{h}_0$ .

**Latent NODE (E) : champ de vecteurs augmenté non contraint.** La baseline Latent NODE paramètre un champ de vecteurs en temps continu non contraint sur l'état augmenté  $z = [y, h] \in \mathbb{R}^{d_y+d_h}$  :

$$\dot{z} = f_\theta(z), \quad (25)$$

où  $f_\theta$  est un perceptron multicouche (MLP) avec activations tanh :

- **Entrée/sortie** :  $(d_y + d_h) \rightarrow (d_y + d_h)$ .
- **Largeur/profondeur** : largeur cachée  $H = 64$  et profondeur  $D = 3$  (deux couches cachées avec activations tanh, suivies d'une couche de sortie linéaire).
- **Initialisation** : la couche linéaire finale est initialisée à zéro (poids et biais), de sorte que la dynamique initiale soit quasi nulle et que l'entraînement débute à partir de rollouts stables.

Les rollouts sont intégrés à l'aide d'un schéma d'Euler explicite (E).

**Latent HNN (E/S) : dynamique augmentée hamiltonienne structurée.** Les modèles Latent HNN reposent sur une paramétrisation hamiltonienne de la forme

$$H_\theta(y, h) = \frac{1}{2} h^\top M_\theta(y)^{-1} h + V_\theta(y), \quad (26)$$

où  $V_\theta : \mathbb{R}^{d_y} \rightarrow \mathbb{R}$  désigne l'énergie potentielle, et où  $M_\theta(y)^{-1}$  est une matrice symétrique définie positive (SPD) dépendant de  $y$ . Cette paramétrisation est implémentée comme suit :

- **Réseau de potentiel** :  $V_\theta(y)$  est modélisé par un MLP  $d_y \rightarrow 1$ , de largeur  $H = 64$ , de profondeur  $D = 3$ , avec activations tanh. La dernière couche est initialisée avec de petits poids aléatoires (gaussiens, écart-type  $10^{-3}$ ) et un biais nul.
- **Réseau de masse** :  $M_\theta(y)^{-1}$  est paramétrée via un facteur de Cholesky  $L_\theta(y)$ . Les coefficients triangulaires inférieurs de  $L_\theta(y) \in \mathbb{R}^{d_h \times d_h}$  sont prédits par un MLP  $d_y \rightarrow \frac{d_h(d_h+1)}{2}$ , de largeur  $H = 64$ , profondeur  $D = 3$ , avec activations tanh ; la couche de sortie est initialisée à zéro. Les éléments diagonaux de  $L_\theta(y)$  sont passés à travers une transformation softplus et décalés par  $\varepsilon$  afin de garantir une positivité stricte.
- **Construction SPD** : nous définissons

$$M_\theta(y)^{-1} = L_\theta(y)L_\theta(y)^\top + \varepsilon I, \quad \varepsilon = 10^{-6}. \quad (27)$$

**Discretisation temporelle.** Deux schémas d'intégration sont considérés :

- **Euler explicite (E)**, défini par  $z_{k+1} = z_k + \Delta t \dot{z}_k$ .
- **Euler symplectique « kick-drift » (S)**, employé pour les dynamiques hamiltoniennes et cohérent avec (4)–(5). Sous observabilité complète ( $y = [q, p]$ ), la mise à jour est appliquée après séparation des variables canoniques  $(q, p)$ .

## E Détails des expériences

TABLE 3 – Balayage détaillé des fenêtres sur le système de Duffing : RMSE de test (moyenne  $\pm$  écart-type sur les graines aléatoires) pour différentes valeurs de  $(C, W)$  et pour chaque méthode. Les valeurs en gras indiquent le meilleur couple  $(C, W)$  (RMSE moyenne minimale) pour chaque ligne.

Duffing — Observations propres									
Model	$W = 10$	$C = 10$ $W = 25$	$W = 50$	$W = 10$	$C = 25$ $W = 25$	$W = 50$	$W = 10$	$C = 50$ $W = 25$	$W = 50$
Latent NODE (E)	0.2920 $\pm$ 0.1727	0.1794 $\pm$ 0.1665	0.1540 $\pm$ 0.1428	0.0302 $\pm$ 0.0051	0.0291 $\pm$ 0.0044	0.0267 $\pm$ 0.0039	0.0562 $\pm$ 0.0147	0.0389 $\pm$ 0.0060	<b>0.0238 <math>\pm</math> 0.0022</b>
Latent HNN (E)	0.0465 $\pm$ 0.0129	0.0212 $\pm$ 0.0203	0.0168 $\pm$ 0.0075	0.0359 $\pm$ 0.0178	0.0188 $\pm$ 0.0117	0.0173 $\pm$ 0.0038	0.0134 $\pm$ 0.0039	0.0129 $\pm$ 0.0054	<b>0.0071 <math>\pm</math> 0.0005</b>
Latent HNN (S)	0.0470 $\pm$ 0.0201	0.0211 $\pm$ 0.0180	0.0165 $\pm$ 0.0060	0.0326 $\pm$ 0.0246	0.0212 $\pm$ 0.0078	0.0171 $\pm$ 0.0048	0.0143 $\pm$ 0.0071	0.0128 $\pm$ 0.0078	<b>0.0035 <math>\pm</math> 0.0011</b>
Duffing — Observations bruitées									
Model	$W = 10$	$C = 10$ $W = 25$	$W = 50$	$W = 10$	$C = 25$ $W = 25$	$W = 50$	$W = 10$	$C = 50$ $W = 25$	$W = 50$
Latent NODE (E)	0.1514 $\pm$ 0.0002	0.1413 $\pm$ 0.0032	0.0806 $\pm$ 0.0038	0.1157 $\pm$ 0.0204	0.0442 $\pm$ 0.0051	<b>0.0354 <math>\pm</math> 0.0023</b>	0.2433 $\pm$ 0.0811	0.0608 $\pm$ 0.0142	0.0398 $\pm$ 0.0057
Latent HNN (E)	0.1171 $\pm$ 0.0170	0.0971 $\pm$ 0.0062	0.1263 $\pm$ 0.0042	0.1331 $\pm$ 0.0663	0.0481 $\pm$ 0.0042	0.0365 $\pm$ 0.0033	0.1050 $\pm$ 0.0798	0.0418 $\pm$ 0.0044	<b>0.0295 <math>\pm</math> 0.0030</b>
Latent HNN (S)	0.1237 $\pm$ 0.0349	0.0950 $\pm$ 0.0063	0.1261 $\pm$ 0.0031	0.1499 $\pm$ 0.0856	0.0452 $\pm$ 0.0012	0.0355 $\pm$ 0.0025	0.1244 $\pm$ 0.1061	0.0401 $\pm$ 0.0007	<b>0.0292 <math>\pm</math> 0.0032</b>

TABLE 4 – Balayage détaillé des fenêtres sur le système de Hénon : RMSE de test (moyenne  $\pm$  écart-type sur les graines aléatoires) pour différentes valeurs de  $(C, W)$  et pour chaque méthode. Les valeurs en gras indiquent le meilleur couple  $(C, W)$  (RMSE moyenne minimale) pour chaque ligne.

Hénon — Observations propres									
Model	$W = 10$	$C = 10$ $W = 25$	$W = 50$	$W = 10$	$C = 25$ $W = 25$	$W = 50$	$W = 10$	$C = 50$ $W = 25$	$W = 50$
Latent NODE (E)	0.2049 $\pm$ 0.0201	0.1807 $\pm$ 0.0234	0.1590 $\pm$ 0.0204	0.1847 $\pm$ 0.0030	0.1714 $\pm$ 0.0002	0.0994 $\pm$ 0.0548	0.1895 $\pm$ 0.0087	0.1350 $\pm$ 0.0369	<b>0.0381 <math>\pm</math> 0.0049</b>
Latent HNN (E)	0.0870 $\pm$ 0.0029	0.0526 $\pm$ 0.0041	0.0450 $\pm$ 0.0056	0.0584 $\pm$ 0.0091	0.0455 $\pm$ 0.0048	0.0424 $\pm$ 0.0052	0.0685 $\pm$ 0.0164	0.0437 $\pm$ 0.0024	<b>0.0367 <math>\pm</math> 0.0060</b>
Latent HNN (S)	0.1196 $\pm$ 0.0215	0.0559 $\pm$ 0.0067	0.0425 $\pm$ 0.0053	0.0743 $\pm$ 0.0130	0.0439 $\pm$ 0.0047	0.0409 $\pm$ 0.0044	0.0805 $\pm$ 0.0237	0.0428 $\pm$ 0.0030	<b>0.0379 <math>\pm</math> 0.0035</b>
Hénon — Observations bruitées									
Model	$W = 10$	$C = 10$ $W = 25$	$W = 50$	$W = 10$	$C = 25$ $W = 25$	$W = 50$	$W = 10$	$C = 50$ $W = 25$	$W = 50$
Latent NODE (E)	0.1987 $\pm$ 0.0060	0.1680 $\pm$ 0.0007	0.1439 $\pm$ 0.0134	0.1887 $\pm$ 0.0153	0.1475 $\pm$ 0.0206	0.1334 $\pm$ 0.0286	0.1814 $\pm$ 0.0073	0.1739 $\pm$ 0.0019	<b>0.1124 <math>\pm</math> 0.0344</b>
Latent HNN (E)	0.1614 $\pm$ 0.0210	0.0818 $\pm$ 0.0072	0.1336 $\pm$ 0.0005	0.1416 $\pm$ 0.0349	0.0707 $\pm$ 0.0058	0.0585 $\pm$ 0.0021	0.1280 $\pm$ 0.0620	0.0697 $\pm$ 0.0026	<b>0.0568 <math>\pm</math> 0.0045</b>
Latent HNN (S)	0.1526 $\pm$ 0.0000	0.0779 $\pm$ 0.0000	0.1340 $\pm$ 0.0000	0.1586 $\pm$ 0.0000	0.0672 $\pm$ 0.0000	0.0577 $\pm$ 0.0000	0.1921 $\pm$ 0.0000	0.0656 $\pm$ 0.0000	<b>0.0524 <math>\pm</math> 0.0000</b>

TABLE 5 – Balayage détaillé des fenêtres sur le système de Kepler : RMSE de test (moyenne  $\pm$  écart-type sur les graines aléatoires) pour différentes valeurs de  $(C, W)$  et pour chaque méthode. Les valeurs en gras indiquent le meilleur couple  $(C, W)$  (RMSE moyenne minimale) pour chaque ligne.

Kepler — Observations propres									
Model	$W = 10$	$C = 10$ $W = 25$	$W = 50$	$W = 10$	$C = 25$ $W = 25$	$W = 50$	$W = 10$	$C = 50$ $W = 25$	$W = 50$
Latent NODE (E)	0.3049 $\pm$ 0.2995	0.2618 $\pm$ 0.2547	<b>0.1884 <math>\pm</math> 0.1707</b>	0.5820 $\pm$ 0.0085	0.4653 $\pm$ 0.0265	0.3631 $\pm$ 0.0230	0.6166 $\pm$ 0.0202	0.4879 $\pm$ 0.0581	0.3630 $\pm$ 0.0361
Latent HNN (E)	0.2499 $\pm$ 0.0755	0.0991 $\pm$ 0.0164	0.0604 $\pm$ 0.0041	0.1555 $\pm$ 0.0319	0.0792 $\pm$ 0.0110	0.0481 $\pm$ 0.0044	0.1561 $\pm$ 0.0285	0.0724 $\pm$ 0.0125	<b>0.0399 <math>\pm</math> 0.0060</b>
Latent HNN (S)	0.1434 $\pm$ 0.0319	0.0645 $\pm$ 0.0112	0.0333 $\pm$ 0.0036	0.1125 $\pm$ 0.0199	0.0549 $\pm$ 0.0075	<b>0.0324 <math>\pm</math> 0.0021</b>	0.1197 $\pm$ 0.0141	0.0613 $\pm$ 0.0077	0.0419 $\pm$ 0.0024
Kepler — Observations bruitées									
Model	$W = 10$	$C = 10$ $W = 25$	$W = 50$	$W = 10$	$C = 25$ $W = 25$	$W = 50$	$W = 10$	$C = 50$ $W = 25$	$W = 50$
Latent NODE (E)	0.5872 $\pm$ 0.2940	0.3991 $\pm$ 0.0160	<b>0.3494 <math>\pm</math> 0.0322</b>	0.5023 $\pm$ 0.2027	0.3875 $\pm$ 0.0109	0.3562 $\pm$ 0.0270	0.6118 $\pm$ 0.3321	0.3944 $\pm$ 0.0084	0.3514 $\pm$ 0.0445
Latent HNN (E)	0.3518 $\pm$ 0.1097	0.0748 $\pm$ 0.0131	0.0619 $\pm$ 0.0122	0.3582 $\pm$ 0.1096	0.0735 $\pm$ 0.0100	0.0519 $\pm$ 0.0059	0.3040 $\pm$ 0.1007	0.0800 $\pm$ 0.0031	<b>0.0516 <math>\pm</math> 0.0005</b>
Latent HNN (S)	0.3690 $\pm$ 0.0957	0.1122 $\pm$ 0.0061	<b>0.0523 <math>\pm</math> 0.0132</b>	0.3496 $\pm$ 0.1209	0.1230 $\pm$ 0.0083	0.0547 $\pm$ 0.0059	0.3077 $\pm$ 0.0597	0.1318 $\pm$ 0.0001	0.0692 $\pm$ 0.0000

TABLE 6 – Balayage détaillé des fenêtres sur le système Pendule : RMSE de test (moyenne  $\pm$  écart-type sur les graines aléatoires) pour différentes valeurs de  $(C, W)$  et pour chaque méthode. Les valeurs en gras indiquent le meilleur couple  $(C, W)$  (RMSE moyenne minimale) pour chaque ligne.

<b>Pendule — Observations propres</b>									
<b>Model</b>	$C = 10$			$C = 25$			$C = 50$		
	$W = 10$	$W = 25$	$W = 50$	$W = 10$	$W = 25$	$W = 50$	$W = 10$	$W = 25$	$W = 50$
Latent NODE (E)	0.1301 $\pm$ 0.1059	0.1178 $\pm$ 0.0979	0.1082 $\pm$ 0.1002	0.0505 $\pm$ 0.0034	0.0287 $\pm$ 0.0007	<b>0.0109 <math>\pm</math> 0.0035</b>	0.0495 $\pm$ 0.0051	0.0303 $\pm$ 0.0017	0.0147 $\pm$ 0.0032
Latent HNN (E)	0.0842 $\pm$ 0.0455	0.0589 $\pm$ 0.0045	0.0477 $\pm$ 0.0028	0.0619 $\pm$ 0.0070	0.0512 $\pm$ 0.0058	<b>0.0445 <math>\pm</math> 0.0033</b>	0.0674 $\pm$ 0.0105	0.0539 $\pm$ 0.0054	0.0510 $\pm$ 0.0042
Latent HNN (S)	0.0286 $\pm$ 0.0206	0.0155 $\pm$ 0.0001	0.0037 $\pm$ 0.0004	0.0128 $\pm$ 0.0021	0.0062 $\pm$ 0.0015	<b>0.0035 <math>\pm</math> 0.0005</b>	0.0132 $\pm$ 0.0020	0.0065 $\pm$ 0.0003	0.0040 $\pm$ 0.0001
<b>Pendule — Observations bruitées</b>									
<b>Model</b>	$C = 10$			$C = 25$			$C = 50$		
	$W = 10$	$W = 25$	$W = 50$	$W = 10$	$W = 25$	$W = 50$	$W = 10$	$W = 25$	$W = 50$
Latent NODE (E)	0.0628 $\pm$ 0.0177	0.0408 $\pm$ 0.0029	0.0342 $\pm$ 0.0018	0.0856 $\pm$ 0.0118	0.0367 $\pm$ 0.0024	<b>0.0227 <math>\pm</math> 0.0014</b>	0.1047 $\pm$ 0.0118	0.0396 $\pm$ 0.0009	0.0251 $\pm$ 0.0028
Latent HNN (E)	0.0774 $\pm$ 0.0049	0.0886 $\pm$ 0.0000	0.0578 $\pm$ 0.0008	0.0808 $\pm$ 0.0088	0.0509 $\pm$ 0.0036	<b>0.0483 <math>\pm</math> 0.0018</b>	0.1000 $\pm$ 0.0135	0.0600 $\pm$ 0.0072	0.0554 $\pm$ 0.0069
Latent HNN (S)	0.0501 $\pm$ 0.0168	0.0552 $\pm$ 0.0011	0.0295 $\pm$ 0.0010	0.0589 $\pm$ 0.0097	0.0280 $\pm$ 0.0034	0.0166 $\pm$ 0.0002	0.0455 $\pm$ 0.0041	0.0293 $\pm$ 0.0005	<b>0.0158 <math>\pm</math> 0.0010</b>