

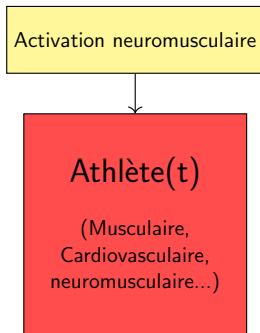
Estimer la RPE à partir de données de puissance

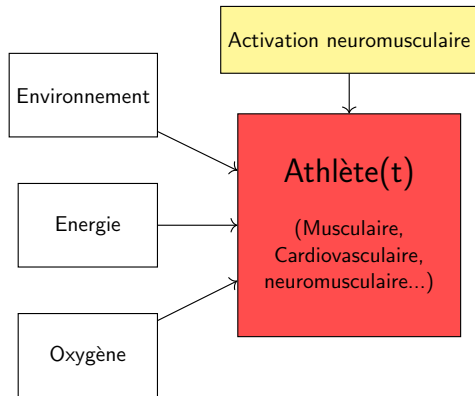
Pierre Marrec

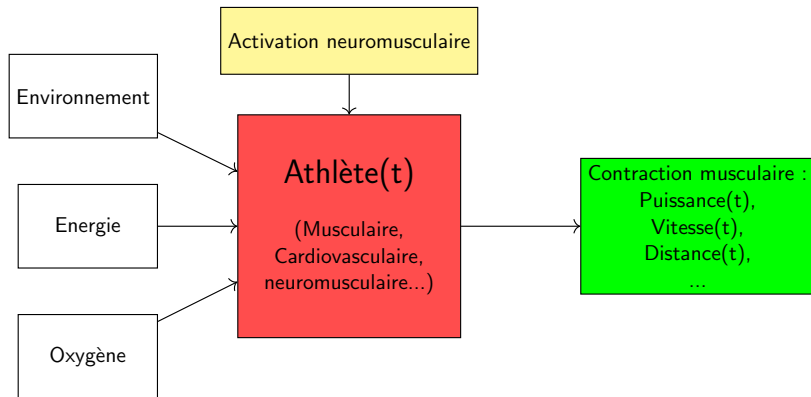
4 juillet 2024

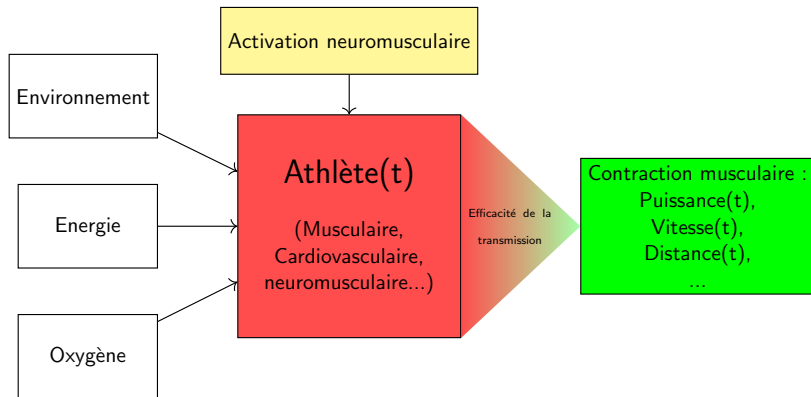
Athlète(t)

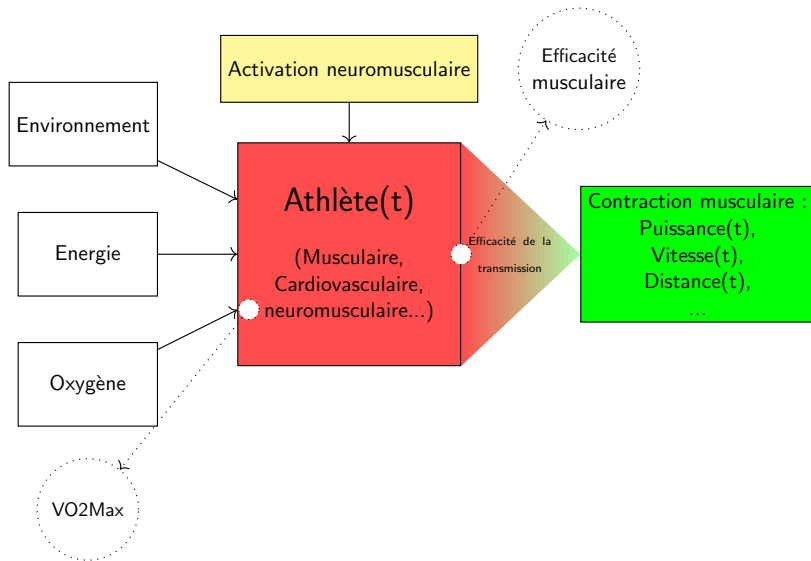
(Musculaire,
Cardiovasculaire,
neuromusculaire...)

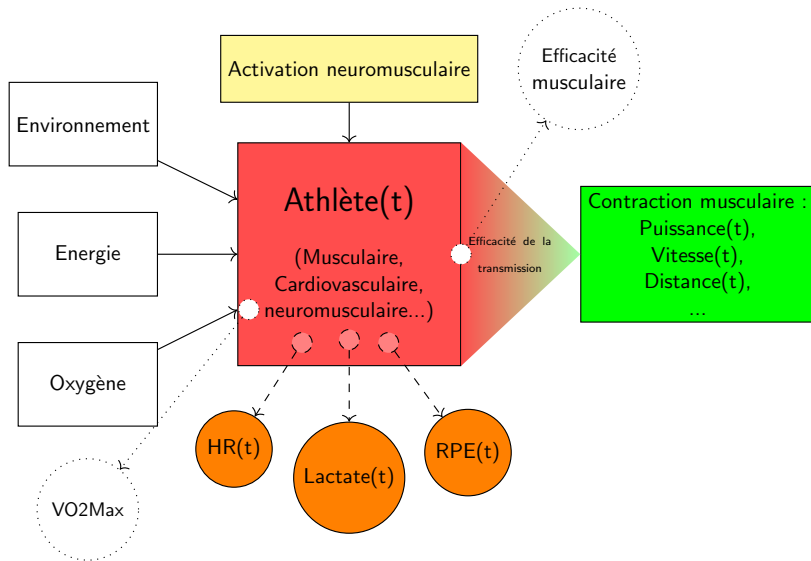


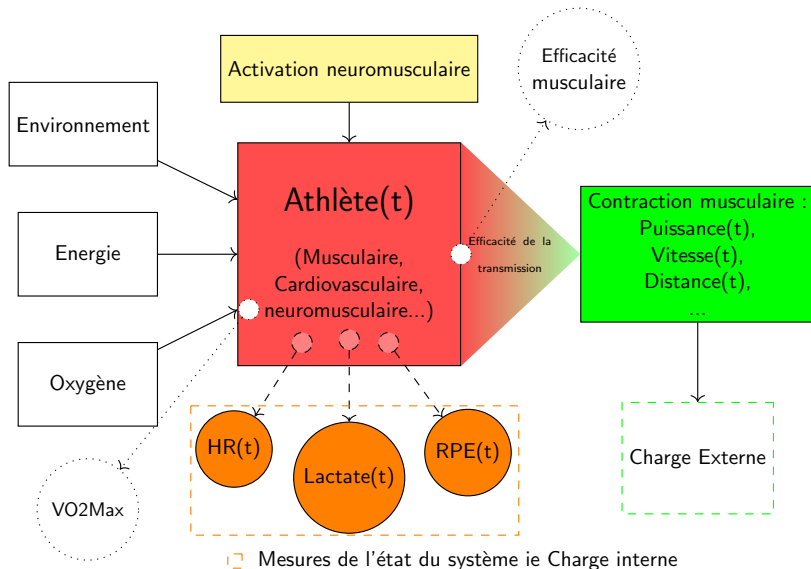








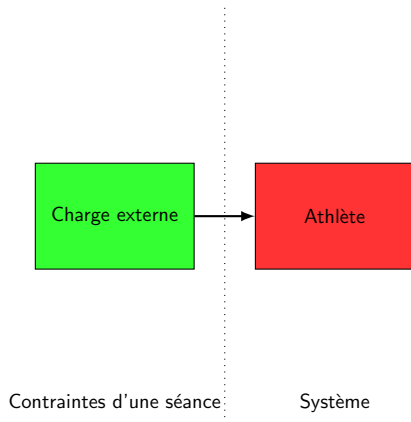


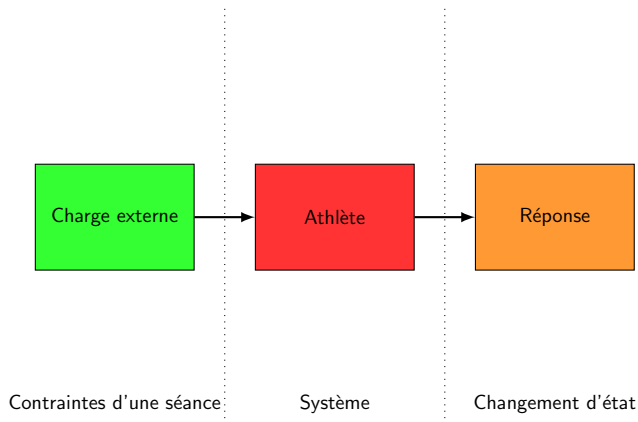


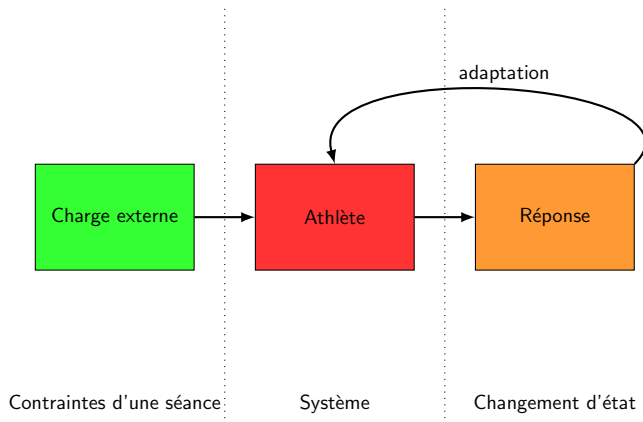


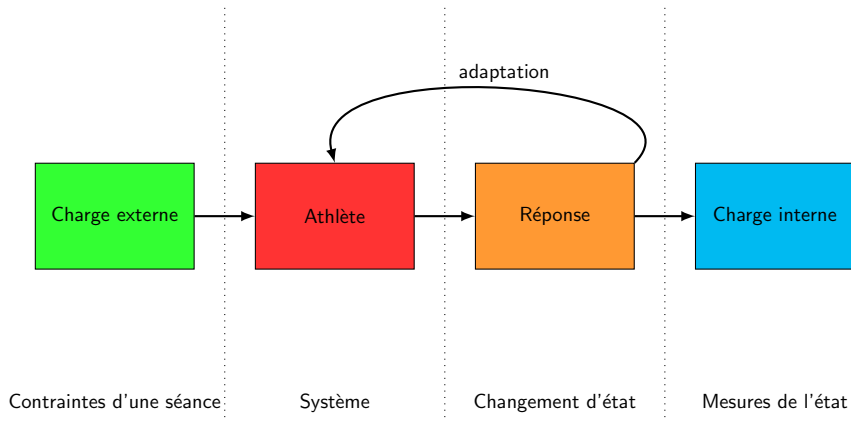
Athlète

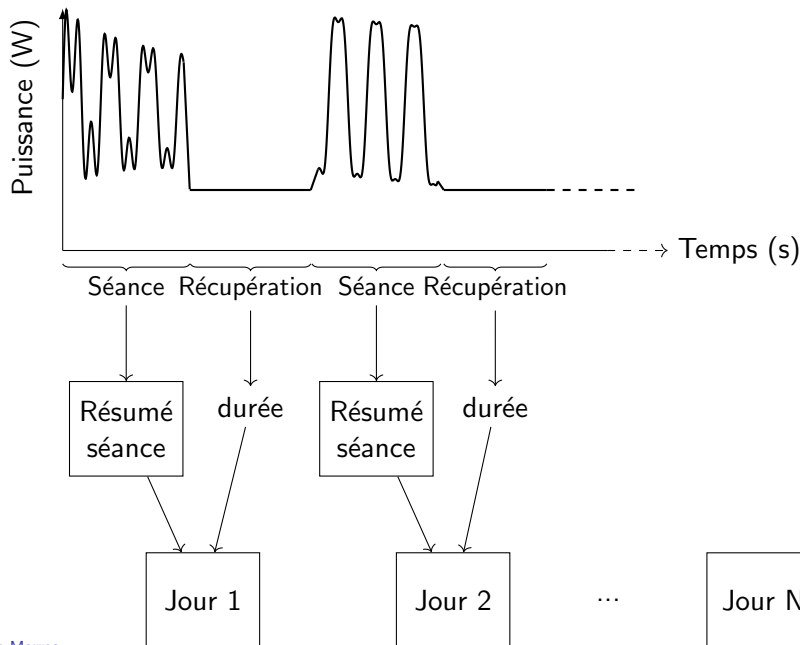
Système











1. Comprendre et nettoyer les données
2. Résumé des séances : Modèle de markov caché (MMC).
3. Prédiction de la RPE : Utilisation de l'encodeur-décodeur pour prédire la RPE à partir des résumés de séance.

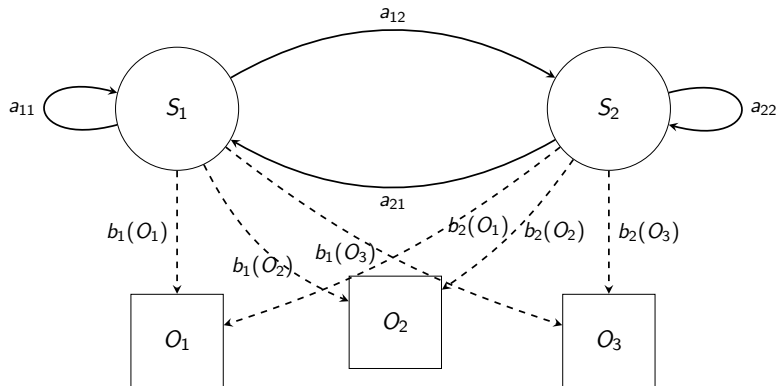
1. 1 cycliste de juin 2020 à avril 2024

Nettoyage :

- ▶ suppression des séquences sans données $> 5\text{min}$,
- ▶ interpolation au maximum,
- ▶ suppression des valeurs aberrantes et interpolation,
- ▶ séance trop courtes ($< 10\text{ min}$),
- ▶ séances avec des moyennes absurdes (négatives, trop basses).

etat	nombre de points	nombres de séances
complète	9440286	1534
nettoyée	4239646	715

2. 13 athlètes de juin 2020 à avril 2022 : 63M de points 13,709 séances
(39M de points nettoyés, 5,292 séances nettoyées)



S_i : États cachés

O_i : Observations

a_{ij} : Probabilités de transition

$b_i(O_j)$: Probabilités d'émission

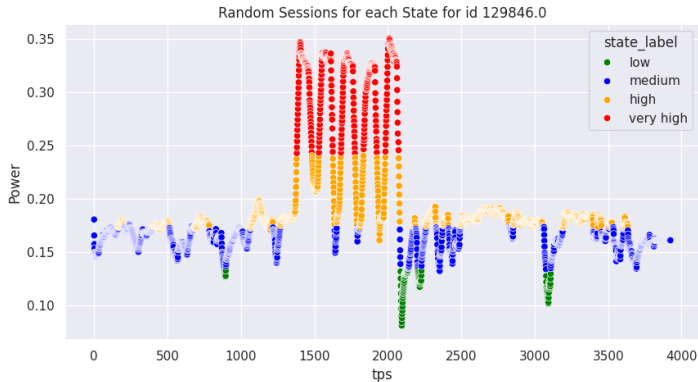


Figure 1 – Les zones décrite par le HMM pour une séance

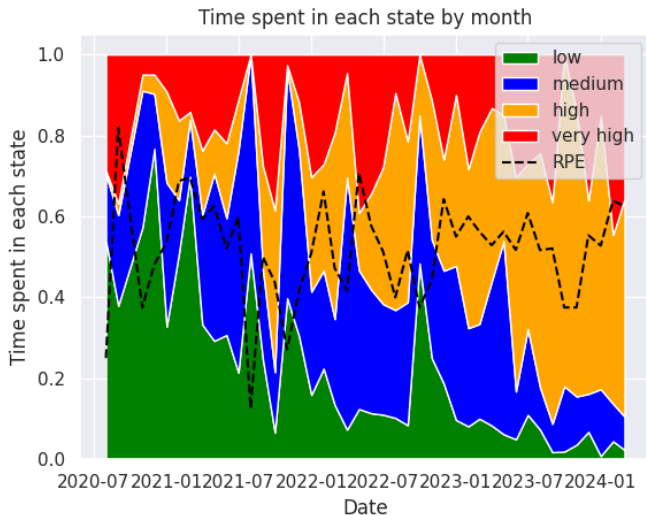


Figure 2 – Temps passé dans chaque zone au cours du temps

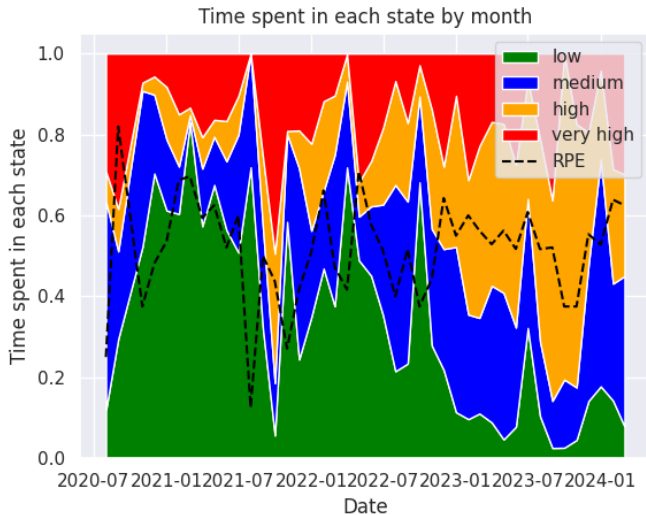
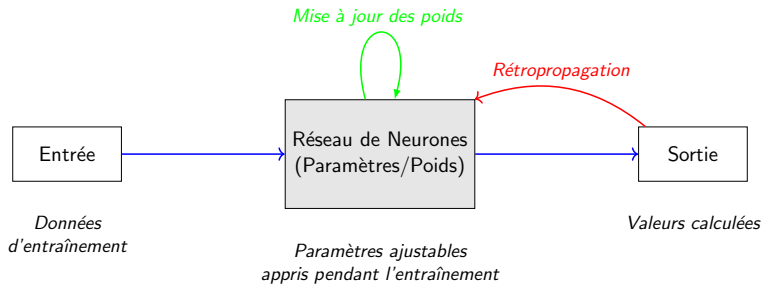
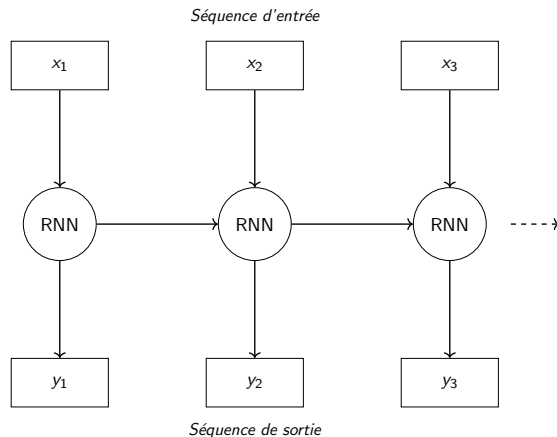
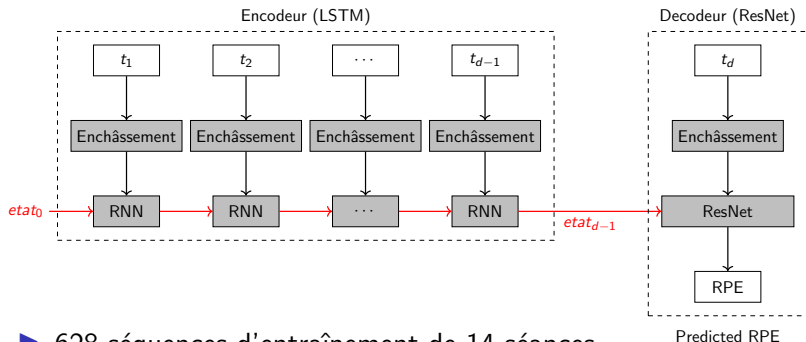


Figure 3 – Temps passé dans chaque zone au cours du temps





Le RNN traite les séquences en appliquant les mêmes opérations de manière répétée à chaque pas de temps



- ▶ 628 séquences d'entraînement de 14 séances
- ▶ 114 séquences de test de 14 séances
- ▶ L'état caché de sortie est sauvegardé pour la séquence suivante.
- ▶ Entrée : durée et puissance moyenne dans les 4 zones de puissance du MMC + temps entre les séances.

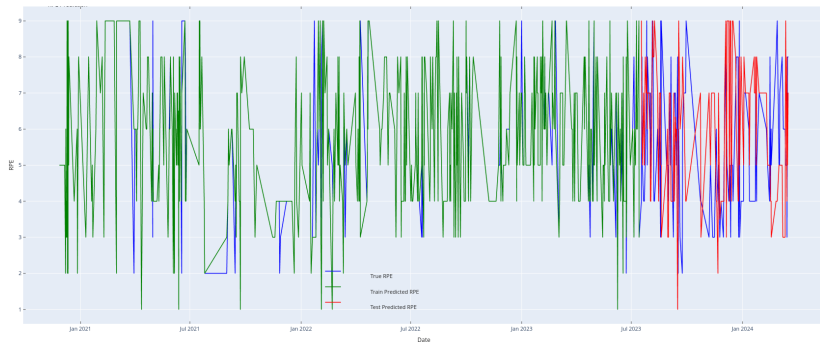


Figure 4 – Prédiction de la RPE pour des séquences de 14 séances

Métrique	Train	Test
Accuracy	92,0%	15,6%
Mean Absolute Percentage Error (MAPE)	5,0%	45%
Weighted Average Percentage Error (WAPEE)	3,7%	39%



Figure 5 – Prédiction de la RPE pour des séquences de 14 séances

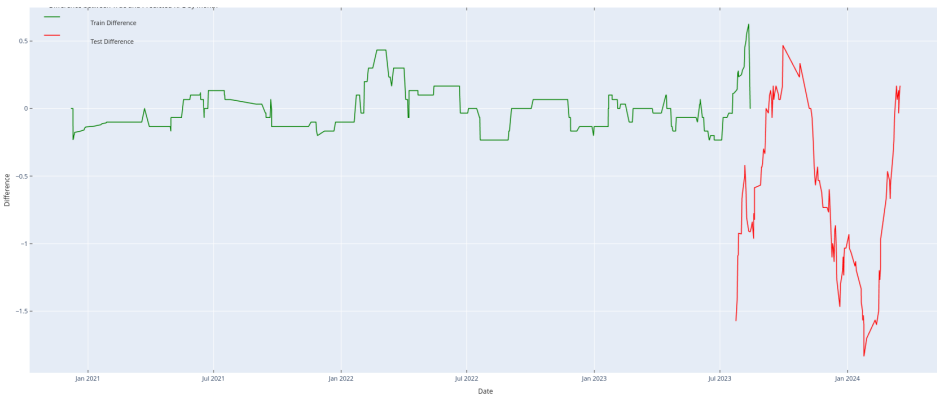


Figure 6 – Différence entre la prédiction et la vraie RPE



Figure 7 – Prédiction de la RPE pour des séquences de 14 séances regroupées par mois

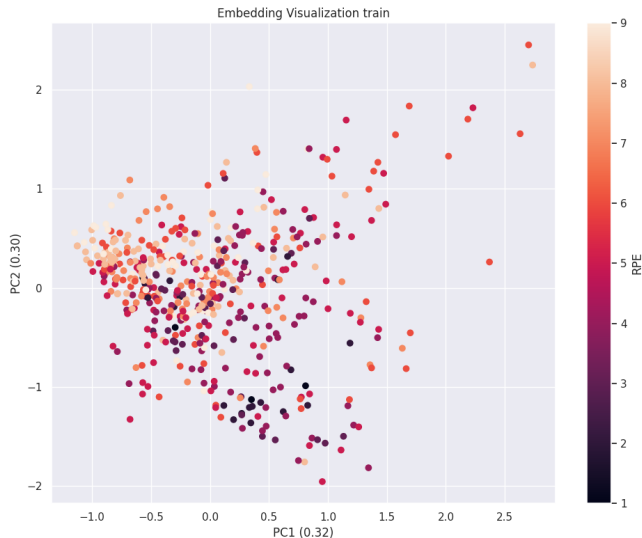


Figure 8 – Projection des séances dans l'espace latent pour les 2 plus grandes composantes

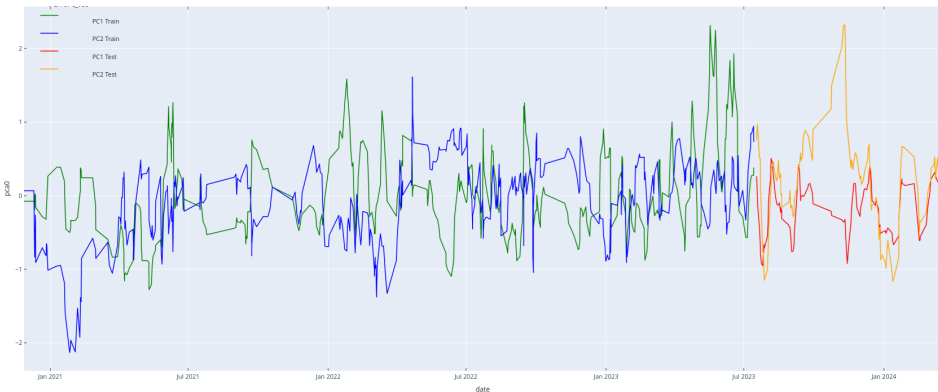


Figure 9 – Composantes principales de la représentation interne au cours du temps

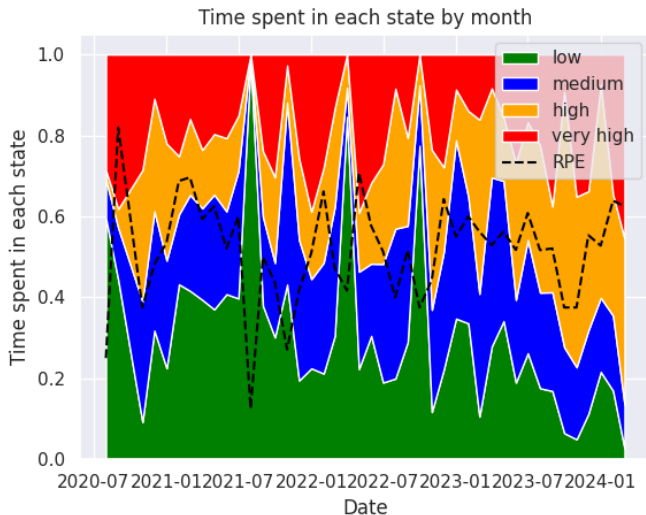
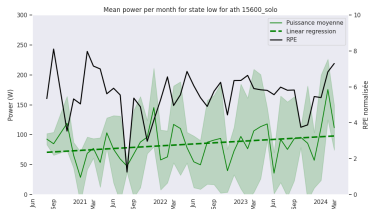
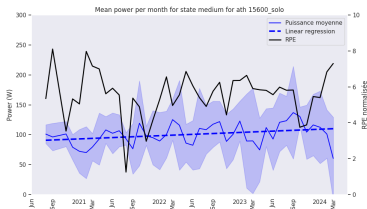


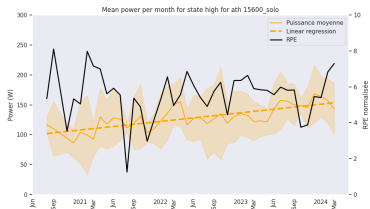
Figure 10 – Temps passé dans chaque zone au cours du temps



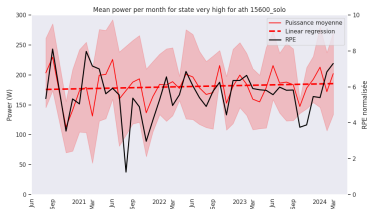
(a) low



(b) medium



(c) high



(d) very high

Puissance moyenne par état au cours du temps avec régression linéaire

► Résultats :

1. La prédiction sur le jeu d'entraînement est bon mais très mauvais sur le jeu de test. (surentraînement ?)
2. Légère séparation des séances dans l'espace latent et tendance croissante de la représentation interne.
3. Visualisation de la progression à partir d'états cachés tiré de la fréquence cardiaque.

► Limites :

1. Résumé de séance : pas assez représentatif ?
2. Les données : manque des séances (nettoyée, pas trackée), le test est trop différent du training.
3. Le modèle : pas adapté au nombre de données, pas suffisamment efficace, surentraînement

► Amélioration :

1. MMC avec des émissions variables / effet aléatoire. Sur des fenêtres glissantes ?
2. Apprendre une représentation implicite des séances (State Spaces Models)
3. Utiliser des modèles de prédiction plus complexes (Convolutions temporelles, transformer, SSM)

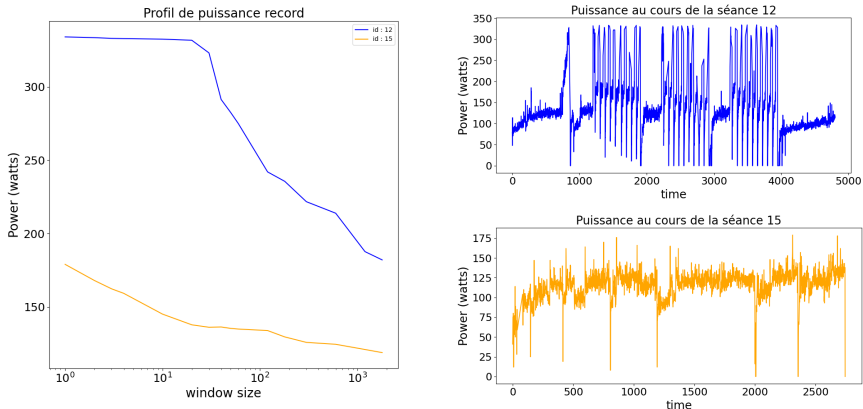


Figure 12 – Comparaison des PPR de deux séances

- ▶ Décrit assez bien l'intensité de la séance
- ▶ Ne décrit pas le temps passé dans ces intensités

- ▶ Problème pour agréger les séances : max ou moyenne ?
- ▶ On définit 4 zones à partir de la PPR à 30 secondes, 10 min, 1h30
 - ⇒ Pas cohérent avec les zones de puissance classique et trop dépendant de la manière d'agréger.

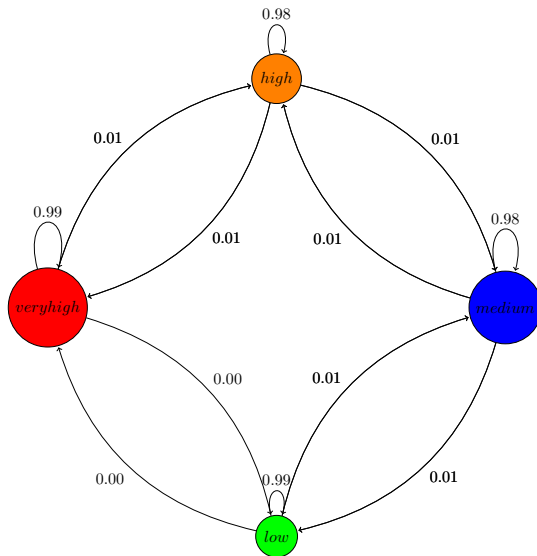


Figure 13 – Matrice d'émission pour les 4 zones

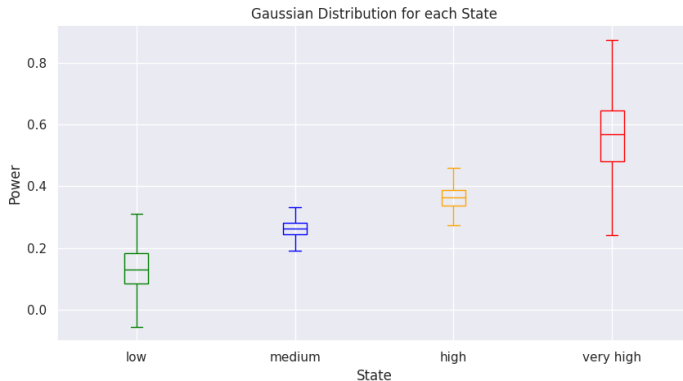


Figure 14 – Box plot de la puissance moyenne par zone

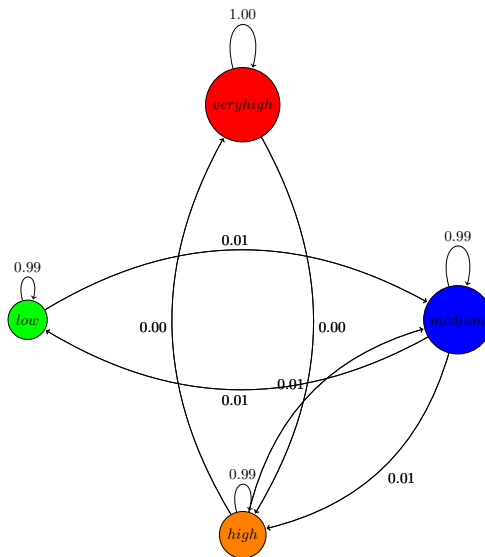


Figure 15 – Matrice de transition pour les 4 zones

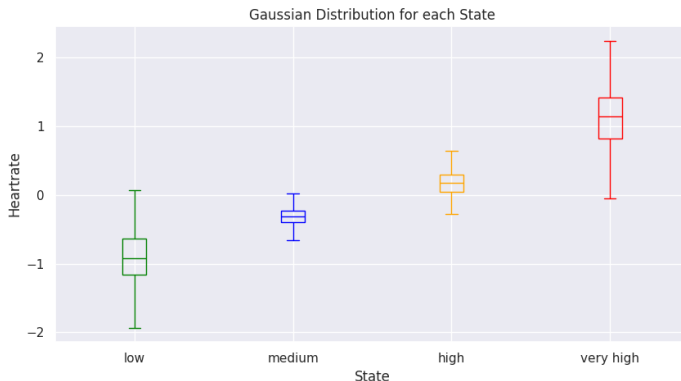
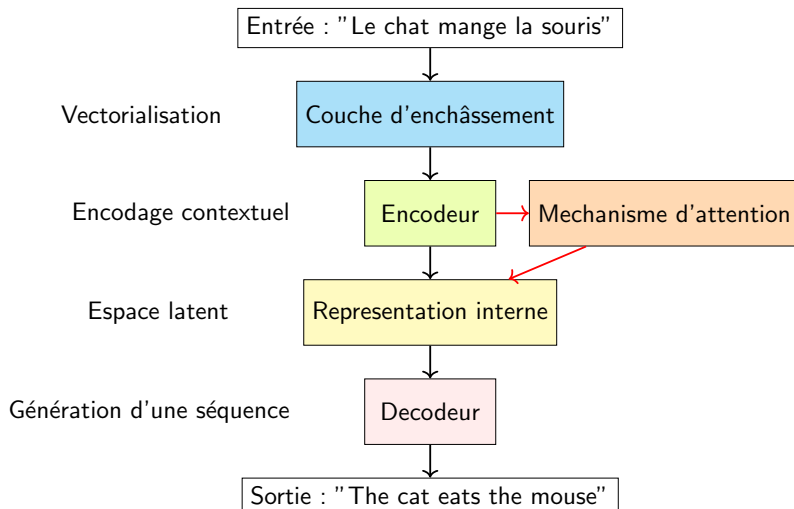


Figure 16 – Box plot de la fréquence cardiaque moyenne par zone



- ▶ 13 cyclistes
- ▶ de juin 2020 à avril 2022

Metric	Points	Nettoyé	Sessions	Nettoyé
Mean	4,558,783	2,771,228	979	378
Std	1,544,788	1,178,575	292	163
Min	2,077,688	843,538	415	85
Max	8,282,392	5,535,237	1,580	696
Total	63,822,957	38,797,199	13,709	5,292

Table 1 – Résumé des données pour les 13 cyclistes