

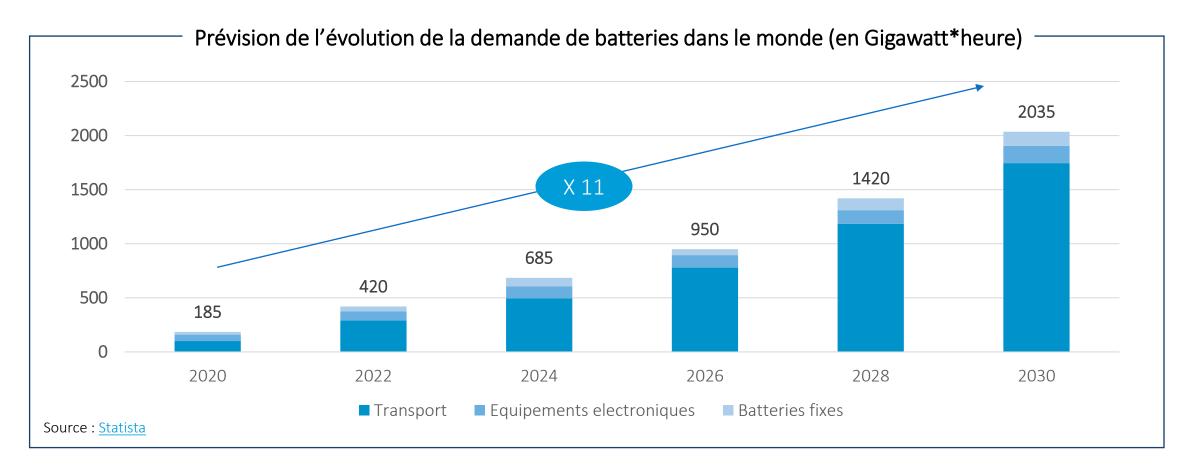
Prédiction de la durée de vie de batteries à l'aide de modèles de machine learning

- 1 Les batteries, un enjeu majeur
- 2 Introduction du projet et des données
- Problématique 1 : classification dès les premiers cycles
- 4 Problématique 2 : prédiction tout au long de la durée de vie
- 5 Ouvertures

- 1 Les batteries, un enjeu majeur
- 2 Introduction du projet et des données
- Problématique 1 : classification dès les premiers cycles
- 4 Problématique 2 : prédiction tout au long de la durée de vie
- 5 Ouvertures

Les besoins en batteries dans le monde vont connaître une croissance sans précédent dans les années à venir, notamment pour soutenir le développement de la mobilité électrique

Entre 2020 et 2030 les besoins en batteries installées vont être multipliés par 11 Ce qui ne préjuge en rien de la capacité du marché à répondre à cette demande.



### Cette augmentation des besoins soulève des enjeux majeurs, notamment environnementaux mais aussi (géo)-stratégiques, notamment pour la maitrise des technologies et de la production





- Utilisation de matériaux rares, éventuellement polluants (notamment le lithium)
- Fin de vie et recyclage



Enjeux économiques

- Contrôle des chaines d'approvisionnement
- Localisation des usines



#### Enjeux technologiques

- Amélioration de la durée de vie
- Limitation de l'utilisation des ressources
- Ruptures technologiques (et pourquoi pas ?)

Dans ce contexte, la prédiction de la durée de vie des batteries est un point crucial, notamment pour accélérer la R&D et améliorer le pilotage opérationnel des parcs de batteries en service.

#### Pourquoi prédire ?

#### Acteurs concernés

Accélérer la R&D

Peut-on déterminer dès les premiers cycles de charge/décharge si une batterie va être performante (afin d'éviter de passer plusieurs semaines/mois à épuiser la batterie pour déterminer si elle l'est ou non) ?

- Les services R&D des fabricants de batteries
- Les chercheurs

2 Anticiper les remplacements et les défaillances

Peut-on prédire précisément combien de cycles de vie reste-t-il à une batterie (quel que soit le nombre de cycles déjà effectués) ?

- Les exploitants de flottes de véhicules électriques (loueurs, services de leasing)
- Les exploitants de parc de stockages (par exemple d'installations non raccordées au réseau)

- 1 Les batteries, un enjeu majeur
- 2 Introduction du projet et des données
- Problématique 1 : classification dès les premiers cycles
- 4 Problématique 2 : prédiction tout au long de la durée de vie
  - 5 Ouvertures

Le projet utilise des données issues d'un article paru dans Nature Energy. Différentes mesures ont été sauvegardées tout au long des cycles de charge/décharge.

#### Vue d'ensemble des données utilisées

nature energy

**Source**: Article de Nature Energy\*1



**135** modules de batteries de voitures électriques



Cycles de charge/décharges complets, jusqu'à épuisement des batteries



**Plusieurs** mesures par batterie à chaque cycles

1\*: <u>Lien vers l'article</u> / Severson *et al*. Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy* volume 4, pages 383–391 (2019). Principaux organismes impliqués : MIT / Toyota Research Institute

#### Description (simplifiée) du protocole

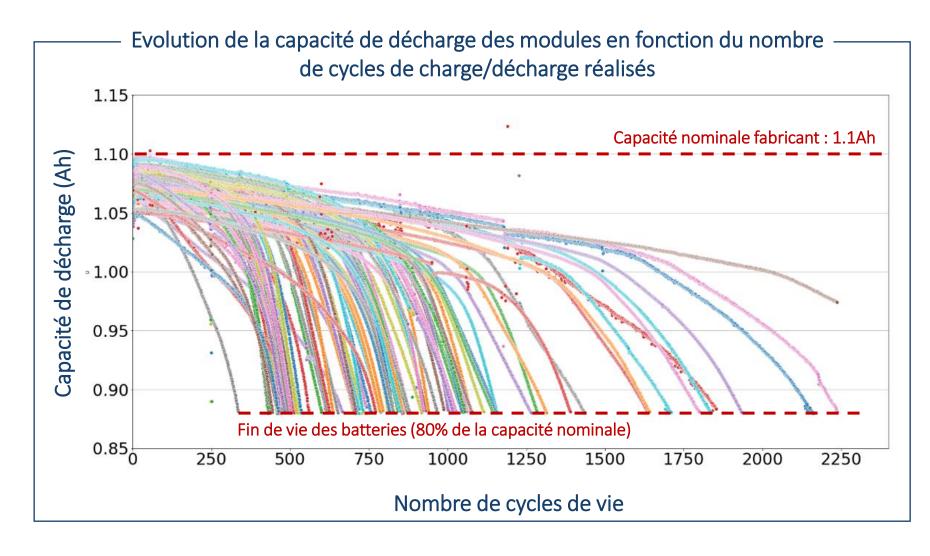
- Les batteries utilisées sont des modules de batterie de voiture (tous identiques) disposant d'une capacité nominale théorique (donnée par le constructeur) de 1,1Ah
- Chaque module est complètement chargé/déchargé (un cycle = une charge + une décharge complètes) jusqu'à ce que sa capacité de décharge tombe en dessous de 80% de sa capacité nominale. A partir de ce moment, le module de batterie est considéré comme ayant atteint sa fin de vie et les mesures sont arrêtées.
- De nombreux paramètres sont mesurés / calculés à chaque cycle :
  - Capacité de charge
  - Capacité de décharge
  - Energie de charge
  - Energie de décharge
  - Rendement de charge
  - Rendement de décharge
  - Efficacité énergétique

- Durée de charge
- Température moyenne
- Température maximale
- Température minimale
- Résistance dc interne

 La mesure principale du jeu de données est la capacité de décharge, sachant que c'est cette dernière qui permet de déterminer la « fin de vie » de la batterie.

Note : Pour plus de détails sur le protocole de mesure, se référer à l'article.

### Une première analyse des données montre une grande variabilité dans la durée de vie des modules



#### Commentaires

- La capacité initiale des modules est un peu inférieure à la valeur théorique donnée par le fabricant (1.07Ah observé en moyenne, contre 1.1Ah en théorie).
- La Durée de vie des modules est très variable (de 363 cycles pour le moins robuste à plus de 2200, avec une moyenne de 860 cycles et une déviation standard de 400 cycles)
- Cette variabilité peut s'expliquer par des variations de qualité des modules (liées à la fabrication), ou aux variations de protocole de charge/décharge (qui pouvait différer d'un module de batterie à l'autre).

# Le projet s'intéresse à deux modèles prédictifs (i) classification binaire selon le nombre de cycles de vie, (ii) régression sur le nombre de cycles de vie restants

1 Accélérer la R&D

2 Anticiper les remplacements et les défaillances

#### Classification binaire

Prédire si un module va durer plus de 550 cycles, en observant uniquement les 5 premiers cycles de vie

#### Calcul de régression

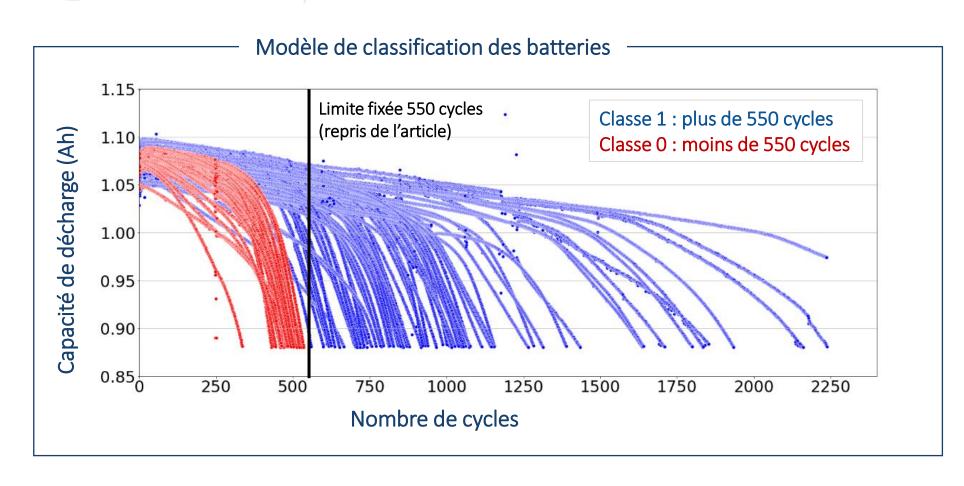
Prédire le nombre de cycles de vie restants, en observant tous les cycles passés.

- 1 Les batteries, un enjeu majeur
- 2 Introduction du projet et des données
- Problématique 1 : classification dès les premiers cycles
- 4 Problématique 2 : prédiction tout au long de la durée de vie
- 5 Ouvertures

### Le premier objectif est de construire un modèle de classification binaire, en prenant comme données d'entrée les 5 premiers cycles.

Accélérer la R&D

Prédire si une batterie va durer plus de 550 cycles, en observant uniquement les 5 premiers cycles de vie

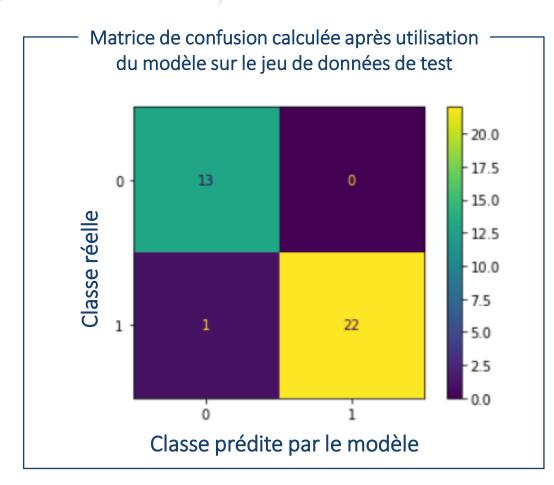


L'objectif du modèle est de prédire si chaque module sera de la classe 1 ou 0, à partir des données des 5 premiers cycles uniquement

### Le modèle de classification est très performant sur le jeu de données de tests. Il classe correctement 35 batteries sur les 36 du jeu de données.

Accélérer la R&D

Prédire si une batterie va durer plus de 550 cycles, en observant uniquement les 5 premiers cycles de vie



Le modèle atteint des performances très satisfaisantes. Il classe correctement 35 des 36 échantillons du jeu de donnés de tests.

Ainsi la performance finale retenue pour le modèle est de 97% (mesure de l' 'accuracy')

A titre de comparaison, la baseline du problème est de 67%. Le modèle est donc largement plus performant que la baseline.

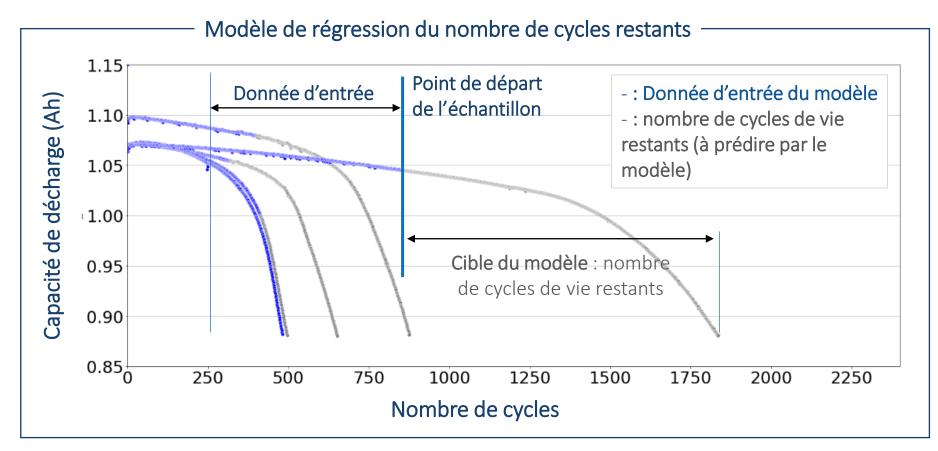
Pour plus de détails techniques sur la méthode de construction des modèles et le choix du meilleur modèle, se référer aux annexes ou au projet sur github (lien en annexe)

- 1 Les batteries, un enjeu majeur
- 2 Introduction du projet et des données
- Problématique 1 : classification dès les premiers cycles
- 4 Problématique 2 : prédiction tout au long de la durée de vie
- 5 Ouvertures

### Le modèle de régression a pour objectif de prédire le nombre de cycles de vie restants, à partir d'un point sélectionné aléatoirement.

2 Anticiper les défaillances

Prédire le nombre de cycles de vie restants, en observant tous les cycles passés.



L'objectif du modèle est de prédire le nombre de cycles de vie de fonctionnement restants pour une batterie (à partir du 'point de départ', choisi de façon arbitraire).

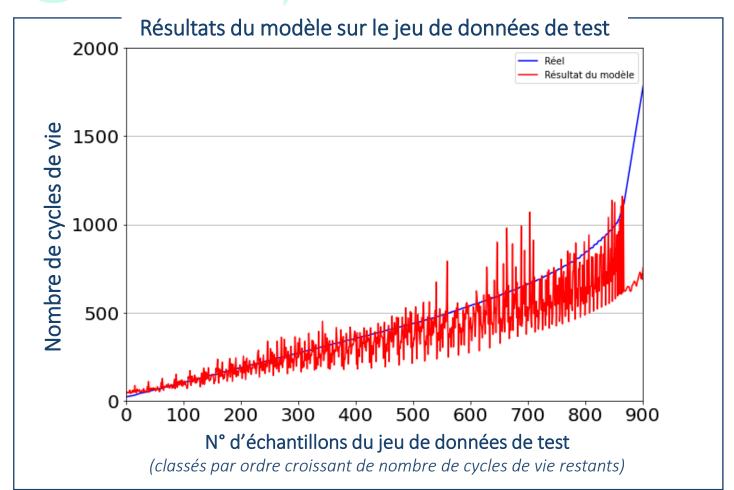
La donnée d'entrée est composée des mesures des cycles de vie précédant le 'point de départ', sur une fenêtre de taille fixe.

Chaque mesure (d'un module) donne lieu à plusieurs échantillons, avec la sélection de plusieurs points de départ

# Le modèle de régression semble cohérent, il est néanmoins plutôt pessimiste et a plus de difficultés lorsqu'il reste beaucoup de cycles de fonctionnement

Anticiper les défaillances

Prédire le nombre de cycles de vie restants, en observant tous les cycles passés.



La tendance du modèle est bonne, plus la durée de vie à prédire est grande, plus le modèle prédit une valeur haute.

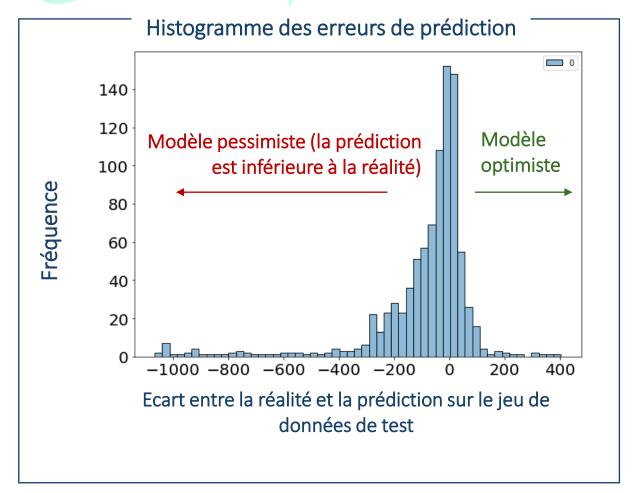
Nous pouvons émettre deux hypothèses en observant ce graphique :

- Le modèle semble prédire plus souvent des valeurs inférieures à la réalité, il est donc plutôt pessimiste
- Par ailleurs, plus le nombre de cycles de vie est grand, plus le modèle les écarts de prédictions semblent importants.

### L'analyse des écarts confirme que le modèle est pessimiste. Cette observation constitue une piste de réflexion pour améliorer le modèle.

2 Anticiper les défaillances

Prédire le nombre de cycles de vie restants, en observant tous les cycles passés.



En moyenne le modèle commet une erreur de -89 cycles, il est donc plutôt pessimiste.

Cette observation pourrait être une piste à explorer pour améliorer le modèle.

### Le modèle apporte une nette amélioration de performance par rapport à la baseline calculée en début de projet.

2 Anticiper les défaillances

\*RMSE: Root Mean Squared Error

Prédire le nombre de cycles de vie restants, en observant tous les cycles passés.

RMSE\* de la baseline : 323

RMSE\* du modèle sur le test set : 212

Le modèle apporte une nette amélioration de la RMSE par rapport à la baseline. La méthode est donc concluante d'un point de vue purement scientifique.

Il n'est néanmoins pas possible de conclure sur la 'qualité' générale de ce résultat, sachant qu'il dépend nécessairement de l'application et des besoins opérationnels.

Pour plus de détails techniques sur la méthode de construction des modèles et le choix du meilleur modèle, se référer aux annexes et au projet sur github (lien en annexe)

- 1 Les batteries, un enjeu majeur
- 2 Introduction du projet et des données
- Problématique 1 : classification dès les premiers cycles
- 4 Problématique 2 : prédiction tout au long de la durée de vie
- 5 Ouvertures

### En conclusion, nous pouvons proposer deux ouvertures, qui permettraient d'étendre le champ d'application des modèles construits dans ce projet

#### Généralisation à d'autres modèles/types de batteries.

➤ Peut-on trouver un modèle capable de faire des prédictions pour différents types de batteries ?

#### Modèle de prédiction en condition réelle

Comment prédire la durée de vie de batteries soumises à des cycles de charge/décharge incomplets ?

En effet les données ont été construites en suivant des cycles de charge/décharge complets. Les batteries à usage opérationnel ont bien souvent des cycles beaucoup plus aléatoires et incomplets. Pensez par exemple à la batterie de votre téléphone.

# Annexes

#### Protocole utilisé pour le modèle de classification



# Accélérer la R&D

Prédire si une batterie va durer plus de 550 cycles, en observant uniquement les 5 premiers cycles de vie

Nettoyage des données Split des données et construction de la baseline

Entrainement de modèles avec différentes combinaisons d'hyperparamètres (avec la méthode de crossvalidation)

Identification du meilleur modèle Validation de la performance

Nombre donné

d'échantillons après nettoyage : 118 Training set = 70% (82 échantillons)

Test set = 30% des données (36 échantillons)

Baseline = 67% (prédiction systématique de la classe majoritaire) **4 combinaisons de données d'entrée** (1, 3, 6 et 9 paramètres utilisés, en gardant systématiquement la capacité de décharge)

Modèle de LogisticRegression. Hyperpamètres testés :

- Solveur
- Normes de régularisation
- Coefficients de régularisation

Modèle de RandomForest. Hyperpamètres testés :

- Nombre d'estimateur
- Min\_sample\_split
- Min sample\_leaf

Identification du meilleur modèle sur la base des résultats des crossvalidation

Le choix du meilleur modèle est basé uniquement sur les résultats obtenus avec le training-set Validation de la performance avec le test-set

#### Protocole utilisé pour le modèle de régression



### Anticiper les défaillances

Prédire le nombre de cycles de vie restants, en observant tous les cycles passés.

Nettoyage des données

Split des données

Création des échantillons pour le réseau RNN

Entrainement des modèles

Identification du meilleur modèle Validation de la performance

Nombre

après

d'échantillons

nettoyage: 118

Training set = 80%

Test set = 20% des données

Mesure de baseline = 323(RMSE) (cf code pour plus de détails sur le calcul de la baseline)

Création des échantillons d'entrée du modèle

Variables utilisées:

- l'écart entre deux points de départ sur deux échantillons consécutifs (appelé 'offset' dans le code)
- La profondeur de la taille de la fenêtre pour les données d'entrée (appelé 'deep' dans le code)

Modèle: réseau RNN

Utilisation d'un 'earlystopping' afin d'éviter la sur-spécialisation du réseau.

Paramètres testés:

- Nombre de couches du réseau
- Nombre de neurones dans chaque unité du réseau
- Ratio de dropout

Identification du meilleur modèle Le choix du meilleur modèle est basé uniquement sur les résultats obtenus avec le training-set

Validation de la performance avec le test-set

# Github du projet