Modeling battery health degradation with uncertainty quantification

Benjamin Larvaron

UNIVERSITÉ DE LORRAINE

École doctorale IAEM Lorraine

Modeling battery health degradation with uncertainty quantification

THÈSE

présentée et soutenue publiquement le 21 mai 2024

pour l'obtention du

Doctorat de l'Université de Lorraine

(mention mathématiques appliquées)

par

Benjamin Larvaron

Composition du jury

Président :	David Brie	Université de Lorraine
Rapporteurs :	Jérémie Bigot	Université de Bordeaux
	Badih Ghattas	Université d'Aix - Marseille
Examinateurs :	Myriam Tami	CentraleSupélec
	Stéphane Raël	Université de Lorraine
	Georges Oppenheim	Université Paris Est
Invités :	Sébastien Benjamin	SAFT
	Louis Verny	TotalEnergies
Directrice :	Marianne Clausel	Université de Lorraine
Co-Encadrant :	Antoine Bertoncello	TotalEnergies

Laboratoire Lorrain de Recherche en Informatique et ses Applications — UMR 7503
Institut Élie Cartan de Nancy

Remerciements

Je remercie tout d'abord mes encadrants, Marianne, Antoine, Sébastien et Georges, ainsi que Clément qui nous a rejoints en chemin, pour ces trois années d'échanges, à la croisée des mondes académiques et industriels mais aussi des statistiques et de la physique. Ce manuscrit est le résultat de cet audacieux mélange, défi scientifique, passionnant et exigeant, qui m'aura tant appris.

Je remercie également Louis pour avoir pris la relève pour la conclusion de la thèse. Les membres de mon comité de suivi de thèse, Rémi Flamary et Stéphane Raël pour leurs précieuses suggestions. Mais aussi l'ensemble des membres de mon jury pour leur temps et leurs retours.

Je remercie les équipes de la plateforme numérique de Total puis TotalEnergies où s'est déroulée la partie industrielle de ma thèse, que ce soit à NanoInnov, au Playground, au NEXT ou même à Coupole. Je vous remercie pour tous ces échanges au quotidien, autant au cours des différentes réunions de présentation de projet ou MATHIAS, qu'autour d'un café voire même d'un Mario Kart. Je remercie également l'équipe BMM de SAFT pour son accueil chaleureux et sa pédagogie avec les néofites des batteries que nous étions. Surtout je remercie l'ensemble des doctorants, notre aîné, Naoufal l'expert optimisation qui nous a montré la voie ; Baptiste, Cheikhna et Amin avec qui je me suis lancé dans la thèse ; Yagnik, Ali, Elie et Nouha l'infiltrée qui ont donné de la vie à ces journées en présentiel (et en dehors) avec toutes ces discussions sur la thèse et ces échanges multiculturels passionnants ; mais aussi Wassil, Khalid et Amar ; tous ces compagnons de route du doctorat.

Seconde facette de ma vie de doctorant, je remercie l'ensemble des équipes et personnels administratifs de l'institut Elie Cartan, pour tous les séminaires, journées et pots partagés. Ici aussi, ma thèse n'aurait pas été ce qu'elle a été sans les doctorants. Je pense aux anciens, Pierre, Christophe et Thomas qui nous ont initiés à la thèse après un début perturbé par la crise du COVID. A mes co-bureaux, le ponctuel et grand dessinateur Victor, ainsi qu'au furtif Nathan T. A Raphaël, Anouk, Valentin, Yann, Pierrick, Clara, Nathan G. ... pour ces discussions prolongées lors des pauses cafés, ces soirées jeux et parties de tennis aux trajectoires plus ou moins contrôlées.

J'ai également une pensée pour mes trois années passées à l'ENSAI, cette école qui m'a donné le goût des statistiques. A Salima El Kolei qui dirigeait la filière génie statistique pour ses conseils, à Camille Saumard rencontrée lors d'un Data Challenge à Pau qui nous a menés plus loin que prévu, mais aussi à Léandre Brault, tuteur durant mon stage de fin d'étude qui m'a initié au travail à l'interface entre domaines. Je pense à l'ensemble de mes camarades de projets ou de moments d'amusements. En particulier, au Bus du RU ; la thèse n'est finalement qu'un escape game géant, domaine dans lequel nous sommes passés maîtres.

Je remercie Paul et Martin pour leur humour intarissable. Romain pour les indémodables Bastilles. Charles, Pax, Aurélien, Baltheus, Karn, la Dyade, pour ces soirées où nous avons réussi (ou presque) à leur donner un faux sentiment de sécurité et ces après-midi/soirées jeux de société. A Constantin, pour nos trajectoires étrangement parrallèles depuis l'entrée à Saint Louis en passant par la thèse.

Enfin, je finirai ces lignes en m'adressant à mes parents et à ma famille. Votre affection et votre soutien constant toutes ces années est ce qui m'a permis d'arriver jusque-là. Je vous dédie ce travail.

Table des matières

List of Figures

Résumé en français

Le changement climatique provoque d'ores et déjà des bouleversements majeurs sur l'environnement. Il a désormais des effets tangibles sur l'amplitude et la fréquence des événements extrêmes avec des conséquences sur les populations et la biodiversité [1]. Les pratiques actuelles sont encore insuffisantes pour endiguer le réchauffement. Des mesures importantes doivent être prises dans les années à venir afin d'éviter des conséquences irréversibles sur la sécurité de larges populations.

Ces mesures vont imposer la transformation de tous les secteurs de l'économie. En particulier, le secteur des transports va devoir remplacer le parc de véhicules thermiques par des alternatives moins émettrices de CO_2 comme les voitures électriques. La production d'énergie va devoir radicalement évoluer passant d'une large majorité d'énergies fossiles à d'autres solutions, nucléaires ou renouvelables. Le développement des voitures électriques et des énergies renouvelables soulève un besoin crucial de batteries électriques. Elles permettent d'alimenter les véhicules et d'améliorer la régulation des réseaux électriques face à des énergies intermittentes et distribuées. Du fait de sa haute densité d'énergie et de puissance, la solution plébiscitée actuellement pour ces problèmes est la technologie Lithium-ion.

En particulier cette thèse s'intéresse à la modélisation de la dégradation des batteries Lithium-ion. La connaissance de cette dégradation est un enjeu majeur autant pour les manufacturiers que pour les utilisateurs. Elle leur permet de prévoir la performance du modèle au cours du temps et d'anticiper leur remplacement. Un point important dans cette modélisation est l'inclusion d'une quantification des incertitudes. Une simple prévision de la valeur la plus probable peut en effet être trompeuse dans des contextes avec peu d'informations ou d'importantes variabilités. Cette problématique sera donc prise en compte tout au long de cette thèse.

La dégradation des batteries électriques varie selon la chimie et la configuration considérée. L'électrode positive est le paramètre principal de la chimie d'une batterie avec de nombreuses variantes des technologies LCO, NMC, NCA ou LFP. Le choix de l'électrode négative et de l'électrolyte jouent aussi un rôle important. La configuration quant à elle, comprend de la dimension des batteries et de leurs électrodes ainsi que leur design (cylindrique, prismatique, « pochette » ou « pile bouton »). Une fois le type de batterie fixé, le vieillissement dépend encore des conditions d'utilisation. La température ambiante ainsi que les courants de charge et de décharge jouent un rôle important sur la vitesse de dégradation.

En fonction de ces éléments, les batteries sont sujettes à de nombreux mécanismes de vieillissement [2]. Certains réduisent la quantité de lithium passant d'une électrode à l'autre, d'autres jouent sur la quantité pouvant être stockée dans les électrodes. Les mécanismes les plus souvent

observés sont la formation d'une couche de SEI (Solid Eletrolyte Interphase) sur l'anode et le "lithium plating". Si le premier peut avoir l'effet bénéfique de ralentir le vieillissement, le second peut provoquer des chutes soudaines de performance.

L'évolution de la performance des batteries est quantifiée à l'aide de deux principaux indicateurs : la capacité, liée à l'énergie que la batterie peut stocker, et la résistance interne, liée à la puissance que la batterie peut délivrer. Des expériences sont menées en laboratoire pour mesurer l'évolution de ses grandeurs tout au long de la vie des batteries. Ces tests sont souvent réalisés sous différentes conditions expérimentales, variant la température ambiante ou la politique de charge/décharge. Il peut s'agir de tests calendaires où les batteries sont laissées au repos ou bien en cyclage avec des phases répétées de charge et de décharge. De plus en plus de jeux de données issus de ces tests mis à disposition publiquement afin de favoriser la recherche sur la modélisation du vieillissement.

L'objectif global est d'utiliser ces données expérimentales afin de prédire une courbe de dégradation de performances en fonction de caractéristiques et de conditions d'utilisation des batteries. Idéalement on souhaiterait obtenir des prévisions même pour des conditions et caractéristiques de batteries non testées. Cet objectif étant trop large au vu des données disponibles, nous avons considéré dans cette thèse des sous-problèmes plus atteignables. Tout d'abord, nous avons modélisé l'évolution temporelle de la performance à une condition expérimentale de référence pour un unique modèle de batterie. Cela inclut l'interpolation temporelle, dans la plage de temps observée et le forecasting de l'évolution future. Ensuite nous avons élargi ce problème afin de pouvoir prédire l'évolution de la performance en fonction d'une condition expérimentale.

Dans chaque cas, plusieurs sources d'incertitudes sont à prendre en compte. Celles-ci se décomposent en deux types, l'incertitude aléatoire liée à la variabilité intrinsèque du phénomène et l'incertitude épistémique reflétant notre manque de connaissances à cause de la faible quantité de données. L'incertitude aléatoire peut avoir de multiples causes, la précision des capteurs, une variabilité dans la production des batteries ou dans la réalisation des expériences. Nous nous concentrerons en particulier sur la variabilité inter-éléments (cell-to-cell variations) [3], quantifiant les différences de performances entre des batteries supposément identiques et testées dans les mêmes conditions. En parallèle l'incertitude épistémique est une question importante pour les prévisions dans des domaines peu ou pas observés, notamment le forecasting et l'interpolation de conditions expérimentales.

De nombreuses approches existent en pratique pour modéliser la dégradation des batteries. Ces approches sont souvent divisées entre les approches "model-based" et celles "data-driven" [4]. Les approches "model-based" utilisent un modèle mathématique de la dégradation fourni à priori, généralement à partir de connaissances physiques. A l'inverse, les approches "data-driven" apprennent directement le modèle à partir des données, sans utiliser la connaissance physique. Dans cette thèse nous sommes parties d'un point de vue data-driven, cependant le manque de données en forecasting et dans l'interpolation de conditions expérimentales nous a incité à compléter cette approche avec de la connaissance physique.

Nous avons tout d'abord considéré le problème de la dégradation temporelle pour un unique modèle de batterie et à une condition expérimentale de référence. Nous avons choisi d'aborder ce problème à l'aide de méthodes par processus gaussiens [5]. Ces méthodes possèdent la particularité de permettre une grande flexibilité dans l'estimation du modèle, propre aux approches machine learning tout en fournissant naturellement une quantification des incertitudes. Ces avantages ont rendu populaires les approches basées sur les processus gaussiens dans la communauté batterie [6],

et ce d'autant plus que ces méthodes sont particulièrement adaptées au contexte des séries temporelles avec peu d'observations, ce qui est souvent le cas pour des données expérimentales.

Les processus gaussiens sont une extension des distributions gaussiennes au cadre fonctionnel. Ce sont des outils utiles pour estimer et quantifier les incertitudes sur des données fonctionnelles comme le sont nos courbes de dégradation de la capacité des batteries. Un processus gaussien est paramétrisé par une fonction moyenne et d'une fonction de covariance, le noyau. Le choix de ce dernier est particulièrement important, définissant plusieurs propriétés comme la régularité, la stationnarité et l'isotropie des prévisions. Il donne également la structure des incertitudes associées.

La régression par processus gaussien, introduite par Rasmussen et Williams [7], est la méthode à base de processus gaussien la plus connue en machine learning. Utilisant un formalisme bayésien elle pose un prior de processus gaussien sur la fonction de tendance à estimer. La vraisemblance utilisée étant également gaussienne, une formule explicite de la loi a posteriori et donc des prévisions peut être calculée. On obtient ainsi un modèle flexible, intégrant directement les incertitudes et calculable de manière exacte.

Des difficultés liées au temps de calcul et à la volonté d'utiliser des modèles plus complexes ont cependant conduit au développement d'approches d'approximation. Les plus connues sont les simulations par méthode de Monte-Carlo, l'inférence variationnelle, la méthode de Laplace ou l' "Expectation Propagation" (EP). Elles peuvent permettre l'utilisation de grands volumes de données ou de modèles complexes avec des vraisemblances non gaussiennes pouvant dépendre de plusieurs processus gaussiens. Dans notre cas nous nous sommes particulièrement intéressés au cadre de la régression par processus gaussiens chaînés [8] reposant sur l'inférence variationnelle.

Pour modéliser la courbe de dégradation de l'état de santé, nous sommes donc partis de l'état de l'art en utilisant la régression par processus gaussien. Nous avons proposé un choix spécifique de design de noyau permettant de décomposer les différentes sources d'incertitudes, notamment la variabilité inter-éléments. Cela nous a fourni une estimation précise de la tendance de dégradation et une décomposition des incertitudes, cependant nous avons pu observer que l'hypothèse classique de stationnarité restreignait fortement la modélisation des incertitudes, empêchant la prise en compte de l'évolution de la variabilité inter-éléments au cours du temps.

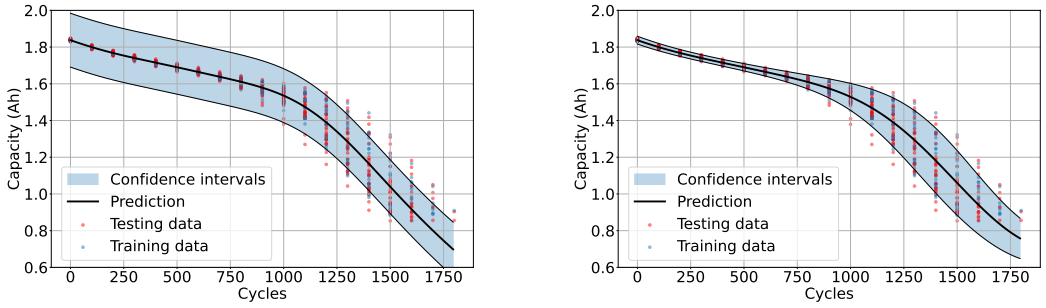


FIGURE 1 – Comparaison de la prévision de la courbe de dégradation (a) par regression par processus gaussien (b) par processus gaussiens chaînés

L'utilisation de la régression par processus gaussiens chaînés nous a permis de dépasser cette contrainte. Par l'introduction d'une vraisemblance adaptée, avec un processus gaussien dédié à la modélisation de l'évolution de variabilité inter-éléments, cette méthode nous a permis d'obtenir une modélisation non-paramétrique fine de cette dernière. Cela a nettement amélioré la modélisation des incertitudes par rapport à la régression par processus gaussien classique comme permet de le voir la figure Fig 1.

Les approches par processus gaussien sont réputées pour les questions d'interpolation. Cependant elles possèdent souvent des limites en ce qui concerne l'extrapolation. Dans notre cas, bien que l'interpolation temporelle fournisse de bons résultats, nous avons constaté des difficultés à faire du forecasting, des prévisions pour des cycles futures non observés. Les prévisions avaient en effet des comportements incohérents d'un point de vue physique avec de très larges incertitudes. Pour pallier ces lacunes nous avons proposé d'intégrer des contraintes [9] traduisant la connaissance physique sur la dégradation. Nous nous sommes concentrés sur des contraintes liées à la dérivé de la tendance de dégradation, monotone et sujet à une accélération à mesure que des mécanismes comme le Lithium plating prennent l'ascendant sur la dégradation.

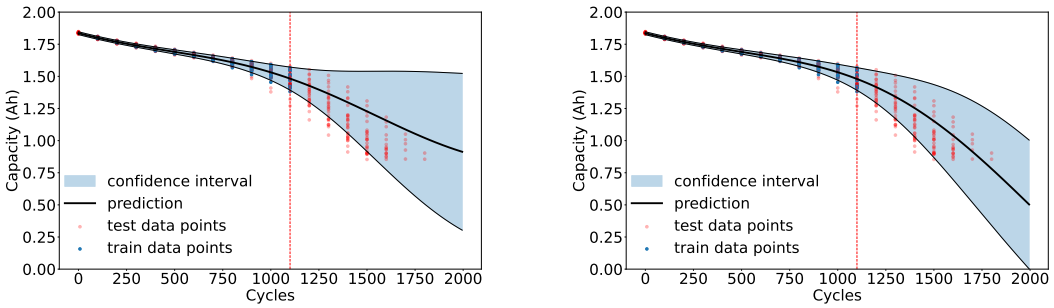


FIGURE 2 – Comparaison des prévisions en forecasting de la regression par processus gaussien (a) sans et (b) avec contrainte

Nous avons donc proposé une adaptation du modèle de régression par processus gaussiens chaînés permettant d'intégrer ces contraintes. Nous avons ainsi pu obtenir des prévisions davantage en accord avec les connaissances physiques et avec des performances nettement améliorées, illustrées Fig. 2. Cela permet d'améliorer la robustesse des modèles et d'obtenir des prévisions à plus long terme. Pour les expérimentateurs cela ouvre la possibilité de réduire le temps de test des batteries et donc les coûts associés.

La modélisation de la dégradation des batteries à une condition expérimentale de référence fournit un premier moyen d'évaluer la performance d'un modèle. Cependant, ces performances dépendent fortement des conditions expérimentales, notamment de la température. La performance d'un modèle de batterie peut être supérieure à celle d'un autre modèle pour un type de condition mais être inférieur dans d'autres situations. Il est donc nécessaire de prendre en compte l'effet des conditions expérimentales dans la modélisation de la dégradation. La difficulté était qu'en pratique peu de conditions sont observées, avec potentiellement d'importantes variations de performances entre elles. Il faut donc une méthode suffisamment robuste pour interpoler entre les conditions observées.

L'approche la plus directe pour répondre à cette problématique correspondait à adapter l'approche précédente par processus gaussiens. On peut en effet étendre la modélisation précédente en ajoutant les facteurs expérimentaux comme une entrée supplémentaire. Les contraintes sur les dérivées peuvent également être adaptées, par exemple pour contraindre la performance à décroître à mesure que la température ambiante augmente. Cette approche avait déjà été utilisée dans la littérature et fournit des résultats intéressants dans des cas où les dégradations observées restent relativement similaires. De plus elle demande peu d'efforts supplémentaires par rapport à la modélisation à une condition de référence. Cependant, nous avons observé les limites de cette approche dans des cas plus complexes où les dégradations observées sont nettement différentes. Dans ce cas, l'utilisation d'un modèle unique est limitée car elle contraint à partager des hyperparmètres, ce qui rend le modèle moins performant que des modèles ajustés individuellement sur chaque condition.

Pour éviter ce problème nous avons proposé une autre approche en deux temps. Tout d'abord, le modèle par processus gaussien est appliqué à chaque condition observée, permettant d'avoir une modélisation précise dans chaque cas. Ensuite, ces prévisions sont utilisées comme données d'entrée pour modéliser l'effet des conditions expérimentales à l'aide d'une autre modélisation. La difficulté de cette seconde étape est qu'elle consiste à apprendre un modèle de regression avec des sorties complexes, dans notre cas des distributions de probabilité gaussiennes multivariées.

Plusieurs stratégies existent pour apprendre un tel model. La plus répandue aujourd'hui consiste à réaliser une étape préliminaire d'"embedding" (plongement) faisant correspondre à chaque donnée complexe une valeur vectorielle plus standard sur laquelle les méthodes classiques peuvent être utilisées. Cette approche a rencontré de nombreux succès notamment dans le domaine du traitement du langage ou des images. Cependant, elle ne permet pas d'utiliser toutes les propriétés des objets concernés. Nous avons donc privilégié une autre approche adaptant directement les méthodes de modélisation aux objets concernés. Dans notre cas où les données sont des distributions de probabilités, la théorie du transport optimal [10, 11] est particulièrement indiquée. Elle fournit un cadre mathématique solide pour traiter ce type d'objet et a connu de nombreux succès ces dernières années sur des problématiques appliquées.

La théorie du transport optimal vise à minimiser le coût de transport d'une distribution vers une autre. Ce transport peut être déterministe, il s'agit du problème de Monge, ou probabiliste, c'est le problème de Kantorovitch. Cependant, dans ces versions, le problème fournit seulement l'allocation optimale de chaque masse élémentaire de probabilité de la première distribution vers la seconde. La formulation continue du transport optimal élargi le problème en optimisant toute la trajectoire d'une distribution de probabilité vers la seconde, le chemin d'interpolation entre les deux distributions. Dans le cas de deux distributions il s'agit de l'interpolation de McCann [12] qui se généralise à plusieurs distributions grâce aux barycentres de Wasserstein [13]. Tous ces éléments peuvent être calculés dans le cas de distributions gaussiennes. De plus la recherche en transport optimal est en plein essor et de nombreuses variantes existent, permettant de réduire les coûts de calcul, d'améliorer la stabilité des algorithmes d'optimisation ou de prendre en compte des distributions de structure différentes.

Ces développements nous donnent les outils pour construire un prédicteur. Les barycentres de Wasserstein nous fournissent une façon rigoureuse de se déplacer dans l'espace des distributions de probabilité en interpolant celles qui ont été observées. Ces barycentres dépendent de coordonnées qui définissent à quel point le barycentre est proche de chacune des distributions utilisées pour le calculer. L'idée consiste à choisir judicieusement ces coordonnées en fonction des conditions expérimentales. Cela conduit à un barycentre conditionnel, généralisation de l'espérance conditionnelle,

utilisée classiquement pour les prévisions sur données vectorielles.

Deux méthodes sont ensuite présentées pour sélectionner les coordonnées du barycentre conditionnel. La première généralise l'approche classique en machine learning [14]. Elle propose un modèle paramétrique sur les coordonnées et en optimisant ses paramètres par minimisation du risque empirique sur les données d'apprentissage. Dans ce contexte, la fonction de perte utilisée est la distance de Wasserstein. Dans notre cas, le nombre de conditions étant faible, nous avons choisi d'utiliser sur les poids un modèle empirique physique permettant d'intégrer l'accélération de la dégradation avec la température.

Cela fournit une première approche de prévision mais présente quelques inconvénients. L'utilisation d'un modèle physique sur les coordonnées devient compliquée à mesure que l'on utilise davantage de distributions pour calculer le barycentre. Cette approche nécessite également qu'une partie des distributions observées servent uniquement à l'ajustement des paramètres du modèle utilisé sur les coordonnées, ce qui cause une perte d'informations importante dans notre cas.

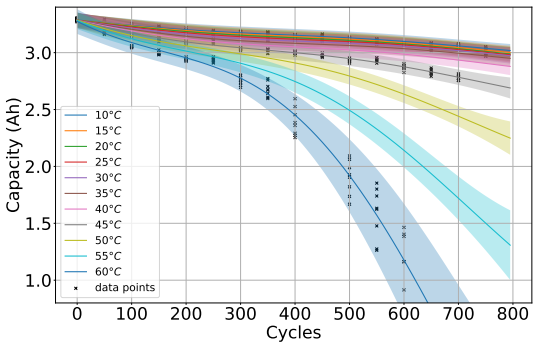


FIGURE 3 – Interpolation des courbes de dégradation entre les températures observées grâce à la regression de Fréchet

Nous avons ainsi proposé l'utilisation d'une autre approche, la régression de Fréchet [15]. Celle-ci est une généralisation de la régression linéaire et possède l'avantage de fournir une expression explicite des coordonnées. Dans sa formulation initiale la régression de Fréchet est trop restrictive, son caractère linéaire ne permet pas de modéliser l'accélération de la dégradation selon la température. Cependant, elle peut être adaptée en une version polynomiale nettement plus flexible. Sa version surparamétrée notamment permet de se ramener au modèle physique empirique précédent tout en exploitant complètement toutes les dégradations observées. Cela fournis une interpolation intéressante prenant directement en compte les incertitudes comme illustré Fig. 3 avec l'effet de la température ambiante. Cette méthode s'adapte également au cas de plusieurs facteurs expérimentaux permettant de prendre en compte leurs interactions.

Nous soulignons tout de même que les barycentres conditionnels de Wasserstein restent une méthode récente et que plusieurs problèmes théoriques devront être traités à l'avenir pour assurer leur utilisation. Dans le cas gaussien, les barycentres de distributions dégénérées restent un problème ouvert. De plus les barycentres sont principalement pensés pour interpoler les distributions, cependant l'extrapolation est également un problème fondamental d'un point de vue appliqué, une généralisation de cet objet serait donc d'un grand intérêt.

Cette thèse aura donc traité de la question de la dégradation des batteries électriques en abordant plusieurs problématiques d'intérêt statistiques, avec la quantification des incertitudes, la combinaison d'approches data driven avec des connaissances physiques (physics informed machine learning), ainsi que de la régression sur objets complexes (structured regression).

Le problème de la modélisation de la dégradation à une condition de référence semble actuellement bien avancé dans la littérature. En revanche, beaucoup d'efforts restent à faire pour la prise en compte de l'effet des facteurs expérimentaux. Cela est principalement dû au manque encore criant de données expérimentales d'autant plus marqué si l'on souhaite considérer des conditions complexes, plus proches des conditions d'utilisation réelles. La modélisation de l'effet des caractéristiques de batteries reste à ce stade un problème ouvert la plupart des expériences étant fait pour un type de batterie fixé. Enfin de nombreux problèmes connexes nécessiteront une attention particulière à l'avenir, comme le passage d'une simple cellule à un pack de batterie, voire un conteneur de stockage, l'anticipation des événenents extrêmes pour améliorer la sécurité ou les méthodes en ligne pour optimiser l'utilisation d'une batterie en temps réel.

Ce manuscrit se décompose en trois principales parties :

Le chapitre 1 décrit en détails le fonctionnement des batteries Lithium-ion et leurs principales caractéristiques. Il présente également les principaux modes de dégradation ainsi que la mesure de celle-ci. Le problème de la modélisation de la dégradation est formalisé, en mettant l'accent sur la quantification des incertitudes. Enfin, les différentes approches de modélisation présentent dans la littérature seront passés en revue.

Le chapitre 2 traite de la modélisation de la dégradation à une condition expérimentale de référence par des méthodes de processus gaussiens. Après une présentation générale de ces méthodes, un premier modèle de régression par processus gaussien avec une décomposition des incertitudes est présenté. Cependant, face à ses difficultés à modéliser l'évolution de l'incertitude inter-éléments, nous proposerons une amélioration du modèle précédent à l'aide de la méthode plus générale de la regression par processus gaussiens chaînés. Finalement, nous nous intéresserons en particulier au problème du forecasting, où l'inclusion de connaissance physique permettera de corriger l'approche purement data-driven.

Le chapitre 3 étend la problématique précédente en prenant en compte l'influence des conditions expérimentales. Dans ce cas, les méthodes par processus gaussien, modélisant simultanément l'évolution temporelle et l'effet des conditions ne fournissent pas de résultats satisfaisants. Nous proposons à la place une approche en deux étapes. L'évolution temporelle est modélisée par les méthodes précédentes par processus gaussien, tandis que l'effet des facteurs expérimentaux est modélisé grâce à la théorie du transport optimal avec les barycentres conditionnels de Wasserstein. Deux modèles de barycentres conditionnels seront présentés. Tous deux reposent sur la connaissance physique de la dégradation permettant une modélisation robuste de l'effet des facteurs.

General introduction

1 Context

In its most recent report for policymakers, the Intergovernmental Panel on Climate Change (IPCC) stated that "Human activities, principally through emissions of greenhouse gases, have unequivocally caused global warming, with global surface temperature reaching 1.1°C above 1850-1900 in 2011-2020" [16]. Global warming is no longer a question of speculation ; it has tangible effects, including the rise of sea level, the increase of extreme events, heat waves, droughts, and precipitation, with strong consequences on population and biodiversity.

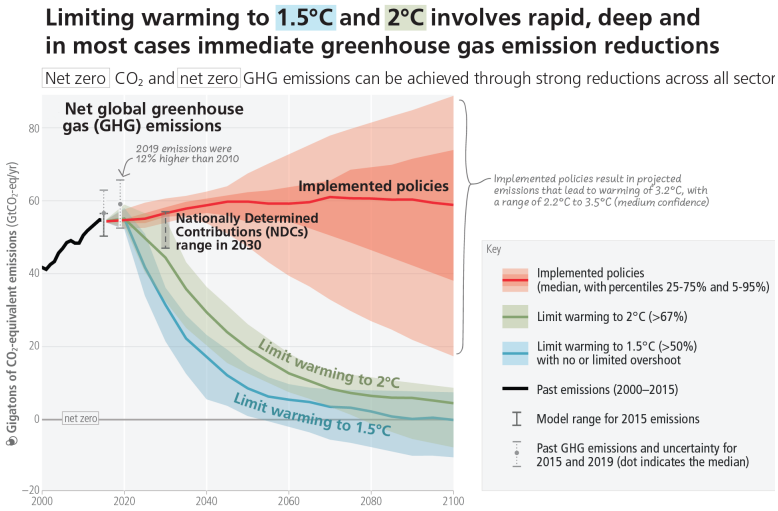


FIGURE 4 – Forecast of greenhouse gas emission, from [16]

Moreover, the IPCC anticipates a further increase of global warming depending on society's reaction. The special report of IPCC of 2018 [17] already detailed the consequences of a 1.5 ° C increase. However, Fig 4 taken from their last report [16] illustrates the greenhouse gas emissions associated with currently implemented policies would lead to a much higher warming of 3.2 ° C. The report details the consequences of such warming. The heat-humidity risk would directly threaten a large part of the world population, the food and water security of large areas would be threatened, and extreme events and amplitude and frequency would lead to major global impacts.

Consequently, important decisions to reduce rapidly greenhouse gas emissions are required.

In response to these incoming consequences, world leaders decided of different objectives to limit carbon emissions from human activities. Notably, the Paris Agreement of 2015, following COP 21 fixed the target of limiting global warming to 1.5°C. Such a goal requires a fast pace reduction of greenhouse emissions. Given that the current energy mix is still dominated by carbonated energy, coal, oil, and natural gas, the efforts must concentrate on the increase of alternative cleaner energies, nuclear, hydroelectricity, biomass but also wind, solar, and geothermal energies. In addition to this substitution, the issue of limiting energy consumption should also be tackled. These efforts have consequences in all the main economic sectors, in the industry, transports, residential, agriculture, and services.

The transition away from fossil energies leads to a rapid growth of electrical energy usage. However, a critical challenge associated with electrical energy relies on its storage which cannot be tackled as easily. Batteries are currently seen as one of the main tools to handle this problem. Notably, its use is rising in the transport sector, driven by the development of electric vehicles. Additionally, batteries play a crucial role in integrating renewable energy sources into the grid through stationary storage solutions. Because of its high energy and power density, the Lithium-ion battery technology concentrates most of the interest in implementing these transformations.

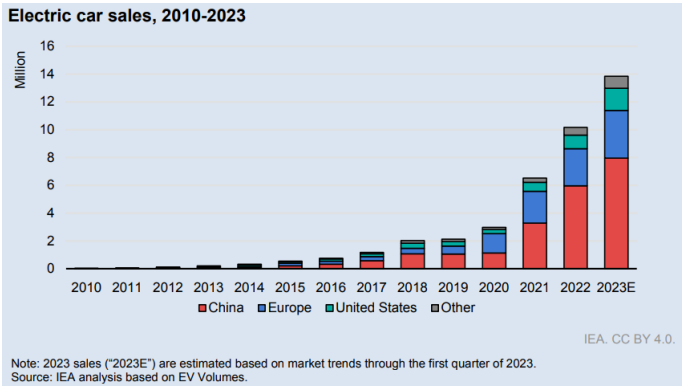


FIGURE 5 – Rise of electric vehicles sales between 2010 and 2023, from [1]

According to the International Energy Agency (IEA), the transportation sector globally accounts for approximately one-fourth of CO2 emissions, and this proportion is even higher in the United States and the European Union. As a response, many developed countries decided to impose regulations on internal combustion engine vehicles. Notably, the European Union decided the interdiction of new thermal vehicle sales after 2035, and many developed countries took similar decisions. This imposes a race for car manufacturers to replace their car production in the next years.

In addition to these regulations, governments have implemented key policies to expedite the growth of alternative vehicles. These measures span from funding research to offering purchase incentives to consumers. Electric vehicles benefit strongly from these policies [1] with currently an exponential growth of new vehicles sales see Fig. 5, principally in the US, Europe, and China, but are also emerging in other markets. Thus, from a negligible market share in 2010, electric vehicles has reached in 2023 a 18% share of the global sales market, with a sustained anticipated growth for the next years.

FIGURE 6 – An example of Battery energy storage systems site, from [19]

The rise of electrical vehicles has led to a boom of the battery production, in particular Lithium-ion batteries. Batteries represent the key technological and business bottleneck of EVs. Despite a gradual decrease in cost, lithium batteries still made up a third of the vehicle's price in 2020 [18], making EVs more expensive, on average than thermal vehicles. This price tends to decrease with the mass production trend and improvement of mineral supply chains. Moreover, it poses a technical challenge to produce electrical vehicles competitive with thermal vehicles in terms of performance, power availability, autonomy, and speed of recharge.

Concerning the grid, balancing electricity production and consumption is an old problem. The demand for electricity fluctuates throughout the day, with higher usage during the day than at night, during the week compared to weekends, and throughout the year depending on the weather. These variabilities could be handled thanks to the use of relatively controllable electricity productions means, thermal power stations or nuclear reactors. However, the decarbonisation of the economy induces significant changes in both electricity demand and supply.

Regarding the supply, the use of thermal power stations, relying on coal, gas or fuel combustion, shall be reduced significantly. In place, renewable energy, notably solar and wind energies, has to be increased. However, these energies create an important temporal variability in electricity production. Solar energy is obtained only during the day and varies with cloud cover. Wind turbines production depends on wind speed and direction which can face important variations. These energies also face an important spatial variability. While a few thermal power stations or nuclear reactors could be used, each producing a significant amount of energy, renewable energy production is much more decentralized with many small production sites. So overall, the control of energy production and grid stabilization is much more complex than before. Besides, the energy demand also faces important transformations. Notably, the energy transition in the transport sector, described above induces an important shift from fossil energy sources to electrical energy, causing an important increase and alteration of the demand.

In order to avoid blackouts or power failures due to overproduction or underproduction, the solutions are either to trade energy with neighboring markets or to store it. However, purchasing energy incurs significant additional costs. Also, the storage demand is increasing. Traditionally, a vast majority of energy storage was done thanks to pumped hydroelectric systems [20]. However, this method is limited regarding the new needs. Besides the important

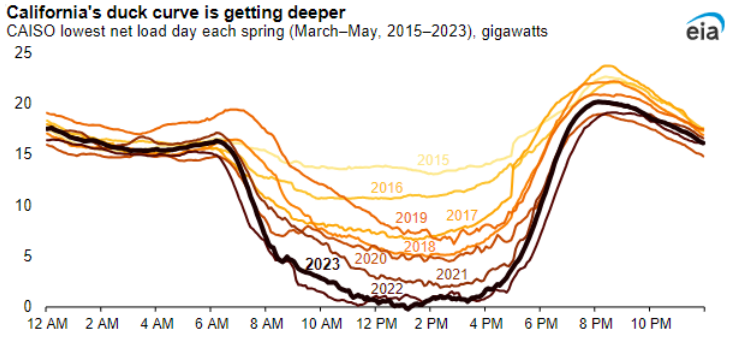


FIGURE 7 – Evolution of the Duck curve in California, from [23]

costs associated with the construction of new facilities, hydroelectric systems are not flexible, requiring important infrastructures, they typically provide high power, corresponding to the scale of a region with important delays. Among alternatives (such as compress air energy storage, flywheels, or capacitors), electrical batteries and mainly Lithium-ion batteries are becoming the most suited solutions to the new issues associated with renewable energies production [21, 22]. Their ease of installation, modularity, and short response, make them suited for more distributed installations at more specific locations, with much more precise control from operators. In practice, we speak of battery energy storage systems (BESS), based on containers full of Lithium-ion batteries, where the number of containers can be adapted to the storage need. Fig. 6 shows an example of such sites connected to the grid.

A clear example of the imbalance between production and demand can be seen in California, with the famous duck curve, illustrated in Fig. 7. In this state, electricity is importantly produced by solar panels which reach a peak production at midday. At this moment it compensates for the global load causing even overproduction risks. However, the production drops with sunset needing a sudden increase of 20 gigawatt production from other energy sources. To mitigate this sudden increase, causing important stress on the grid, and avoiding the use of carbonized energy production or production curtailment, a massive storage need is required, and California developed massively BESS usage. This phenomenon is not specific to California. The early development of solar energy in this state anticipates future disequilibrium of other American states and developed countries.

Additionally, thanks to their multiple advantages, lithium-ion batteries are used for other applications supporting the grid [21]. Lithium-ion batteries are of important use for grid stabilization in frequency, requiring high power spikes with a fast response. Lithium-ion batteries permit also the development of microgrids, autonomous from the global grid, often connected to some renewable energy productions. Finally, many projects look for solutions to optimize the balance between production and storage with smart-grids [24] possibly integrating electric vehicle batteries as a way to stabilize the grid [25].

2 Outline and contributions

This work is the result of an industrial thesis (CIFRE thesis) with TotalEnergies. As a major actor in the energy domain, it is fully involved in the aforementioned context. To respect new regulations and to address new customer demands TotalEnergies is developing its electricity energy production and storage capacities for both the grid and the electric vehicles. More precisely this work was done in collaboration with SAFT, an historical battery manufacturer and subsidiary of TotalEnergies since 2016. SAFT's R&D is responsible for the production of new competitive battery models for different applications.

This thesis focuses on the modeling of the degradation of new battery chemistries. This is a key element of their performance, determining the number of batteries to be used and the maintenance costs. Historically this problem was tackled by SAFT thanks to physic-based models. This thesis will mainly consider data-driven methods as a complementary set of approaches winning popularity in recent years. The main goal was to study in detail the uncertainty quantification associated with battery degradation. The idea was to go beyond the classical mean prediction methods with the prediction of confidence intervals. The ambition is to permit the assessment of the financial risk related to performance guarantees.

Apart from this general introduction and ending conclusion, this thesis is composed of three main chapters

Chapter 1 will present in detail the working principle of batteries and their main characteristics. It will also shed light on the mechanisms of degradation and how they are measured. After an explicit formalization of the degradation modeling goal and a focus on the uncertainty quantification we will present an overview of the existing modeling approaches.

Chapter 2 focuses particularly on the degradation modeling under a reference condition using Gaussian processes methods. After a general presentation of these methods a particular adaptation of the Gaussian process regression to obtain with a detailed uncertainty quantification will be proposed. Then observing the limits of the above framework a more general Chained Gaussian process model is presented, enabling us to model precisely the time-evolution of the uncertainty. Finally, in order to improve the forecasting ability of the models we will present the inclusion of physical prior knowledge to complete the data-driven approach.

Chapter 3 expands the scope of the problem to include the effects of operating conditions. Given the difficulties of Gaussian process-based methods to simultaneously model time and condition effects, we propose a two-step approach. Time modeling is still done through previous GP-based methods, but experimental factors are modeled through optimal transport theory with conditional Wasserstein barycenters. Two dedicated conditional Wasserstein barycenters models, exploiting the available physical knowledge, will be derived permitting a robust modeling of experimental factors.

The contributions of this thesis have been patented by TotalEnergies. It also led to the three following publications on which chapters 2 and 3 rely extensively :

— Larvaron, B., Clausel, M., Bertoncello, A., Benjamin, S., & Oppenheim, G. (2023). Chained Gaussian processes to estimate battery health degradation with uncertainties. Journal of Energy Storage, 67, 107443.

— Larvaron, B., Clausel, M., Bertoncello, A., Benjamin, S., & Oppenheim, G. (2023). Chained Gaussian processes with derivative information to forecast battery health degradation. Journal of Energy Storage, 65, 107180.

— Larvaron, B., Clausel, M., Bertoncello, A., Benjamin, S., Oppenheim, G., & Bertin, C. (2024). Conditional Wasserstein barycenters to predict battery health degradation at unobserved experimental conditions. Journal of Energy Storage, 78, 110015.

Chapter 1

Modeling battery health degradation

Contents

1.1 Introduction

The word "battery" is used in several meanings in practice. Rigorously, a battery is a combination of devices storing chemical energy and converting it into electrical energy [26]. For example in the automotive context, the basic elements are called battery cells. These cells are combined to form a module, which are combined again to form a battery pack. In this thesis, we will concentrate on the fondamental level of a single cell. The electrical and mechanical combination of battery cells is a key element of the performance of the final battery pack, and estimating pack performance based on the one of cells is an important challenge [27], yet we will not cover this issue here and in the following, we will use the terms “cell” and “battery” indifferently.

Batteries’ history started at least two centuries ago [26]. After a controversy with Galvani, who claimed the discovery of "animal electricity", the first battery is attributed to Volta in 1800. The practical use of the Voltaïc cell was rather limited with a quickly fading current. Soon, William Cruickshank proposed an improvement with the trough battery which paved the way to battery mass production. In 1836, Daniell developed an eponymous battery with a steadier current, allowing practical applications. It became the industrial standard associated notably with telegraphs development. At this time, all batteries were primary batteries meaning that

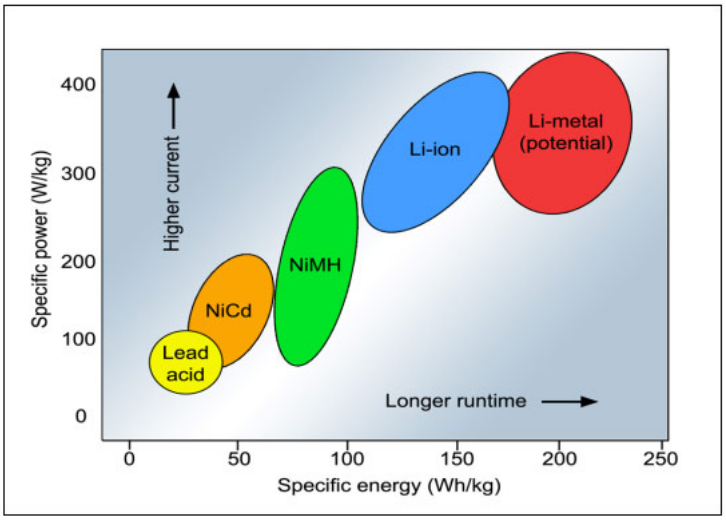


FIGURE 1.1 – Comparison of the specific power and specific energy of the main battery chemistries, from [28]

they could not be recharged. The first secondary battery, meaning rechargeable, came with the work of Wilhelm Josef Sinsteden in 1854 and improvements by Planté in 1859 leading to the lead-acid battery. This kind of battery can provide instantaneously a high current which has long been used to start automotive engines. However, it is limited by its low energy density and lead toxicity. Leclanché cells patented in 1866 introduced the modern cylindric cells, still in use today. It was the first dry cell with a non-liquid electrolyte, limiting the possible leakage. In 1899 Waldmar Jungner from Sweden, introduced the nickel-cadmium (NiCd) batteries. Even though it has a relatively high cost and Cadmium is a highly pollutant material, it was for long the main battery model for mobile applications, replacing lead-acid batteries. More recently, just before the development of Lithium-ion batteries, Nickel-metal hydride batteries, Ni-MH, were developed replacing the Cadmium of the NiCd battery by an hydrogen-absorbing alloy. Besides reduced toxicity compared to NiCd batteries, Ni-MH have a lower cost, a longer lifetime and a reduced memory-effect. However, their main drawback is an important self-discharge.

This thesis concentrates on another type of battery which has now taken the lead over other chemistries, the Lithium-ion technology. Lithium presents the major physical properties of having the highest electronegativity among Alkali metals while being the lightest of them. As illustrated in Fig. 1.1 these advantages allowed the creation of batteries with a significant energy density and power density [29, 30].This explains their widespread use in mobile applications, electronic devices since the 2000s, and electric vehicles in the present days.

Development of Lithium batteries began in the 1960s, initially using directly Lithium metal in non-rechargeable batteries [31]. However, its use remained limited due to safety issues and high initial costs. The rise of lithium batteries came with the replacement of Lithium metal by Lithium-ions inserted in another component improving the battery stability at the cost of a reduced energy density. The invention of the insertion mechanism is attributed to Stanley Whittingham, but the first commercially usable model was developed by John Goodenough [32] who introduced the Lithium Cobalt chemistry. In addition to their high energy and power density, this battery presented over the other advantages of a lack of memory effect, relatively long cycle

life, or low maintenance [29], even if their cost and the safety concerns were higher. Lithium-ion batteries were then introduced in the customer market by Sony in 1991 for electric devices and quickly became the dominant battery technology. In 2019, Goodenough alongside Stanley Whittingham and Akira Yoshino, was awarded the Chemistry Nobel prize to express their decisive contribution to the development of the Lithium-ion technology.

In this chapter, we will start in Sec. 1.2 by introducing the necessary physical background and vocabulary on batteries. We will present the fundamentals of electrochemical reactions, and describe their main battery elements and characteristics. This will allow us to discuss their aging behavior and its quantification. Then Sec. 1.3 will explain our goal regarding the modeling of this degradation. We will explicit the problem, highlighting the importance of quantifying associated uncertainties, and present the different solutions existing in the literature.

1.2 Background on batteries

1.2.1 Battery presentation

Lithium-ion cell operation

Before delving into the ageing of a battery we need to understand its basic operation. A battery turns chemical energy into electrical energy through a Redox reaction [33]. At the oxidation a species called the reducer, R, loses electrons and becomes an oxidant, O. The reduction is the opposite reaction, where an oxidant gains electrons and becomes a reducer. Thus, oxidation corresponds to the left to right reaction, and the reduction to the right to left reaction of the following equation

$$\mathrm{R} \rightleftharpoons \mathrm{O} + xe^{-}. \tag{1.1}$$

The Redox reaction is the combination of an oxidation and a reduction. Two couples are considered $(\mathrm{R}_1, \mathrm{O}_1)$ and $(\mathrm{R}_2, \mathrm{O}_2)$ and perform simultaneously an oxidation and a reduction operation. R_1 oxides into O_1 and O_2 reduces into R_2, corresponding to the global reaction,

$$\mathrm{R}_1 + \mathrm{O}_2 \rightleftharpoons \mathrm{O}_1 + \mathrm{R}_2. \tag{1.2}$$

Importantly, the reaction needs to be reversible to produce a rechargeable battery. This is denoted by the two-sided arrows.

To facilitate this redox reaction, a battery is composed of two main electrodes a *cathode* and an *anode*. The reduction takes place at the cathode while the oxidation takes place at the anode. Their role is reversed between the charge and the discharge of a battery as the reaction is reversed. Yet in practice we will use the discharge convention, calling the electrodes by their role during the discharge.

In Lithium-ion batteries both reactions involve the $(\mathrm{Li}/\mathrm{Li}^{+})$ couple. In general, Lithium is not utilized in its pure metallic state. Its high reactivity leads to safety issues and accelerated

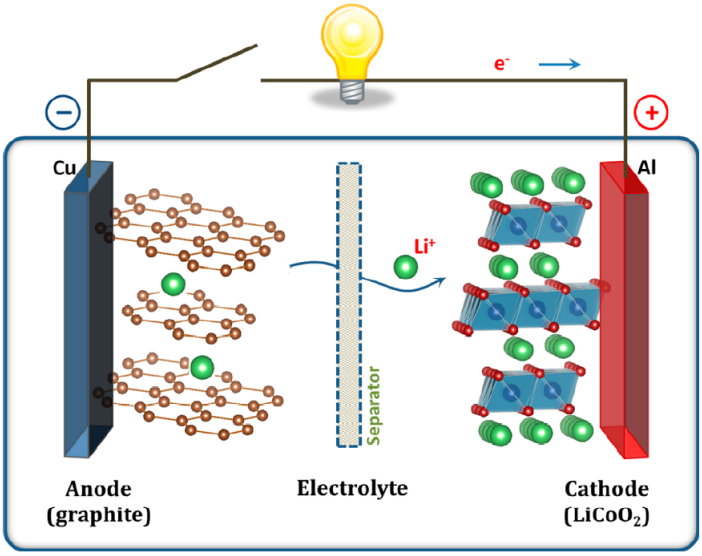


FIGURE 1.2 – Sketch of a LCO battery, from [34]

ageing with dendrite formation leading to short-circuits between electrodes. Thus, in most modern Lithium batteries, Lithium is inserted in another material.

As an example, the LCO battery, introduced by Goodenough and co-workers, stands for a Lithium Cobalt Oxide cathode, $LiCoO_2$. The anode is composed of graphite, a common choice in Lithium-ion batteries. These batteries correspond to the popular first generation of batteries, massively used in mobile electronic devices such as cell phones. As sketched in Fig. 1.2 during a discharge a proportion y of the Lithium contained in the Graphite anode, is oxidized, producing Li^+ ions and electrons

$$Li_{x+y}C_6 \rightharpoonup Li_xC_6 + yLi^+ + ye^- \tag{1.3}$$

Simultaneously, a reduction occurs at the cathode, leading to the insertion of an equivalent proportion of Lithium-ions within the Lithium Cobalt Oxyde.

$$Li_zCoO_2 + yLi^+ + ye^- \rightharpoonup Li_{z+y}CoO_2 \tag{1.4}$$

The electrodes are not connected directly. Through a binder, the anode is connected to a current collector generally in copper. It is connected to a load, and then another current collector in Aluminum, makes the connection to the cathode. This allows the flow of electrons. The positive charges pass through an electrolyte, a material allowing ionic flow but not electronic flow. The cell also contains a separator avoiding short circuits while allowing a reduced distance between the electrodes. During the charge, the opposite reaction occurs.

Name	Material	Comments
LCO	Cobalt Oxyde	High volumetric density but high Cobalt price
LMO	Manganese Oxyde	Cheap but low lifespan
NMC	Nickel Manganèse Cobalt Oxide	Rising alternative to LCO, cheaper and safer
NCA	Nickel Cobalt Aluminium Oxyde	Close to NMC but lower lifespan
LFP	Iron Phosphate	Safe but low volumetric density

Table 1.1 – Main cathode chemistries

Lithium-ion batteries characteristics

By presenting Lithium-ion cell operation, we introduced the main components of a battery, namely the electrodes, cathode and anode, the electrolyte, the separator, the binders, and the current collectors. The chemistry of a battery is defined by the choice of all these elements with variations in each model. The main choices rely on the electrode and the electrolyte.

For Lithium-ion batteries, the anode is most of the time made of graphite. This electrode has major advantages over other possible options [35], it is cheaper and compensates for the high reactivity of alkali metal anodes improving stability. Different kinds of graphite can be used, it can be natural or artificial. Its natural version is cheaper, being highly abundant on earth, but the artificial version has better performance. The combined use of graphite with Silicon gained popularity in recent years but few others are used in practice. Lithium Titanate, LTO, can for some applications compensate its higher price and lower specific energy by a much longer lifespan, with higher safety guarantees and the possibility of fast charges. Purely metallic Lithium anodes are also still really appealing with a much higher energy density. But the dendrite formation which can break the separator and short-circuiting the cell is still an unsolved problem.

The majority of research efforts are focused on the selection of the cathode [36], such that a battery is often called by the chemistry of its cathode. This electrode is generally a lithiated metal oxide or lithiated metal phosphate, a list of the most popular materials is provided in Table 1.1. As presented above the first lithium-ion batteries, introduced by Goodenough and coworkers used a Lithium Cobalt Oxyde (LCO), $LiCoO_2$, cathode. They were developed for portable electric devices and remained for long the most popular chemistry for these applications. This is due to its high volumetric energy density allowing to reduce the size of the devices. However, they are progressively replaced by other chemistries notably because of the high price and environmental issues linked to Cobalt. Moreover, they are less suited for electric vehicle applications where the gravimetric density is a more important factor.

A proposed alternative without Cobalt is the Lithium Manganese (LMO), $LiMn_2O_4$, cathode. Thanks to a low resistance they can be charged quickly but their lifetime is reduced and they are unstable at high temperatures. After difficult tests with Lithium Nickel Oxydes, $LiNiO_2$, a Nickel-Manganese-Cobalt (NMC), $LiNiMnCoO_2$ chemistry combining the benefits of Nickel, Manganese, and Cobalt was produced. They are safer and less expensive than LCO cells with a reduced use of Cobalt. They are highly popular currently since they are better suited to electric vehicles and energy storage systems. Another alternative with properties close to NMC is the Nickel-Cobalt-Alluminium (NCA), $LiNiCoAlO_2$. Its specific energy can be higher, yet it has a lower lifetime. Finally, Iron phosphate batteries (LFP), $LiFePO_4$, propose another interesting compromise. Even if their volumetric energy is lower than the one of NMC and NCA they are

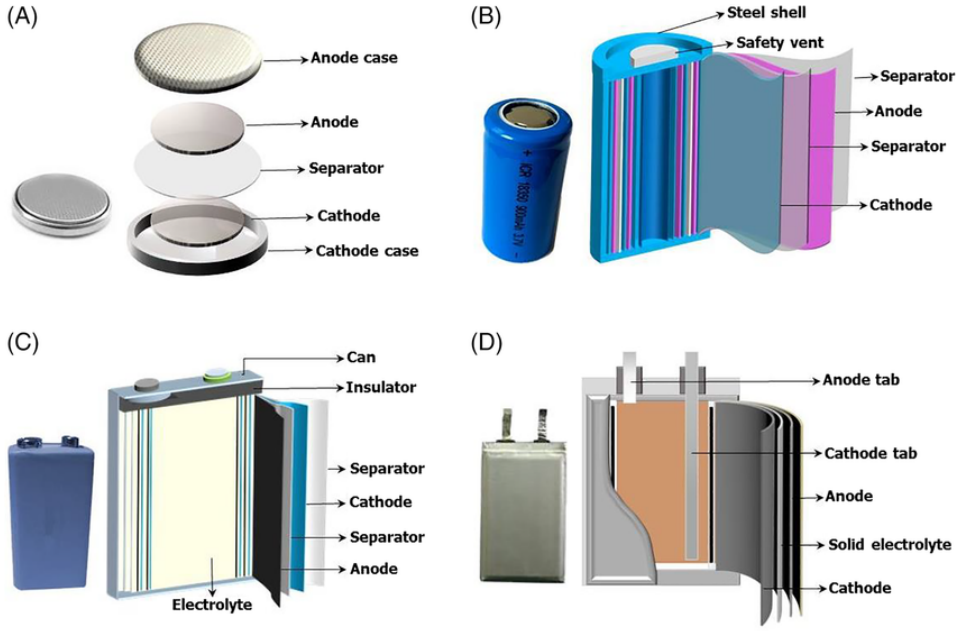


FIGURE 1.3 – Picture and sketch of main battery configurations, A) button/coin, B) cylindrical, C) prismatic and D) pouch cell, from [38]

very safe and have a long lifespan. They are gaining popularity recently because of the lower cost of iron.

Apart from the electrodes the influence of the electrolyte should not be neglected. The electrolyte is important to facilitate the ionic flow. It is often liquid but non-aqueous because of the high reactivity of Lithium with water. It is generally composed of a Lithium salt in an organic solvent. Depending on the manufacturers many additives can be used to increase the conductivity or facilitate the SEI formation. Since electrolyte are the causes of safety issues, being flammable and thermally unstable, important researches are currently led to improve them notably with the development new solid electrolyte [37].

To perfectly define a battery characteristics we need, in addition to its chemistry, to precise its design. This includes the sizes, electrode lengths and contact surface, and configuration. There exist four main types of battery configurations, pictured and sketched in Fig. 1.3, button/coin, cylindrical, prismatic, or pouch cell, each existing with different sizes and chemistry.

Button (or coin) cells are miniaturized batteries with limited applications nowadays, except in some specific products such as watches and memory backup. Typically, one of their limitations was the need for a long charge because of swelling issues. The Cylindrical cell, with its standard 18650 format, is probably the most used battery format. They correspond to the batteries used in our everyday life for example for remote control devices and are currently massively used for electric vehicles. The cylindrical shape is not optimal for stacking, but it eases manufacturing, reducing the production cost and improving the mechanical stability. Prismatic cells, due to their slim size, are the preferred choice for devices like cell phones and laptops. This shape eases stacking, making them an alternative to cylindrical cells for battery packs. Finally, pouch-cells

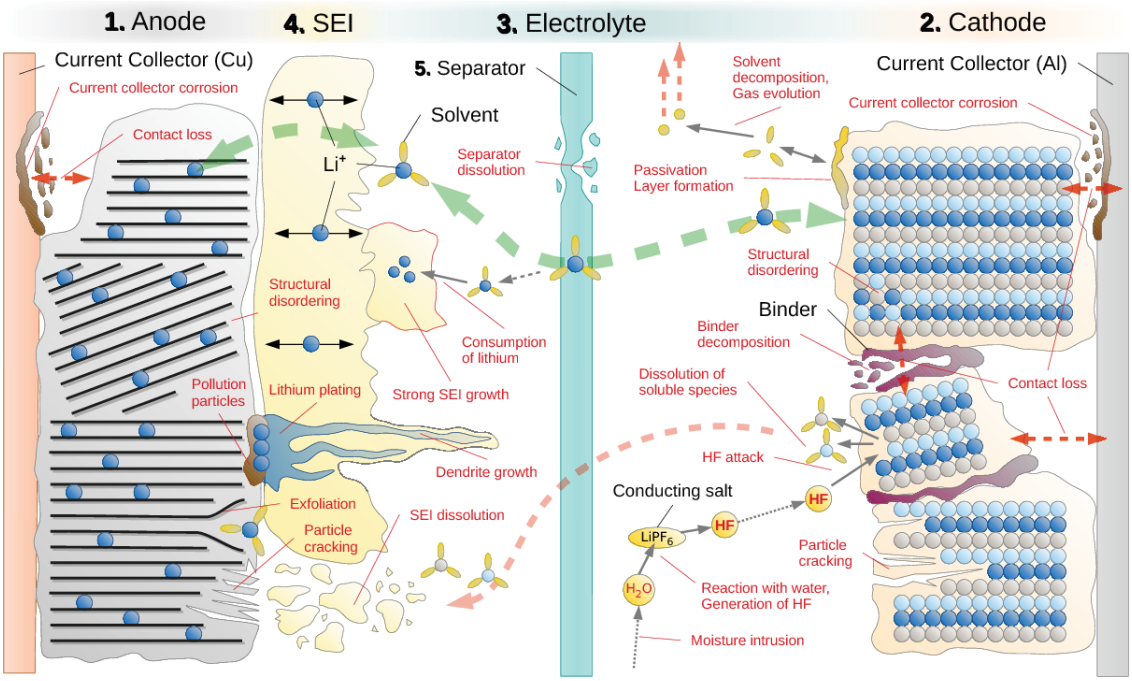


FIGURE 1.4 – Sketch of main degradation mechanisms, from [39]

have the specialty of being flexible with any possible sizes. Moreover, their reduced packaging makes them lighter than other battery types. They are used for cell phones and applications requiring specific shapes.

1.2.2 Batteries degradation

Battery degradation is a complex topic mixing many physical and chemical mechanisms at various scales. In this thesis, we will focus on "standard" aging processes. Extreme events related to safety issues are a major topic in batteries applications notably for the Lithium-ion chemistry [40], but we will not cover it here.

Ageing mechanisms

The main degradation mechanisms are illustrated in Fig. 1.4 provided by Schlasza *et al* [39]. All the components of the battery are concerned. The electrodes can undergo structural disordering or particle cracking. External layers can also affect the electrolyte, mainly the Solid Electrolyte interphase (SEI) at the anode, but also the Cathode electrolyte interphase (CEI). Many reactions can occur at these interphases. The electrolyte can lose its additive and solvent in side reactions. The separator is degraded because of dissolution or mechanical stresses increasing the risk of a short circuit because of a contact between the two electrodes. Corrosion of the current collector and decomposition of the binder can also reduce the contact with the electrode. Finally, besides

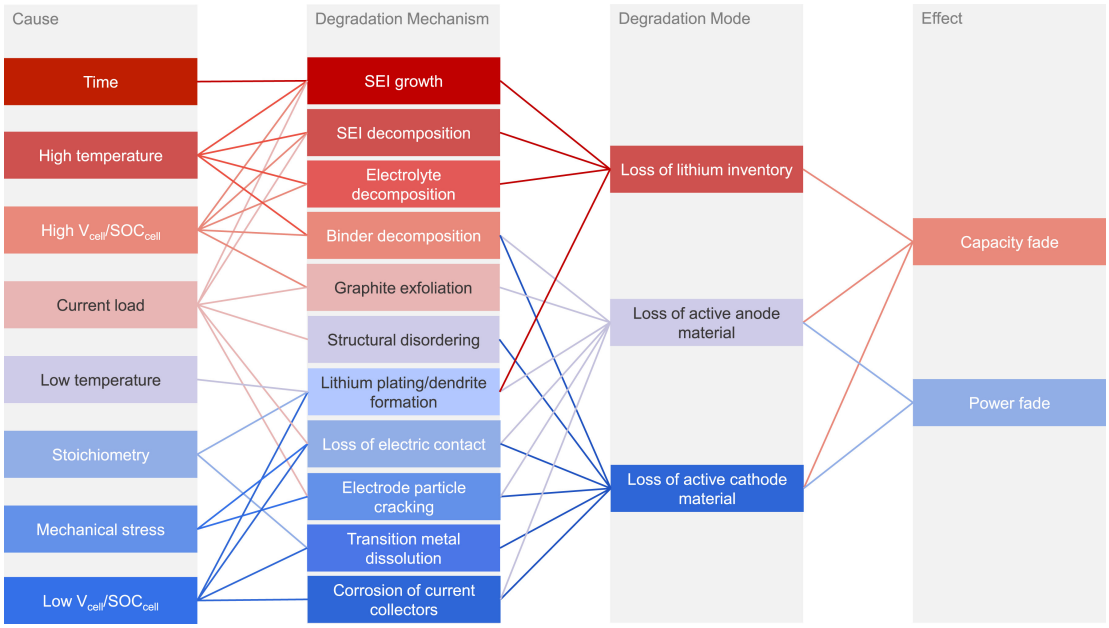


FIGURE 1.5 – Nomenclature of the different degradation mechanisms, from [2]

these standards elements of the battery, exterior elements can also age, cell terminals, cell housing, or isolation.

As presented in Fig. 1.5 this diversity of ageing mechanisms can be resumed in two main categories 1) the loss of Lithium inventory and 2) the Loss of active material [2]. The loss of Lithium inventory concerns the quantity of Li^+ ions that can move from one electrode to the other. The loss of active material concerns the reduction of insertion sites available in the electrodes to store the Lithium ions. It can be decomposed between reactions in the negative electrode and the positive electrode and contains all the particle cracking and active site blocking. Some additional reaction modes can be considered [33] such as the Ohmic resistance increase caused by loss of contact with current collectors or the increase of resistive layers as the SEI.

Among all these degradation mechanisms, two will typically have a major impact on the degradation trend. At the beginning of life, the SEI layer at the anode starts to increase. As a side reaction, it consists of a loss of Lithium inventory. Yet it has the positive effect of protecting the anode from corrosion slowing down the aging, and is thus desired to some extent by manufacturers. On the opposite, lithium plating, consisting of a solid Lithium deposit on the anode blocking Lithium-ion deinsertion and possibly leading to dendrite growth, occurs later in the battery life and can lead to a quick fade in battery health.

The degradation mechanisms depend on the conditions of use of the battery as summarized in Fig. 1.5. The primary division is between a storage and cycling application [41]. In the former scenario, the battery is only subject to a calendar ageing resulting from thermodynamic instability of materials. In the latter case, the degradation is coupled with kinetically induced effects. In any case, the ambient temperature is a key experimental factor. High temperatures tend to accelerate chemical reactions and the decomposition of the elements. In cycling, low temperatures may also have an important impact causing lithium plating with a fast degradation of performances.

The charging/discharging policy also has an important impact. In calendar settings, there is no charge or discharge but one needs to take into account the State of Charge (SoC) of the battery, whether is it left fully or partially charged. In the cycling context, we first need to consider whether the cycles are complete or not. In the latter case we specify the upper and lower bound of this charge and the range is called the depth of discharge (DoD). For instance, one can consider an experiment where a battery is fully discharged and then only charged to half of its capacity, fully discharged and so on. In that case, there is a DoD of 50%. Small DoD and more central lower and upper bounds will generally degrade less the battery.

Then many charging and discharging profiles can be used. But globally, higher currents and more extreme voltages degrade more the battery. We should also be careful that even in a cycling context the battery may not be charging and discharging all the time. Some resting time could occur and enhance temporary the battery health, by reducing its internal temperature for example.

Degradation measurement

We described qualitatively the main degradation mechanisms. In order to compare battery models we also need to quantify this degradation. For this, we use some aggregated indicators called State of health (SoH) indicators. The most oftenly used are the battery capacity and the internal resistance. Such indicators also provide a summarized idea of the battery usefulness for practical applications and when it can be considered out of service.

The capacity is the amount of charge that a battery can deliver during a discharge. It is equal to the integral of the current curve, $i(t)$, during a complete discharge

$$C = \int_{t_{min}}^{t_{max}} |i\,(t)|\,dt, \tag{1.5}$$

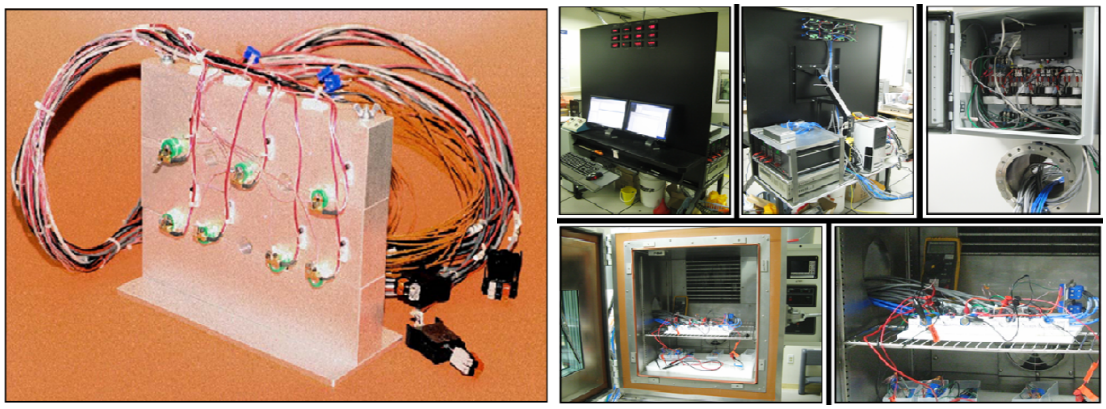
which is expressed in Ampere-hour (Ah). A 1 Ah battery contains enough charges to deliver during one hour a 1 A current or equivalently a 0.5A current for 2 hours. Starting from an initial value the capacity will progressively decrease with time and the number of charges and discharges. So a user will progressively get less and less energy from the battery and will have to recharge it more frequently. In the automotive sector, a common practice is to consider that a battery can no longer be used once it has reached a 20% capacity loss compared to its initial value. We remark that this definition corresponds to the discharge capacity, the charge that a battery can deliver. We can similarly define a charging capacity, the amount of charge that can be stored in a battery, which may differ slightly.

The other important state-of-health indicator is the internal resistance. While the capacity is linked to the amount of energy that a battery can supply, the internal resistance is linked to the power it can deliver. A high internal resistance limits the maximal power that a battery can deliver, and reduces the ability to deliver precise power spikes. The internal resistance will typically increase during the battery life.

These two primary indicators can be normalized in various ways in practice, such as by dividing by the nominal or initial value. The main interest among these two indicators may vary depending on the considered application. In this thesis, we will mainly consider capacity modeling,

as it is more commonly studied in the literature notably for automotive applications. In addition to these two main indicators, the raw measurement of current, temperature, and voltage can also be useful in practice. On the opposite, more detailed measurements such as Electrochemical Impedance Spectroscopy (EIS) can also be performed.

Remark For a given battery at a given time, the measured SoH depends on the temperature and the charging/discharging policy. This can be an issue when one wants to compare the performance of batteries operated under different conditions. That is why, in experimental settings, some control cycles are performed regularly at a standardized condition (potentially different from the testing condition).



(a) (b)

FIGURE 1.6 – Experimental setups with (a) a thermal block with cells and (b) a NASA ARC prognostic testbed, from [42]

The battery degradation may be measured in two types of contexts. They can either be experimental data or on-field data. Experimental data are realized in a laboratory by specialists in controlled operating conditions as depicted in Fig. 1.6 with cells in a thermal block, Fig. 1.6a, and the NASA ARC prognostic testbed in Fig. 1.6b, from [42]. Detailed measurements are generally done but they are costly and time-consuming leading generally to a limited number of tested batteries. On the other hand, on-field datasets come from real-life situations. They are directly linked to the end-user application and a higher volume of batteries can be included. Yet, these data are noisier, there are more uncertainties on the operating conditions and on the state-of-health. Experimental data are most of the time unrealistic, with fixed charging policy as CC-CV (Constant Current - Constant Voltage, see Appendix A) and temperature, but it can also be argued that on-field data, while specific to one type of application, may not be representative of other usages.

Researchers try progressively to fill the gap between experimental and on-field data providing more complex experimental conditions with varying currents and temperatures. The number of these datasets remains limited because of the extra costs involved, but significant literature exists in the automotive sector [43]. In this context, different vehicle velocity profiles have been created and associated discharging policies can be imposed with a proportional current.

Currently, research and publicly available datasets are mainly using experimental datasets [43]

although recent studies have begun to address this issue [44]. Even if field data are present in higher volume, they are probably harder to share, being directly linked to commercial usage. In this thesis, because of our goals, we will concentrate on experimental data. Indeed, experimental data correspond to the data available to the manufacturers in practice. Manufacturers produce frequently new battery models and they need to evaluate their performance before selling them to consumers so they need to rely mainly on their own experimental data to evaluate the battery performance. Moreover, on-field data are often available years after the creation of the battery model so to some extent they may already be outdated compared to batteries currently produced. For manufacturers, a crucial issue would be to anticipate on field performance of batteries using only experimental data.

Datasets

In this thesis, we will apply our methods on publicly available datasets [45, 46]. Important progress has been made in the last decade regarding open data. Only a few batteries were tested fifteen years ago, whereas the latest datasets often contain one or two hundred cells. Yet there is still an important need for more exhaustive experiments, particularly regarding the number of tested conditions and their complexity. We will mainly concentrate on two datasets, one provided by the Aachen University and another by the Maryland University.

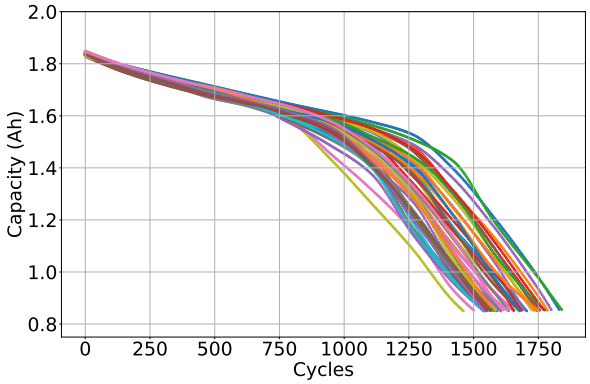
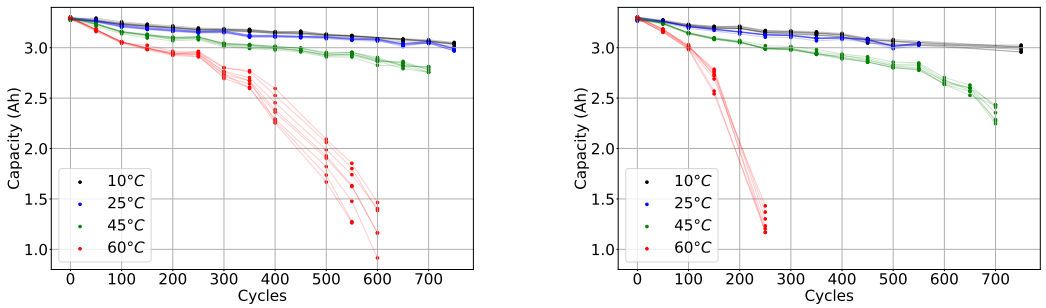


FIGURE 1.7 – Aachen capacity curves

The Aachen dataset, see Fig. 1.7 will permit us to model precisely the health degradation and associated uncertainties at a reference experimental condition. The dataset was introduced in [47] and made publicly available after [48] publication. It consists of 47 tested Panasonic/Sanyo UR18650E lithium-ion batteries, cylindrical 18650 graphite/NMC batteries, with a nominal capacity of 1850 mAh. All the batteries were charged and discharged similarly in 30 minutes at 25°C. The authors used a CC-CV (see Appendix A) policy with CC at 2C until 3.9V and CV until the end of the 30 minutes. Tests are performed until a capacity loss around 50% of nominal capacity. The dataset is publicly available on GitLab [45] and full specifications can be found in table 3 of supplementary materials [48].

The Maryland dataset was proposed by Diao *et al* from the CALCE battery team of Maryland University. Part of it is displayed in Fig. 1.8. It consists of 192 pouch cells of a 3.36 Ah nominal



(a) (b)

FIGURE 1.8 – Capacity degradations from the Maryland dataset for a discharge rate of 0.7 C and a charge current cut-off of (a) 0.2 C or (b) 0.025 C .

capacity, with a graphite anode, a $LiCoO_2$ cathode, and $LiPF_6$-salt mixed with liquid organic solvent electrolyte. Three experimental factors are considered, the ambient temperature, the current cut-off of the CV phase of the CC-CV charge, and the discharge current. The authors considered 4 levels for the temperature (10, 25, 45, and 60°C), 2 for the current cut-off (C/5 and C/40), and 3 for the discharge current (0.7, 1, and 2C). A full factorial experimental plan were used, leading to 24 conditions with tested batteries in each case. Cycling is performed until a maximum number of 800 cycles with control cycles every 50 cycles. With 192 batteries this dataset is one of the biggest currently available.

1.3 Degradation modeling

Now that we have presented the physical causes of the degradation and its measurement we can model it. First, we will define explicitly the available data and the quantities to be estimated and add a special focus on the uncertainty quantification. Then, we will review the existing approaches in the literature to tackle this task.

1.3.1 Thesis problematic

Our main goal is to characterize the typical degradation of a battery design given the experimental conditions. We are interested in modeling the complete degradation curve. So our problem is larger than the lifetime prediction problem which only targets the prediction of the moment when the SoH indicator reaches a threshold. It will answer the need of manufacturers to quantify performance throughout the entire lifespan of the battery.

As previously stated, the degradation of a battery is quantified by a SoH indicator, most of the time the capacity of the battery or its internal resistance. This SoH indicator denoted as y varies with time $t \in \mathcal{T}$, either a cycle number in a cycling context or calendar time for storage applications. The degradation depends on two types of elements, characteristics of the battery and its operating conditions (experimental conditions in our case). As presented in the

previous section, battery characteristics contain their chemistry (electrodes, electrolyte ...) and their design, sizes and configuration (cylindrical, pouch-cells...). The experimental conditions contain the ambient temperature, or all the elements defining the charging and discharging policy, the charge or discharge rate, the depth of discharge, the resting time... In this thesis, we will concentrate on a univariate SoH, but several indicators could also be modeled simultaneously [49].

Denoting $\mathbf{c} \in \mathcal{C}$ the experimental condition and $\mathbf{d} \in \mathcal{D}$ the battery characteristics our general problem takes the following form

$$(\mathbf{c}, \mathbf{d}) \mapsto (t \mapsto y_{\mathbf{c},\mathbf{d}}(t)), \tag{1.6}$$

given conditions and characteristics a time evolution of the SoH is predicted. We will assume $\mathbf{c}$ and $\mathbf{d}$ to be vectorial inputs of fixed sized. This is direct for fixed quantitative factors, such as controlled ambient temperature, charge and discharge current, or battery size, while quantitative variables such as battery configuration or chemistry can be vectorized through an encoding. For experiments considering time-varying conditions, currents, or temperature, this framework can be used in simple evaluations, completely defined by linear coefficients or some fixed levels for example, but more complex profiles would be harder to include.

This corresponds to the ideal problem, however, in practice, experimental conditions and battery characteristics are never perfectly known and each particular battery will have a different degradation. So we need to complete the above problem by adding a dependence in the considered battery element b,

$$(\mathbf{c}, \mathbf{d}) \mapsto ((t, b) \mapsto y_{\mathbf{c},\mathbf{d}}(t, b)). \tag{1.7}$$

At this point, we need to make a distinction between two different problems present in the literature. Experiments are processed on B different batteries with $b \in [\![1, B]\!]$. The two different problems often considered are 1) a prediction for a specific battery present in the dataset $b \in [\![1, B]\!]$, 2) a prediction of the average behavior associated to the battery batch. This last problem is equivalent to making a prediction for batteries outside the training dataset $b \notin [\![1, B]\!]$, without any knowledge of the battery, the average behavior is predicted.

The first problem can be linked to predictive maintenance problems. Using data collected by a battery management system (BMS) the end-of-life of the battery for example can be anticipated. It is linked to the often considered Remaining useful life (RUL) problem, see for example [50]. On the other hand, the second problem focuses on the characterization of a battery design at a particular experimental condition. In link to the industrial context of the thesis, we will mainly focus on the second objective. It enables manufacturers to determine the value of new battery designs before introducing them to the market.

To fulfill our goal, we consider a setting where available data consist of a total of n observations of a SoH indicator y, realized on B batteries. Each battery b is tested at a condition $\mathbf{c}_b$ and has characteristics $\mathbf{d}_b$. An observation is done at a particular time t, so the concatenated input corresponds to $\mathbf{x}_i = (t_i, b_i, \mathbf{c}_{b_i}, \mathbf{d}_{b_i})$ and output $y_i = y(\mathbf{x}_i)$. Finally the whole dataset consists of $(X, \mathbf{y})$, with $X = \{\mathbf{x}_i\}_{i=1}^n$ and $\mathbf{y} = y(X) = (y(\mathbf{x}_1), \cdots, y(\mathbf{x}_n))^T$.

Remark One may be interested in adding supplementary inputs to train the model, such as information from the voltage of the current curve or more precise EIS measurements. This is an interesting possibly for the first problem, these data being recorded simultaneously to SoH. Yet, they cannot be used for the characterization problem since no knowledge is available for future batteries.

The problem of Eq. (1.7) encapsulates the complete goal but it is most of the time unrealistic considering the available experimental data. In practice, the space of experimental conditions $\mathcal{C}$ is reduced to a few factors in a limited range, and most datasets only consider a unique battery model. Additionally, to simplify the modeling we will cut the problem into two subproblems respectively studied in the two following chapters.

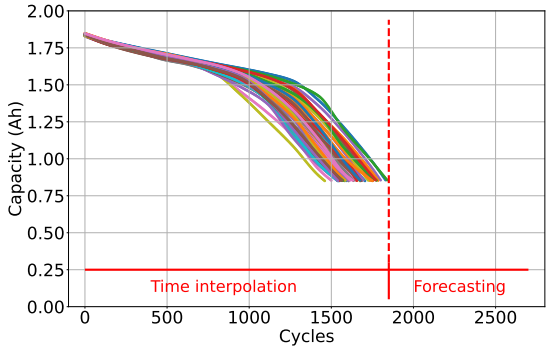


FIGURE 1.9 – The time degradation problem at a fixed experimental condition

The first considered subproblem treated in the next chapter 2 focuses on the time evolution at a unique experimental condition and battery characteristics, $\mathbf{c}$ and $\mathbf{d}$. So we consider the modeling of

$$(t, b) \mapsto y_{\mathbf{c},\mathbf{d}}(t, b), \tag{1.8}$$

where $\mathbf{c}, \mathbf{d}$ subscript will be removed in the following. In that case, inputs reduce to $\mathbf{x}_i = (t_i, b_i)$. As presented in Fig. 1.9, this problem will be split into two cases. Considering observations of state-of-health measurements observed up to a limit time t_{max}, we will furthermore separate the prediction on the data before t_{max}. The first case is easier with available observations, while the second is a time extrapolation, a forecasting task. It's important to note that we define forecasting here as making predictions at times when no observations are available at all. This differs from forecasting for a specific battery at cycles that have been observed for other batteries.

The second subproblem, treated in chapter 3, enlarges the previous one by including dependence in the experimental conditions keeping a unique battery design

$$\mathbf{c} \mapsto ((t, b) \mapsto y_{\mathbf{c},\mathbf{d}}(t, b)). \tag{1.9}$$

Now the available inputs for training consist of $\mathbf{x}_i = (t_i, b_i, \mathbf{c}_{b_i})$. More precisely, we will consider

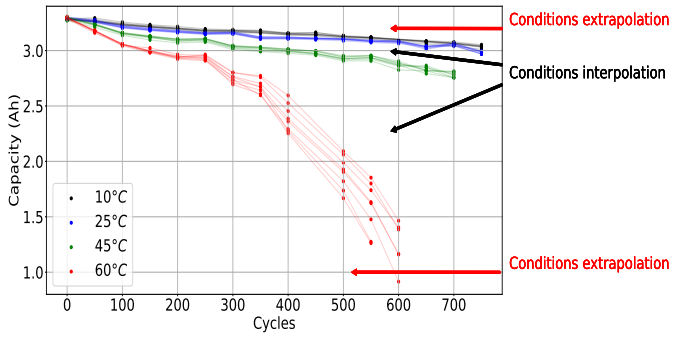


FIGURE 1.10 – Modeling the experimental factor effect

the common cases of K tested experimental conditions where each condition $\mathbf{c}_k$ has B_k tested batteries.

The main interest of this task will be to provide a prediction of the degradation at unobserved experimental conditions, which could not be tested during the experiment. As illustrated in Fig. 1.10, it can also be divided between an interpolation and extrapolation task regarding the experimental conditions. Considering the example of the Maryland dataset, a prediction at 30°C or 50°C corresponds to an interpolation of the conditions, whereas a prediction at 70 °C or 0°C is an extrapolation. The interpolation problem is complex as no data is available at these positions and extrapolation is even harder as monotonicity or acceleration knowledge is not guaranteed. So we will mainly focus on condition interpolation.

Further study should consider the effect of the battery characteristics. Two potential applications were anticipated, one similar to the previous problem, predicting the performance of unobserved battery designs, for example for different electrode lengths, but public experimental data on the detailed battery design are often missing. Another potential problem would be to transfer the knowledge present in a particular chemistry to another chemistry through adaptation domain methods [51]

1.3.2 Uncertainties quantification

A compelling demand from manufacturers is to quantify the uncertainties associated with the degradation modeling. The goal is to predict confidence intervals instead of punctual predictions of SoH loss. Manufacturers can then use them to assess the financial risk related to performance guarantees, for example, based on the probability of having a 20% capacity loss after 1000 cycles. These uncertainties should be estimated as accurately as possible to prevent misguided decisions.

We shall here precise what we call an "uncertainty". Uncertainty is relative to a quantity whose exact value is unknown. The quantity is commonly modeled as a random variable. It generally has a location parameter, corresponding to its mean value, but also a scale parameter linked to its variance. The uncertainty quantification primly corresponds to the estimation of the scale parameter giving the amplitude of the uncertainty. The most important case is the use of a Gaussian distribution defined by the mean and variance parameters. Other distributions may be used for example to quantify other types of variables, discrete or categorical, to model extreme

values or asymmetries.

Uncertainty quantification can be precised by discussing the sources of uncertainties. This is a highly important topic since it improves the interpretability of the model and may provide guidelines to handle particular sources of uncertainty reducing the overall uncertainty. This imposes to detail the modeling, for example with a variance term associated with each source of uncertainty.

To ease the discussion, uncertainty sources are often splitted into division between two types, namely the aleatoric and the epistemic uncertainty [53, 54]. The aleatoric uncertainty corresponds to the intrinsic variability of the phenomenon. Supposing that the same experiment is repeated in the same conditions, the obtained result will always be different. This uncertainty is often qualified as uncompressible, it cannot be avoided since it is naturally present in the process. On the other hand, the epistemic uncertainty reflects our lack of knowledge on the quantity to be estimated. We can generally consider that it would reduce to zero if the amount of data was increasing to infinity.

Hüllermeier *et al.* proposed an illustrative example of these two uncertainties, reproduced in Fig. 1.11. Considering a simple two-class classification problem, the left panel illustrates the aleatoric uncertainty. Part of the support of both classes is shared so in this area there will always be uncertainty in the corresponding class, even with a high number of observations. On the other hand, the right panel shows the effect of epistemic uncertainties. In that case, the boundary between the two classes is not clear because there are only a few observations, causing a lack of knowledge. With more data, the boundary will be precised.

In the battery health degradation context, aleatoric uncertainties can have many different causes [55]. It may be due to variability in the cell behavior itself or to variability in the operating conditions. In this thesis, we will consider in detail the cell-to-cell variations [3]. It refers to the differences between batteries allegedly similar and tested in the same conditions. However, in practice, the operating conditions are never completely controlled and there are still some differences in the manufacturing of the cells. As can be seen in Fig. 1.12a, this uncertainty may be significant with an observed variability close to 20% of the nominal capacity at cycle 1300.

Many other sources of aleatoric uncertainties can be observed in experiments. they can be linked to the measures themselves, in link to the sensor's precision, or to the experimental process.

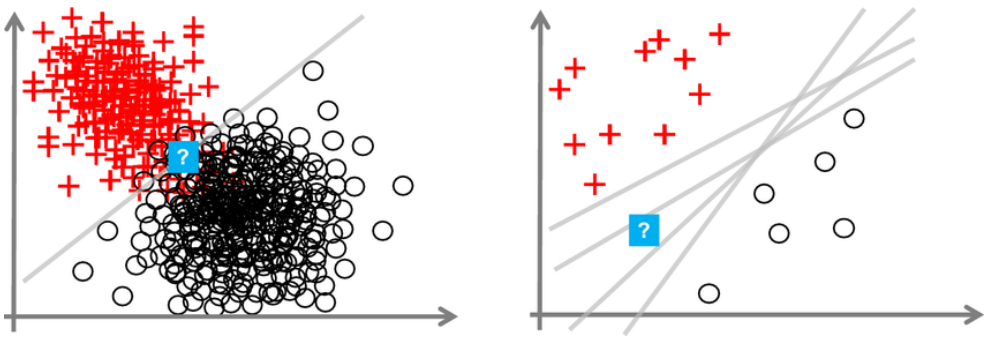
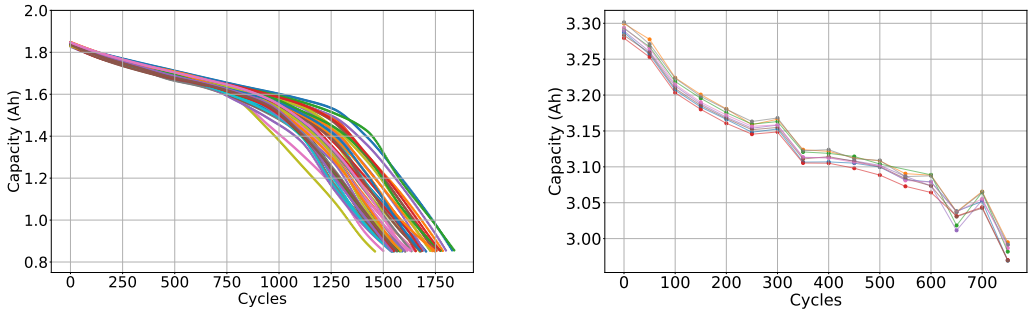


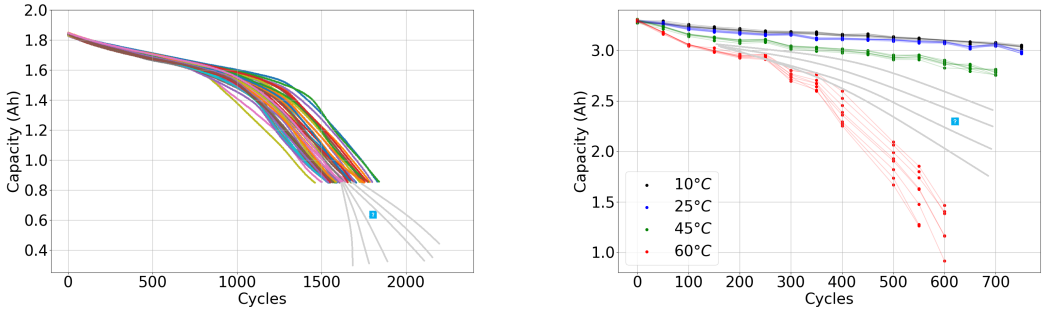
FIGURE 1.11 – Illustration of the effect of aleatoric uncertainties (left) and epistemic uncertainties (right), from [52]



(a) (b)

FIGURE 1.12 – Examples of aleatoric uncertainties with the (a) the cell-to-cell variation in the Aachen dataset and (b) a batch noise in the Maryland dataset

The latter one can affect a particular battery or a whole tested batch, because of a variation of the room temperature or a longer rest because of holidays. Fig. 1.12b focusing on a particular condition in the Maryland dataset illustrates such effects. For example at cycle 300 we see an increase in capacity shared by all the batteries of the batch. This is not due to an independent noise, since it is unlikely to have simultaneously the same increase for all batteries. Thus, the aleatoric noise can have complex structures, partly shared, but also potentially varying with time or asymmetrical.



(a) (b)

FIGURE 1.13 – Examples of epistemic uncertainties in (a) the forecasting task with the Aachen dataset (b) the condition interpolation with the Maryland dataset

The epistemic uncertainty is linked to the amount of available data. Here the problem is the number of tested batteries and the number of measurements. It is a primary concern when we want to make predictions for nonobserved inputs, for us, the forecasting task and the interpolation/extrapolation of experimental conditions. A forecasting problem is illustrated in Fig. 1.13a with the Aachen dataset, after cycle 1800, there is no data available, using only the data, we do not know if the average capacity decreases more or less slowly, and we have less and less knowledge as we make predictions further. This should be taken into account, in the uncertainty quantification. Similarly, taking the Maryland dataset, Fig. 1.13b illustrates the epistemic uncertainty in making a prediction for a temperature between 45 and 60°C. The

associated epistemic uncertainty should be higher for predictions at 52°C than at 46°C, closer to an observed condition.

Either for epistemic or aleatoric uncertainties, an important aspect of our problem is that we want to predict complete degradation curves. In that case, two different options exist, point-wise confidence intervals and full-curve confidence intervals. Considering the prediction of a degradation curve $t \mapsto y_{\mathbf{c},\mathbf{d}}(t, b)$ in both cases we will speak of confidence intervals of level α, corresponding to upper and lower bound functions, denoted $\mathcal{I}_\alpha^{PW}$ in the point-wise case and $\mathcal{I}_\alpha^{FC}$ in the full-curve case. In the point-wise case, this level is valid at each particular time, given $t \in \mathcal{T}$, $\mathbb{P}\left(y(t, b) \in \mathcal{I}_\alpha\right) = 1 - \alpha$. On the other hand, in the full-curve case, we consider that the prediction curve should at all times be inside the confidence interval curve $\mathbb{P}\left(\{\forall t \in \mathcal{T}, y(t, b) \in \mathcal{I}_\alpha\}\right) = 1 - \alpha$.

The full-degradation confidence interval aligns more closely with our task, giving guarantees that the true curve should stay within the confidence bound. However, the most commonly used confidence intervals are point-wise intervals. This is partly due to the ease of computation but also of a major drawback of full-curve intervals being that these intervals can be over-conservative. It is indeed a severe constraint to impose that the predicted curve should never cross the confidence bound. In our case, there is no particular reason to impose such constraint so we will consider the point-wise, and confidence intervals should be considered at each time separately.

1.3.3 State-of-the-art of batteries health degradation modeling

Having defined the problem at hand, we can present an overview of the existing modeling methods. For this overview we will rely on the nomenclature provided in the following review [56]. This nomenclature separates the model-based approaches from the data-driven approaches. In the first case, the function modeling of SoH degradation is specified in advance by the modeler with only calibration of its parameters whereas in the second case, the function is completely learned from the data. The main idea behind this division is whether the modeling is based on physical prior knowledge or learned automatically from the data. The first case leverages domain knowledge which allows to specify a priori a model shape and then fit it on the data. The second case is considered physics-agnostic and relies on the high flexibility of the model to automatically learn complex patterns.

The model-based approach uses the available prior knowledge on battery degradation. It permits a better control of the robustness of the predictions and requires less data than a data-driven approach which is learning the trend from scratch. However, the physical prior knowledge can also be a limit in itself as theoretical modeling may not match the reality of measures. Moreover, a significant increase in battery dataset sizes occurred in the last year. Combined with the exploding popularity of statistical and machine learning methods in many application domains, the data-driven approach gained popularity in recent years.

The model-based approaches can be divided into three different approaches, mechanistic model, equivalent circuit model (ECM), and (semi-)empirical models. They propose a parametric model of the battery behavior. Some, directly model the whole battery health degradation but the parametric nature of these models limit the quality of fit, especially for long-term extrapolation. Thus, the model-based approaches are often combined with filtering approaches such as Kalman filters [57] or particle filtering [58] in order to update regularly the model's parameters. Filtering methods are really popular for battery applications and many variants such as extended Kalman

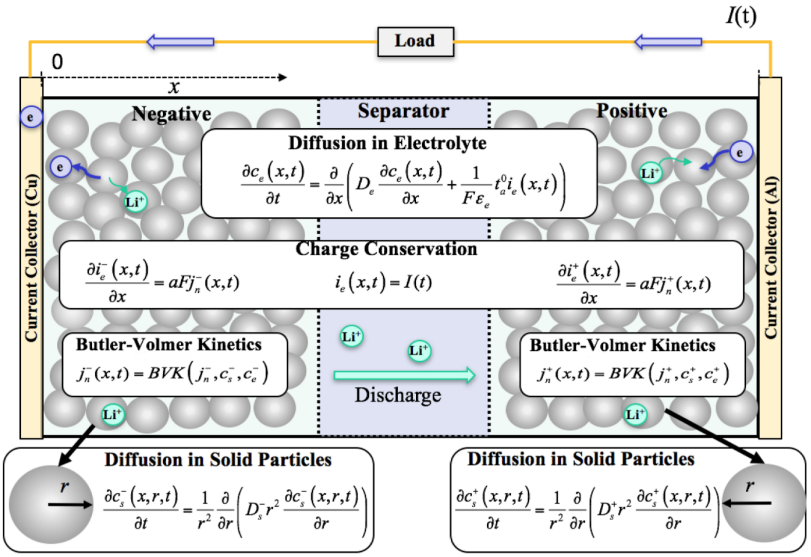


FIGURE 1.14 – Illustration of a mechanistic modeling proposed in [62]

filters [59], unscented Kalman filters [60] or unscented particle filters have been considered [61].

Mechanistic models rely on the physical modeling [63, 64] of degradation mechanism. These models can have different scales and complexities [65] ranging from atomistic models, describing the behavior of each particle, to continuum models with simplified approximation, with homogeneous concentrations and temperature, and simplified geometry. A common model, first introduced for a Lithium-metal battery, is the Doyle Fuller Newman model, also known simply as the Newman model, introduced in [66]. It models the main physical mechanisms, in the different parts of the batteries as illustrated in Fig. 1.14. It is the basis for many specialized developments proposed since then. This approach has the advantage of exploiting the available physical knowledge and the characteristics of the tested battery. However, they can have in important complexity in computational power, needing the resolution of many differential equations, and lack of versatility.

Instead of modeling precisely the physics within the battery, Equivalent circuit models (ECM) [68, 69], approximates its behavior through the construction of an equivalent circuit, using standard elements. Fig. 1.15 shows a detailed modeling of a LMO battery, where each of the battery components has its own equivalent component. To utilize such a model, all parameters - resistance, capacity, and inductance - must be measured. This is typically done through Electrochemical Impedance Spectroscopy (EIS), measuring the response of the battery to different frequencies. Once the parameters are known, a mathematical model of the internal impedance evolution can be derived. Capacity can also be modeled by exploiting its correlation with impedance parameters [42]. However, the ECM model parameters need to be estimated frequently through EIS. These tests need particular cycling which can degrade the battery. Moreover, identification of parameters can be difficult, so regression methods are often applied to estimate battery parameters directly from sensor data.

Lastly, (semi-)empirical models select the model after observation and analysis of the correlation within the data. This consists of a simple exponential or power model with 2 to 4 parameters updated with a filtering method [70, 71], or can model directly the degradation curve with more

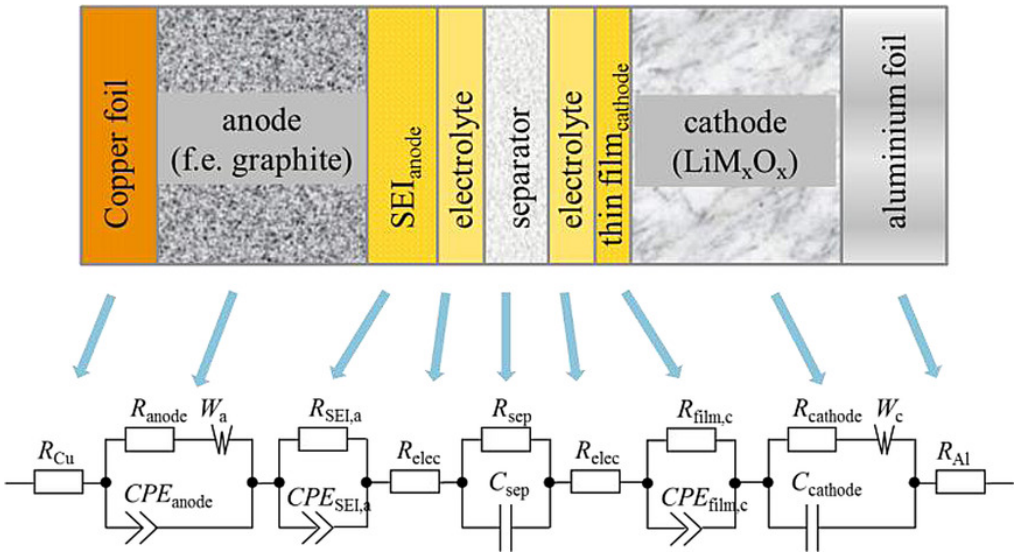


FIGURE 1.15 – Skectch of a LMO battery and associated equivalent circuit model, from [67]

precise modelings [72, 73]. The term "empirical" means that the modeler selects the model after observation and analysis of the data. In practice, this kind of modeling also relies on some physical knowledge as Arrhenius law [74], which is why the term "semi" is often added, to emphasize that the model is not purely empirical.

Contrary to the above methods data-driven methods learn a model, relying only on training data. A common division among them is between statistical methods and machine learning (or artificial intelligence) methods. Machine learning methods mainly focus on learning complex models with the highest accuracy. The classical naive Bayes method has been used to classify batteries between slow and fast degradation [75]. Numerous studies have benefited from the flexibility of the support vector machine (SVM) [76, 77] flexibility, notably in exploiting raw data, even if their computation can become heavy as the amount of data increases. Relevance vector machine (RVM), a probabilistic version of SVM, with a higher sparsity have also been considered [78, 79]. Another probabilistic method that we will consider in detail in this thesis is the Gaussian processes regression (GPR) [80, 81]. Additionally, guided by its ubiquity in almost all applied fields currently, different neural networks [82, 83] have been used. Several architectures suited to time series such as Recurrent Neural networks (RNN) [84] or Long short-term memories (LSTM) [85], have been tested.

Statistical methods are often more restrictive, learning less complex models, but they have the advantage of including uncertainties, being more interpretable and less computational intensive. They have a longer history, being used before machine learning methods. The use of autoregressive models AR is a standard to model stationary time-series. In order to take into account the degradation trend its ARIMA extension handling linear or polynomial trend was considered [86, 87]. Grey models [88, 89] and particularly its $GM(1,1)$ version were also used to model the degradation trend. It is well suited to settings with limited data but it is also limited by a linearity assumption. Wiener process is another kind of stochastic process to model degradation [90, 91]. A particular advantage is that the law of the remaining useful life, the time before the first hitting of a given threshold has a known expression.

This thesis will start from a data-driven point of view. It was a goal in itself to compare traditionally used model-based approaches with more recent data-driven approaches. Moreover this choice is also motivated by our personal background. However, we will see that the frontier between model-based and data-driven approach is not absolute and that a mixture of the two can provide appealing benefits. In any case, the uncertainties quantification discussed in Sec. 1.3.2, will remain an important focus all along this thesis. We highlight again that we decided to model the complete degradation curve not only the lifetime, to provide a complete view of the battery degradation. We will also, in link with the industrial context of this thesis, focus on the characterization problem as opposed to the individual prediction task considered in the remaining useful life literature.

Chapter 2

Battery health degradation prediction at a reference condition with Gaussian processes

Contents

2.1 Introduction

This chapter concentrates on the task of modeling the degradation trend with associated uncertainties at a unique experimental condition and all batteries assumed identical, as introduced in Sec. 1.3.1. This corresponds to a use case often considered in practice, with the condition seen as a reference condition used to compare performances. We also see it as an intermediary step before considering the more complex condition-dependent case studied in the next chapter. This problem will be splitted between the time interpolation case, considering prediction at observed cycles, and the forecasting task, predicting future never observed cycles.

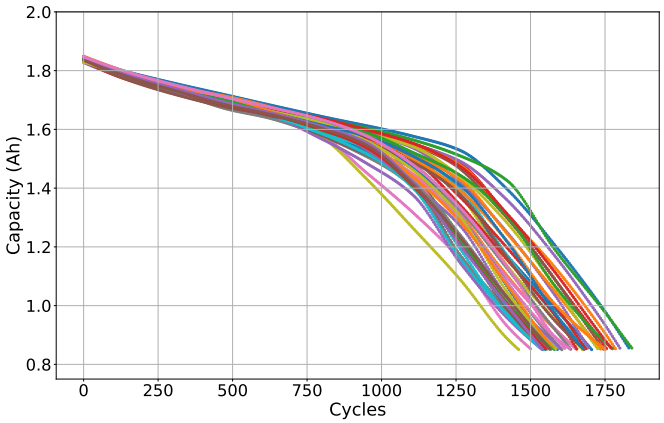


FIGURE 2.1 – Capacity degradation curves of the Aachen dataset

However, as we can see in Fig. 2.1 displaying the Aachen dataset introduced in Sec. 1.2.2, even in that case, important differences in health degradation can be observed. It means that there is an important aleatoric uncertainty on the performance of a new battery, linked to the intrinsic variability of the process. This uncertainty is well known in the literature, referred as the cell-to-cell variations [3]. It can be due to local variations in the experimental conditions, temperature, and charging policy that can never exactly be controlled, or disparities in the manufacturing of each cell. As a result, the performance of some batteries can be much lower than what has already been observed, which can be critical in some applications. In particular, this effect is of high interest for battery packs [92], where identical elements can be used in series and global performance is defined by the worst cell.

More precisely, we see in Fig. 2.1 that this uncertainty increases with time : at cycle 100, the capacity range is approximately 1% of the nominal capacity, whereas, at cycle 1300, it is close to 20%. To our knowledge, the modeling of the time evolution of the cell-to-cell variation has been barely tackled even if it can be observed in many other public datasets, see for example [93], [72], or [94]. An important challenge will be, in addition to the prediction of the mean degradation curve, to model precisely the cell-to-cell variations taking into account its evolution.

To tackle this problem we decided to use methods based on Gaussian processes (GPs), already popular to model the battery health degradation with uncertainties [81, 80]. Using the nomenclature detailed in section 1.3.3 these methods are data-driven, mostly physic agnostic. Yet they combine the advantages of machine learning and statistical approaches. They allow us to learn complex models while naturally modeling uncertainties, as a probabilistic method. Moreover, we will see that GPs can be designed to create more interpretable models and decompose the different sources of uncertainties. These methods are often used in contexts with little data, a common constraint in the R&D context. Additionally, we will see that these methods can be adapted to include physical prior knowledge which can compensate the lack of data as in the forecasting task.

In this chapter, Sec. 2.2 first give some background about Gaussian processes, the base component of our models. This requires the presentation of the main properties of Gaussian

processes, the base component of our models. We then see how they are used in regression within the popular Gaussian process regression (GPR) framework [5] and the more general Chained Gaussian processes (CGP) [95] framework. Then Sec. 2.3 details how we applied these frameworks to model the battery health degradation. The first proposal relies on the GPR framework and will illustrate the limits of this often used framework to model precisely the uncertainties. A more precise uncertainty quantification is thus proposed, exploiting the abilities of the CGP framework. Finally, a particular focus is made to improve the performances of the model in the forecasting task, through the inclusion of physical prior knowledge, exploiting derivative constraints. This chapter is concluded in Sec. 2.4 by a discussion on the strengths and weaknesses of the proposed models.

This chapter contains the results developed in two articles :

— Larvaron, B., Clausel, M., Bertoncello, A., Benjamin, S., & Oppenheim, G. (2023). Chained Gaussian processes to estimate battery health degradation with uncertainties. Journal of Energy Storage, 67, 107443.
— Larvaron, B., Clausel, M., Bertoncello, A., Benjamin, S., & Oppenheim, G. (2023). Chained Gaussian processes with derivative information to forecast battery health degradation. Journal of Energy Storage, 65, 107180.

Both exploit the presented background on Gaussian process methods. The first one focuses on the question of uncertainties quantification while the second focuses on the forecasting task.

2.2 Background on Gaussian processes methods

Gaussian processes have a long history. Their most famous introduction was in the field of stochastic processes theory in the 1940s with the works of Wiener [96] and Kolmogorov [97] with applications to time series. These methods were then used on applied fields such as mining by Krige [98], forestry by Matérn [99], and meteorology by Gandin [100]. These works inspired Matheron considered as the founder of geostatistics [101], naming the main method kriging in reference to Krig's research. Many fruitful developments followed until today, in mining but also more widely in many spatial data applications, and have been reviewed in [102] and [103].

Most current uses of Gaussian processes rely on the more recent developments of Rasmussen and Williams who introduced the Gaussian process regression [104]. It was developed quite independently from the geostatistical field, leaving the restriction to spatial data to consider generic inputs. It was part of the Bayesian machine learning community prolonging Neal's work on Bayesian neural network [105] and adapting the older results of O'Hagan [106]. Since then, the Gaussian processes have been used extensively in the machine learning community with a famous reference book from the authors [5].

However Gaussian process usage is not limited to the Gaussian process regression framework. More generally it can be seen as a tool to estimate one or several unknown functions, for a high variety of models. Among others, Williams described its use for binary or multi-class classification [107], Neal used a student distribution to improve the outlier modeling [108], Diggle adapted its use to model spatialized count data [109]. Simultaneously many developments tackled the scalability issue of Gaussian process methods [110]. As alternative to the standard but costly

Monte Carlo Markov chains methods many model approximation methods such as EP, Laplace approximation, or variational inference have been used. Recent progress in such approximation methods paves the way to more general Gaussian process frameworks coupling many different models and scalability issues [95, 111]. Among these methods, the use of chained Gaussian process [95], will permit us to significantly improve to the uncertainty quantification of battery health degradation.

2.2.1 Gaussian processes

Definition

Before presenting its use for regression tasks we need to define what is a Gaussian process. A *random process*, also called a random function or stochastic process, is a random variable whose realizations are functions. A *Gaussian process* is a particular case of random process, defined as the generalization of the multivariate Gaussian distribution. An example of realizations is displayed Fig. 2.2. For ease of presentation, figures will only display Gaussian processes depending of a univariate input, but they can also be multivariate.

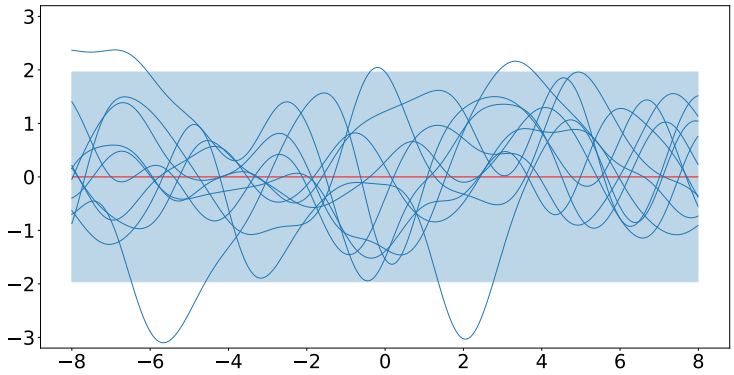


FIGURE 2.2 – Samples of a Gaussian process

In the case of a real value random variable y, we recall that a Gaussian random variable is defined by its mean, m and its variance, σ^2. We denote $y \sim \mathcal{N}(m, \sigma^2)$ and it has the following density

$$p(y|m, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-m)^2}{2\sigma^2}\right). \tag{2.1}$$

This definition extends to the multivariate case, imposing that any linear combination of the components should be a real random variable. In the case of a dimension of d, these random variables depend this time of a mean vector $\mathbf{m} \in \mathbb{R}^d$ and a covariance matrix $\Sigma \in \mathbb{R}^{d,d}$. It is denoted $\mathbf{y} \sim \mathcal{N}(\mathbf{m}, \Sigma)$ with the density

$$p(\mathbf{y}|\mathbf{m},\Sigma) = \left((2\pi)^d \det(\Sigma)\right)^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{m})^T\Sigma^{-1}(\mathbf{y}-\mathbf{m})\right). \tag{2.2}$$

Gaussian processes are an infinite dimension extension of the Gaussian distribution, moving from a vector $\mathbf{y}$ to a function f. Each of its values is indexed by an element of an infinite input $\mathcal{X}$ which is infinite. For each finite sub-collection of $\mathcal{X}$, the associate random vector is a multivariate Gaussian distribution. Explicitly, for any size n, and finite sub-collection $(\mathbf{x}_1, \cdots, \mathbf{x}_n)$ of $\mathcal{X}$, $(f(\mathbf{x}_1), \cdots, f(\mathbf{x}_n))$ is a multivariate Gaussian distribution. As for the finite-dimension case, the Gaussian process only depends on the two first moments. They now correspond to the *mean function* m, the mean at each input and the covariance function k, the covariance between each pairs of inputs, generally called the *kernel*. The process is denoted

$$f \sim \mathcal{GP}(m, k), \tag{2.3}$$

with

$$m(\mathbf{x}) = \mathbb{E}\left[f(\mathbf{x})\right] \tag{2.4}$$
$$k(\mathbf{x},\mathbf{x}') = \mathbb{E}\left[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))\right]. \tag{2.5}$$

As for deterministic functions the notions of continuity, and derivability, extend to random functions using a convergence adapted to random variables, generally the mean-square convergence. Considering for ease of presentation, a univariate Gaussian process, $f(t)$, it is continuous at a location a if the random variables $f(t)$ converges in mean square to the random variable $f(a)$ as t tends to a. If the limit exists, the derivative is defined as the mean square limit of the difference quotient,

$$\frac{df(t)}{dt} = \lim_{h\to 0} \mathbb{E}\left[\frac{f(t+h) - f(t)}{h}\right]^2. \tag{2.6}$$

This definition directly extends to random functions of several variables with partial derivatives, but also to higher-order derivatives (second, third derivatives...). Both continuity and derivability of a random process should not be confounded with the one of trajectories, even if both are related as we will see. Interestingly, if the limit exists, the derivative of a Gaussian process is also a Gaussian process [5], with the following properties [112],

$$\mathbb{E}\left[\frac{df(t)}{dt}\right] = \frac{d\mathbb{E}[f(t)]}{dt}, \tag{2.7}$$

$$\mathrm{Cov}\left[\frac{df(t)}{dt}, f(t')\right] = \frac{d}{dt}\mathrm{Cov}\left[f(t), f(t')\right], \tag{2.8}$$

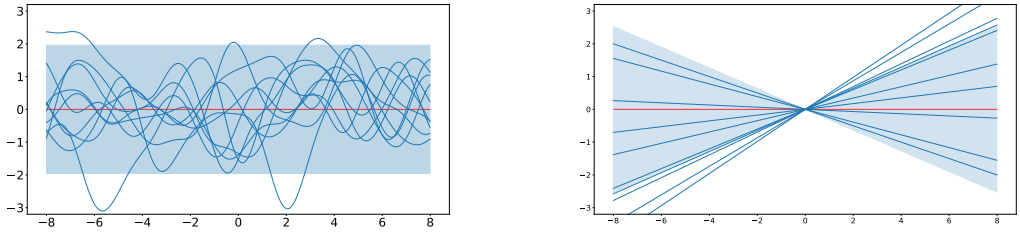
$$\mathrm{Cov}\left[\frac{df(t)}{dt}, \frac{df(t')}{dt'}\right] = \frac{d^2}{dtdt'}\mathrm{Cov}\left[f(t), f(t')\right]. \tag{2.9}$$

Specifying a Gaussian process model corresponds to choosing the mean function and the kernel. Both of them are generally parametric functions, whose parameters are called *hyper-parameters* since they do not directly define the function shape, but only define some global behaviors.

The mean function is often overlooked in the machine-learning community. Usually, it is considered either as a constant, or set as equal to the empirical mean. We will see in our applications that the choice of the mean matters, with a significant influence in extrapolation (for example forecasting). However, the kernel remains the most important element, and has to be carefully selected.

Kernel properties

A kernel is a symmetric and semi-definite function of two variables. It can be interpreted as a way to quantify the similarity of two inputs. Here we will explain the main properties of kernels, and the most popular examples are provided bellow.



(a) (b)

FIGURE 2.3 – Samples from a Gaussian process with (a) a stationary squared exponential kernel, (b) and a non stationary linear kernel

The first characteristic often used in practice is the *stationarity*. A kernel is said stationary if

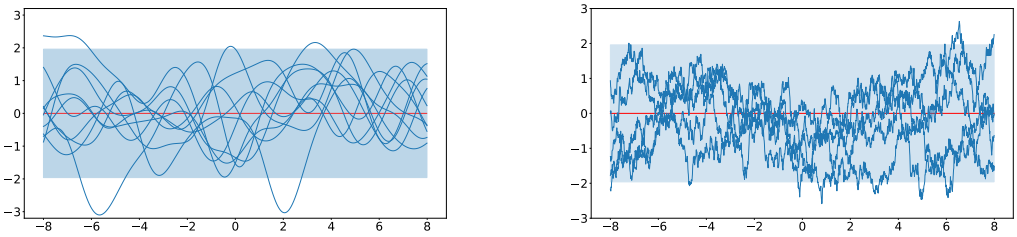
$$k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x} - \mathbf{x}'). \tag{2.10}$$

which implies that the kernel value is invariant up to a translation $k(\mathbf{x} + \mathbf{h}, \mathbf{x}' + \mathbf{h}) = k(\mathbf{x}, \mathbf{x}')$. Also, the properties of the Gaussian process do not depend on the particular location in the input space. Fig. 2.3 illustrates this with samples from Gaussian process with a squared-exponential kernel which is stationary with no particular location, unlike the non-stationary linear kernel, whose behavior depends on the distance to the origin. As can be seen in this example, it has an important consequence on the variance evolution. When one takes $\mathbf{x} = \mathbf{x}'$, $k(\mathbf{x}, \mathbf{x})$ corresponds to the variance of the process at location $\mathbf{x}$. With the stationarity assumption $k(\mathbf{x}, \mathbf{x}) = k(\mathbf{0})$, the variance is constant, independent of $\mathbf{x}$. This particular consequence is called homoscedasticity. We will see in the following, that this impose strong restrictions on the uncertainty quantification.

In the multivariate case, a stronger property of kernels is *isotropy*. A kernel is isotropic if

$$k(\mathbf{x}, \mathbf{x}') = k\left(|\mathbf{x} - \mathbf{x}'|\right). \tag{2.11}$$

That means that the covariance is also independent from direction, it only depends on the distance between the inputs. In practice, isotropy is often too restrictive and an *anisotropic* kernel is used instead. The simplest way of including anisotropy is to use a specific lengthscale for each direction as explained below.



(a) (b)

FIGURE 2.4 – Samples from a Gaussian process with (a) a squared exponential kernel, (b) an exponential kernel

Kernels also dictate different *regularity* to realizations. The continuity and the existence of derivatives of a Gaussian process is completely determined by the choice of kernel. Explicitly, in the univariate stationary case, the kernel is only a function of the distance h between the inputs and $d^r f(t)/d^r t$ exists for any t only if $d^{2r}k(h)/dh^2$ exists and is finite at $h = 0$. Fig. 2.4 compares predictions realizations for a process with a squared-exponential kernel which is infinitely differentiable and an exponential kernel which is said continuous but not differentiable.

Kernel examples

The most often used kernels belong to the Matérn family. These kernels are stationary. In their isotropic version, they are defined as follows

$$k_\nu(\mathbf{x}, \mathbf{x}') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{|\mathbf{x} - \mathbf{x}'|}{l} \right)^\nu K_\nu \left(\sqrt{2\nu} \frac{|\mathbf{x} - \mathbf{x}'|}{l} \right), \tag{2.12}$$

where K_ν is a modified Bessel function, Γ is the Gamma function, ν specifies the kernel smoothness, and σ^2 and l are the variance and lenghtscale hyper-parameters.

In fact, this kernel is considered for several specific values of ν. For $\nu = 1/2$ we obtain the exponential kernel,

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{|\mathbf{x} - \mathbf{x}'|}{l} \right) \tag{2.13}$$

also called the Ornstein-Uhlenbeck kernel in the univariate case. For $\nu = 3/2$ and $\nu = 5/2$ we get respectively the Matérn $3/2$

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \left(1 + \frac{\sqrt{3}|\mathbf{x} - \mathbf{x}'|}{l}\right) \exp\left(-\frac{|\mathbf{x} - \mathbf{x}'|}{l}\right) \tag{2.14}$$

and the Matérn 5/2 kernel

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \left(1 + \frac{\sqrt{5}|\mathbf{x} - \mathbf{x}'|}{l} + \frac{5|\mathbf{x} - \mathbf{x}'|^2}{3l}\right) \exp\left(-\frac{|\mathbf{x} - \mathbf{x}'|}{l}\right) \tag{2.15}$$

Finally, in the case $\nu \to \infty$ we obtain the popular squared-exponential kernel

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{|\mathbf{x} - \mathbf{x}'|^2}{2l^2}\right) \tag{2.16}$$

also called the Radial basis function (RBF) kernel. The ν parameter controls the smoothness of the kernel such that these four kernels are ordered by increasing smoothness. In many applied settings the squared-exponential kernel is a default choice. However the three are interesting alternatives and can provide more realistic predictions in case of less smooth functions.

When applicable, a closed form of Eq. (2.17) and (2.18) can be obtained for these kernels. In particular for the squared-exponential kernel [112], considering again the univariate case,

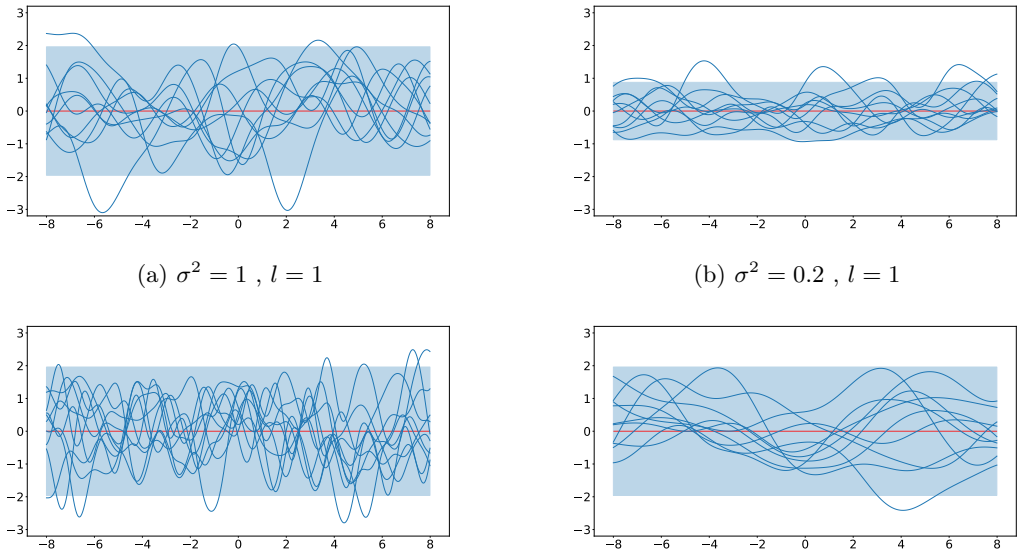
$$\text{Cov}\left[\frac{df(t)}{dt}, f(t')\right] = \sigma^2 \exp\left(-\frac{(t - t')^2}{2l^2}\right) \times (-l^{-2}(t - t')) \tag{2.17}$$

$$\text{Cov}\left[\frac{df(t)}{dt}, \frac{df(t')}{dt}\right] = \sigma^2 \exp\left(-\frac{(t - t')^2}{2l^2}\right) \times \left(-l^{-2}\left(1 - l^{-2}(t - t')^2\right)\right) \tag{2.18}$$

And similar formula can be derived for second-order derivatives. These derivations will be important to set constraints on derivative of the fitted model in Sec. 2.3.3.

As it is the case for many other kernels, all these kernels depend on two hyperparameters denoted l and σ^2, the first one called the lengthscale and the second one the variance. We see that the lengthscale controls the decrease of the covariance with the distance between inputs. When the distance between two inputs is higher than say $2l$ the covariance of the corresponding variables are close to zero so nearly independent. Intuitively, it controls the frequency of the process, high lenghtscales correspond to long variations whereas small lenghtscales lead to short variations. On the other hand, the variance parameter corresponds to the amplitude of the process, it is the value of the variance of the Gaussian process, which is constant for stationary kernels. The influence of these two hyperparameters is illustrated in Fig. 2.5.

In the above version the Matérn kernel is isotropic. In practice, this assumption is too restrictive so one generally adapt the kernel with a particular lenghtscale for each dimension. In the case of the squared-exponential kernel we obtain



(a) $\sigma^2 = 1$, $l = 1$ (b) $\sigma^2 = 0.2$, $l = 1$

(c) $\sigma^2 = 1$, $l = 0.5$ (d) $\sigma^2 = 1$, $l = 2$

FIGURE 2.5 – Samples from a Gaussian process with a Gaussian kernel using different values of variance and lengthscale hyper-parameters

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\sum_{j=1}^{p} \frac{|\mathbf{x}_j - \mathbf{x}'_j|^2}{2\mathbf{l}_j^2}\right), \tag{2.19}$$

replacing the real lengthscale l by a lengthscale vector $\mathbf{l}$.

Remark In the machine learning community the inverse value of the lengthscale is sometimes interpreted as a measure of importance for each variable and called ARD, Automatic Relevance Determination [104]. If a lengthscale value is really large, predictions in that direction will be quite flat so we can think that the variable is not important. But we have to be careful since "large" depends on the value of the variance parameter. A large lengthscale associated with a high variance can correspond to a trend effect which can be of high interest.

Depending on the context other kernels can be used such as the periodic kernel

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{2\pi \sin^2(|\mathbf{x} - \mathbf{x}'|)}{l^2}\right) \tag{2.20}$$

often used to capture periodicity in time series. The constant kernel

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2, \tag{2.21}$$

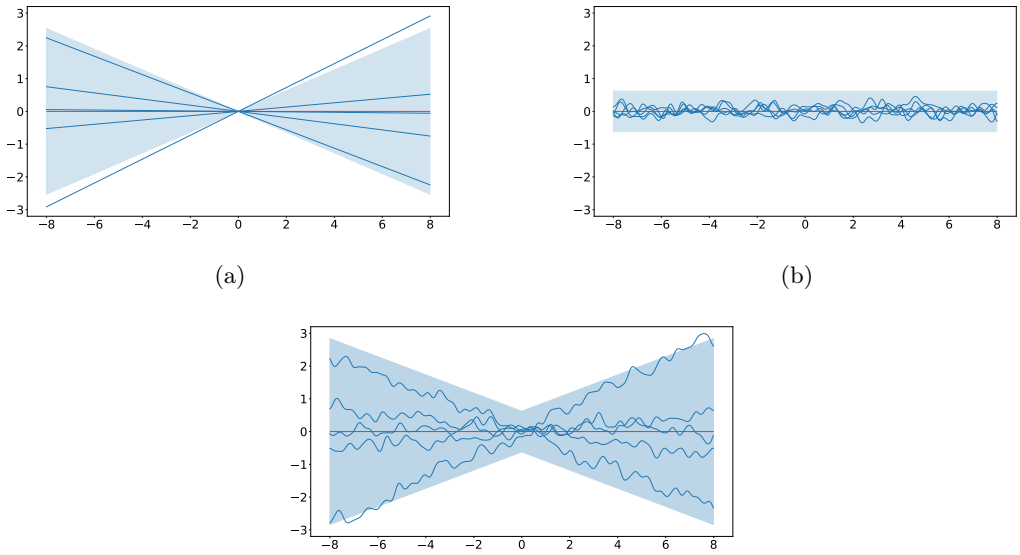
provides constant realization functions. Finally, non-stationary processes can be obtained thanks to the polynomial kernel

$$k(\mathbf{x}, \mathbf{x}') = (\sigma^2 \mathbf{x}^T \mathbf{x}' + \gamma)^p \tag{2.22}$$

including the linear kernel for $p = 1$. More details on the properties of these kernels can be found in section 4.2 of [5].

Kernel design

This corresponds to the basic kernels. In practice, a powerful property is that they can be combined to create more complex ones. This has the two benefits. First, it may improve the fit of the model. Second, it enhances its interpretability. The most standard combinations are products and sums of kernels. In this work, we will mainly consider sums of kernels. For the sum, considering two independent Gaussian processes, the sum is also a Gaussian process with as kernel the sum of the two kernels. More details about kernel design can be found in chapter 2 of Duvenaud's thesis [113] also summarized in his popular kernel cookbook [114].



(a) (b)

(c)

FIGURE 2.6 – Samples from a Gaussian process with (a) a linear kernel, (b) a squared-exponential kernel and (c) combination of the two kernels

Explicitly, we consider a Gaussian process $f \sim \mathcal{GP}(m, k)$ and we suppose that

$$k(\mathbf{x}, \mathbf{x}') = k_1(\mathbf{x}, \mathbf{x}') + k_2(\mathbf{x}, \mathbf{x}') \tag{2.23}$$

and

$$m(\mathbf{x}) = m_1(\mathbf{x}) + m_2(\mathbf{x}) \tag{2.24}$$

It is equivalent to

$$f(\mathbf{x}) = f_1(\mathbf{x}) + f_2(\mathbf{x}), \tag{2.25}$$

with $f_1 \sim \mathcal{GP}(m_1, k_1)$ and $f_2 \sim \mathcal{GP}(m_2, k_2)$ both being independent Gaussian processes.

This decomposition opens many possibilities in model specification. For example, the effect of a variable can be decomposed between a short and a long range effect with each its own lengthscale hyper parameter. Fig. 2.6 illustrates a simple example of this, mixing a linear part with small variance and lenghtscale squared-exponential kernel term. In the case of several variables it can also be used to build additive models [115], with a different function per variable.

We will present the application of this method in our use case below. Detailed and inspiring examples of this approach can be seen in Rasmussen's book [5], section 5.4.3 on the Mauna Loa atmospheric CO2 evolution and with a fully Bayesian perspective in Gelman's reference book [116] at section 21.2 on the daily number of births.

2.2.2 The Gaussian process regression framework

Presentation of the model

We can now see how the Gaussian processes can be used for regression, starting from the popular framework of the Gaussian process regression [5]. Considering a typical regression problem we want to model a random value y as a function of inputs $\mathbf{x}$ and we usually write

$$y(\mathbf{x}) = f(\mathbf{x}) + \varepsilon, \tag{2.26}$$

f being an unknown function modeling the mean function of y and ε a centered experimental noise, independent of f and itself for each $\mathbf{x}$. It is a common choice to assume ε to be Gaussian with constant variance $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$.

The Gaussian process regression is a Bayesian method. f is seen as the unknown parameter of the model and a Gaussian process prior it used

$$f \sim \mathcal{GP}(m, k). \tag{2.27}$$

More precisly, this is refered as a *nonparametric Bayesian* method since no particular shape is specified for f, only general properties through the kernel. The model complexity will depend on the number of training data points.

Moreover, Eq. (2.26) with the assumption of independent Gaussian noise of fixed variance provides the *likelihood* of the model, the distribution of the interest variable given the model

parameters. Given a matrix of inputs, $X = (\mathbf{x}_1, \cdots, \mathbf{x}_n)$ we write $\mathbf{y} = y(X) = (y(\mathbf{x}_1), \cdots, y(\mathbf{x}_n))$ and $\mathbf{f} = f(X)$, we get

$$p(\mathbf{y}|\mathbf{f}) = \mathcal{N}(\mathbf{y}|\mathbf{f}, \sigma_\varepsilon^2 I) \tag{2.28}$$

$$= \prod_{i=1}^{n} \mathcal{N}(\mathbf{y}_i|\mathbf{f}_i, \sigma_\varepsilon^2) \tag{2.29}$$

where f parametrizes the mean of the interest variable.

Computation of hyperparameters

To define the Gaussian process prior distribution we need to choose the mean and the kernel function, taking into account the properties detailed above. Yet, both of them depend on hyperparameters $\boldsymbol{\tau}$, typically the variance and lengthscale terms, that are not known *a priori* and need to be inferred. Different approaches are available to learn these hyperparameters. In a fully Bayesian framework, this problem is solved using a hierarchical model. In the likelihood, the law of $\mathbf{y}$ depends on parameter $\mathbf{f}$ whose itself relies on $\boldsymbol{\tau}$ on which one can set a prior distributions, which can be called hyper-prior. Then using classical approximation methods (typically MCMC) we can estimate the posterior law of $\boldsymbol{\tau}$ and $\mathbf{f}$.

Even if this approach reflects the uncertainty linked to the estimation, it is often considered too complex computationally (scales badly as n increases) and other methods are preferred. The two main other methods are optimization of the marginal likelihood and Cross-validation. We will concentrate here on marginal likelihood optimization.

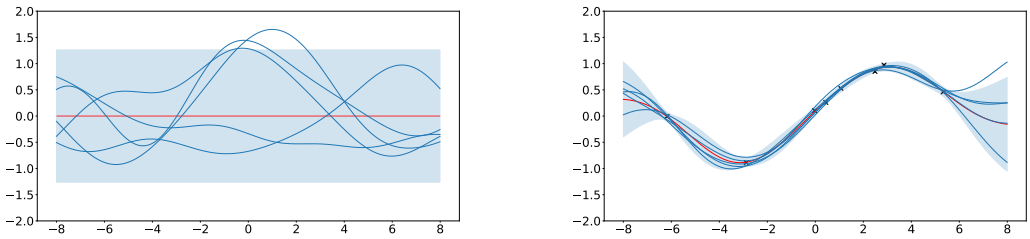
In the Gaussian process regression, the marginal likelihood (or evidence) corresponds to

$$\log p(\mathbf{y}|X, \boldsymbol{\tau}) = -\frac{1}{2}\mathbf{y}^T K_y^{-1}\mathbf{y} - \frac{1}{2}\log\det(K_y) - \frac{n}{2}\log 2\pi \tag{2.30}$$

where $K_y = K_f + \sigma_\varepsilon^2 I = K(X, X) + \sigma_\varepsilon^2 I$, with for $A \in \mathbb{R}^{n_1,p}$, $B \in \mathbb{R}^{n_2,p}$, $K(A, B) \in \mathbb{R}^{n_1,n_2}$ with $(K(A, B))_{ij} = k(A_i, B_j)$, and I is the identity matrix of size n. It is called marginal since it corresponds to the likelihood $p(\mathbf{y}|\mathbf{f})$ integrated over the values of $\mathbf{f}$. As explained in section 5.2 of [5], this is an approximation of the Bayesian approach called type II maximum likelihood (ML-II). The authors highlight the fact that it can lead to overfitting but is sufficient in most cases. In practice, maximization is done through iterative gradient-based methods.

Posterior distribution

Once the hyperparameters computed, the model is fully specified. Since GPR is a Bayesian model, predictions are then made by computing the *posterior distribution*, the law of the parameter given the training data. In our context, we have a vector of observations $\mathbf{y}$, at n locations contained in X. Considering n^* new inputs concatenated in X^* we are interested in the law of $\mathbf{f}^* = f(X^*)$



(a) (b)

FIGURE 2.7 – Comparison of (a) the prior distribution, (b) the posterior distribution

whose posterior is $p(\mathbf{f}^*|\mathbf{y})$. The strength of the Gaussian process regression is that the Gaussian prior and the Gaussian likelihood lead to a Gaussian posterior law with an analytic expression of the mean and the variance, they are conjugated. Explicitly, the posterior has the following expression

$$p(\mathbf{f}^*|\mathbf{y}) = \mathcal{N}(\overline{\mathbf{f}}^*, \mathrm{cov}(\mathbf{f}^*)) \tag{2.31}$$

with

$$\overline{\mathbf{f}}^* = \mathbf{m}^* + K(X^*, X)\left[K(X,X) + \sigma_\varepsilon^2 I\right]^{-1}(\mathbf{y} - \mathbf{m}) \tag{2.32}$$

$$\mathrm{cov}(\mathbf{f}^*) = K(X^*, X^*) - K(X^*, X)\left[K(X,X) + \sigma_\varepsilon^2 I\right]^{-1} K(X, X^*), \tag{2.33}$$

where I is the identity matrix of size n, $\mathbf{m} = m(X)$ and $\mathbf{m}^* = m(X^*)$.

The posterior mean (2.32) is the predicted mean of f at positions X^*, it is the same as the one of the interest variable y. The posterior covariance matrix (2.33) contains the predicted variance at each position. It permits us to quantify uncertainties and compute confidence intervals of the interest variable through the addition of the variance of the independent noise.

It is important to distinguish between the properties of the prior and those of the posterior. Considering a prior with a constant mean function and a stationary kernel Fig. 2.9 compares samples from the prior to some of the posterior, once some training data have been obtained. We see in (2.32) and Fig. 2.7b that it does not impose a constant predicted mean since it is a weighted average of the observations. Similarly, the predicted variance (2.33) is not constant in general. In the case of a stationary kernel, it is equal to the constant prior variance minus a term depending on the distance between the position of prediction and the position of all data points. For the kernels selected previously, the second term decreases toward zero as this distance increases, and the predicted variance increases.

2.2.3 The Chained Gaussian process framework

Presentation of the model

The GPR framework is popular and is often considered to quantify epistemic uncertainties, linked to the lack of data. However, we will see in our study that it presents important limitations regarding the aleatoric uncertainty quantification, linked to the natural variability of the process. Thus we present here a generalization of this framework, the chained Gaussian process (CGP) framework from Saul et al. [95] that will permit us to overcome these limits. While keeping the Gaussian process as prior over unknown function, this framework proposes to use any distribution p as the likelihood (not only Gaussian) potentially depending on several functions $f_1, \cdots, f_c. \cdots, f_C$. The only restriction on the likelihood is that it should be separable. It means that, for any finite collection of observations X, writing $\mathbf{f}_c = f_c(X)$, we have

$$p(\mathbf{y}|\mathbf{f}_1, \cdots, \mathbf{f}_C) = \prod_{i=1}^{n} p(\mathbf{y}_i|\mathbf{f}_{1,i}, \cdots, \mathbf{f}_{C,i}) \tag{2.34}$$

with p being any probability density whose parameters are defined thanks to $\mathbf{f}_1, \cdots, \mathbf{f}_C$.

A Gaussian process prior is set on each component, $f_c \sim \mathcal{GP}(m_c, k_c)$ and they are all assumed independent. We see that the GPR is a special case of the CGP model with one component, and a Gaussian likelihood.

The above presentation of the CGP has been simplified. CGP is a generalization of the sparse Gaussian processes [117], whose focus was in the reduction of computational complexity. This is done, through the introduction of inducing variables summarizing the Gaussian processes at positions $Z \in \mathbb{R}^{mp}$, $\mathbf{u}_1 = f_1(Z), \cdots, \mathbf{u}_C = f_C(Z)$ with generally, $m << n$ to reduce computations. However, in the context of batteries degradation, datasets have a limited size, and the computational cost reduction is not the main concern of our work. So we considered, the CGP framework with $Z = X$ to ensure minimal approximation error. This is equivalent to the above presentation with $\mathbf{u}_c = \mathbf{f}_c$.

The presented chained Gaussian processes model offers more flexibility in the possible model. However, in general, prior and likelihood do not conjugate anymore, so no close form of the posterior is available. Marginal likelihood to optimize hyper-parameters cannot be computed neither. In such cases, computations can either be done through Monte Carlo Markov Chain (MCMC) converging toward the true posterior distribution or approximated. As in this context, MCMC methods can be slow we prefer approximation methods. Many approximation methods have been proposed in the context of the Gaussian process, the most well-known being Laplace approximations [118], Expectation Propagation (EP) [119] and variational inference [120]. For the CGP framework the authors relied on variational inference. This approximation method is gaining popularity in the recent years in the machine learning community [121, 122], thanks to its versatility and relative ease of use.

Model approximation though variational inference

Variational inference [123] is an approximation method relying on an optimization problem. The goal is to approximate the true unknown posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$. The first step consists in selecting a variational family $\mathcal{Q}$, a set of distributions where we will search the approximate distribution. Then an optimization step is performed to find the distribution $q(\boldsymbol{\theta})$, the closest to the true posterior. In variational method, closeness is defined by the Kullback-Leibler divergence. So the goal is to find the distribution $q \in \mathcal{Q}$ which minimizes

$$\mathrm{KL}(q(\boldsymbol{\theta})||p(\boldsymbol{\theta}|\mathbf{y})) = \mathbb{E}_q\left[\log\left(\frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\mathbf{y})}\right)\right] = \int \log\left(\frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\mathbf{y})}\right) q(\boldsymbol{\theta})\, d\boldsymbol{\theta}. \tag{2.35}$$

The family of distribution $\mathcal{Q}$ is generally chosen to ease computations. For example, a typical condition is to impose independence between components of $\boldsymbol{\theta}$, $q(\boldsymbol{\theta}) = \prod_j q(\boldsymbol{\theta}_j)$, we speak of the mean-field variational family. Variational inference remains an approximation method since in general, the true posterior is not included in $\mathcal{Q}$. In practice, the variational family is chosen as a parametric family for each component, e.g. typically Gaussian distributions for continuous variables. Those parameters are called the variationnal parameters. So the variational inference resumes to optimizing (2.35) with respect to these parameters.

However, the Kullback-Leibler divergence cannot be minimized directly, since

$$\mathrm{KL}(q(\boldsymbol{\theta})||p(\boldsymbol{\theta}|\mathbf{y})) = \mathbb{E}_q[\log q(\boldsymbol{\theta})] - \mathbb{E}_q[\log p(\mathbf{y}, \boldsymbol{\theta})] + \log p(\mathbf{y}), \tag{2.36}$$

and $\log p(\mathbf{y})$, the log-evidence is unknown, being the original reason why we cannot compute $p(\boldsymbol{\theta}|\mathbf{y})$. However, since this term does not depend of q, so we can consider an equivalent problem, maximizing the evidence lower bound (ELBO),

$$\mathrm{ELBO} = \mathbb{E}_q[\log p(\mathbf{y}, \boldsymbol{\theta})] - \mathbb{E}_q[\log q(\boldsymbol{\theta})]. \tag{2.37}$$

By definition the Kullback-Leibler divergence remains positive so we can check that it is the lower bound of the (log) evidence.

There are several methods to maximize the ELBO. In the popular case of mean-field variational family, the usual method is coordinate ascent variational inference [124]. However, this does not scale well to massive data. This is the reason why stochastic variational inference [125], which uses gradient-based methods in a stochastic way has become more and more popular for machine learning applications.

In the case of the Chained Gaussian process model we write $\boldsymbol{\theta}$ the concatenation of all parameters, $\boldsymbol{\theta} = (\mathbf{f}_c)_{c\in[\![1,C]\!]}$. A variational family $\mathcal{Q}$ is defined, imposing for any of its elements q to respect the following condition,

$$q(\boldsymbol{\theta}) = \prod_{c=1}^{C} q(\mathbf{f}_c) \tag{2.38}$$

imposing independence between components. Additionally each $q(\mathbf{f}_c)$ is modeled by a multivariate Gaussian distribution, $q(\mathbf{f}_c) = \mathcal{N}(\mathbf{f}_c|\boldsymbol{\mu}_c, S_c)$. So the variational parameters, to be optimized are $(\boldsymbol{\mu}_c, S_c)_{c\in[\![1,C]\!]}$.

Thus, rearranging the ELBO in the following way to ease computation

$$\begin{aligned}\text{ELBO} &= \mathbb{E}_q[\log p(\mathbf{y}|\boldsymbol{\theta})] - \mathbb{E}_q\left[\log \frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta})}\right] &(2.39)\\ &= \mathbb{E}_q[\log p(\mathbf{y}|\boldsymbol{\theta})] - \text{KL}(q(\boldsymbol{\theta})||p(\boldsymbol{\theta})) &(2.40)\end{aligned}$$

exploiting the separability of the likelihood, we finally get

$$\text{ELBO}_{\text{CGP}} = \sum_{i=1}^{N} \int q(\mathbf{f}_{1,i}) \cdots q(\mathbf{f}_{C,i}) p(\mathbf{y}_i|\mathbf{f}_{1,i}, \cdots, \mathbf{f}_{C,i}) d\mathbf{f}_{1,i} \cdots d\mathbf{f}_{C,i} + \sum_{c=1}^{C} \text{KL}(q(\mathbf{f}_c)||p(\mathbf{f}_c)). \quad (2.41)$$

The second term can be computed in closed form as the Kullback-Leibler divergence of two Gaussian distributions. Special cases are computable for the first term but they cannot be computed exactly in the general case. In practice, the integral is approximated but it becomes more difficult for higher numbers of components, it is generally assumed low, 2 or 3 components.

Remark The above detailed variational procedure is built to approximate the variational distribution. However, in Gaussian processes-based methods we also look for an optimization of the hyperparameters of the Gaussian prior distribution. As explained by Salimbeni *et al* [126], section 5.3, a common practice is to optimize the ELBO regarding both variational and Gaussian process prior hyperparameters. In that case, strictly speaking, optimizing ELBO may lead to sub-optimal solutions since $\log p(\mathbf{y})$ is not constant. Still, it is reported to work well in practice, and we followed this approach in the following.

2.3 Gaussian process models of the battery health degradation

The previous section provided all the necessary background to build regression models relying on Gaussian process priors. We will now present the particular model that we proposed for the battery health degradation modeling with uncertainties. Relying extensively on the Aachen dataset, we will put a particular focus on the estimation of the cell-to-cell variations, a major source of aleatoric uncertainty. As Gaussian process methods are relatively computational intensive, some additional implementation details are provided in Appendix B.

2.3.1 Explicit modeling of the cell-to-cell variations with a Gaussian process regression

Proposed Gaussian process regression

The GPR framework is already popular in battery applications [127, 128, 80], notably for its ability to quantify uncertainties. In our application, a standard approach would consider a model taking the following form

$$y(t, b) = f(t, b) + \varepsilon, \tag{2.42}$$

with the two input variables being the time t and the battery indicator b. Yet a limitation of such a model is that it does not explicitly express the cell-to-cell variations. f includes all the uncertainties except the measurement error, so it mixes the epistemic uncertainty linked to the mean estimation with the aleatoric uncertainty.

We want a model taking explicitly into account the cell-to-cell variations. Adapting ideas from Myllymäki *et al.* [129], we precise our model, writing

$$f(t, b) = \mu(t) + y_{centered}(t, b) \tag{2.43}$$

with μ the shared trend of SoH degradation and $y_{centered}$ the noise-free SoH deviation from the mean, i.e., the cell-to-cell variations. Now the model considers three sources of uncertainties, an epistemic uncertainty linked to the estimation of the mean μ, and two aleatoric uncertainties, the cell-to-cell variations with $y_{centered}$ and the measurement error with ε.

This model can be directly learned thanks to kernel design methods described in the previous section. With independent GP priors on μ and $y_{centered}$, $\mu \sim \mathcal{GP}(m_\mu, k_\mu)$, $y_{centered} \sim \mathcal{GP}(m_0, k_0)$, it is equivalent to use (2.42) with prior $f \sim \mathcal{GP}(m, k) = \mathcal{GP}(m_\mu + m_0, k_\mu + k_0)$, in the following we will directly work with such f.

For the trend μ we choose m_μ constant equal to the empirical mean of the data. Concerning k_μ, since the degradation trend is often modeled as an infinitely derivable function, we choose a Squared-exponential kernel. For $y_{centered}$, we suppose that the time evolution of two different batteries are independent. To impose that, we took $k_0(\mathbf{x}, \mathbf{x}') = \hat{k}_0(t, t')I(b = b')$ with I an indicator function. We choose a Matérn $3/2$ or Matérn $5/2$ kernel for $\hat{k}_0$ since the SE kernel must be too smooth to model individual variations. Finally, by construction, $y_{centered}$ has a constant mean equal to zero.

In this model, detailed information can be obtained on each of its components, with the computation of the posterior law of μ and $y_{centered}$. For example, $\mathbf{y}^*_{centered} = y_{centered}(X^*)$ has the following posterior law

$$p(\mathbf{y}^*_{centered}|\mathbf{y}) = \mathcal{N}(\overline{\mathbf{y}}^*_{centered}, \mathrm{cov}(\mathbf{y}^*_{\mathrm{centered}})) \tag{2.44}$$

with

$$\overline{\mathbf{y}}^*_{centered} = K_0(X^*, X)\left[K(X,X) + \sigma_\varepsilon^2 I\right]^{-1}\mathbf{y} \tag{2.45}$$

$$\text{cov}(\mathbf{y}^*_{centered}) = K_0(X^*, X^*) - K_0(X^*, X)\left[K(X,X) + \sigma_\varepsilon^2 I\right]^{-1} K_0(X, X^*), \tag{2.46}$$

Qualitative analysis

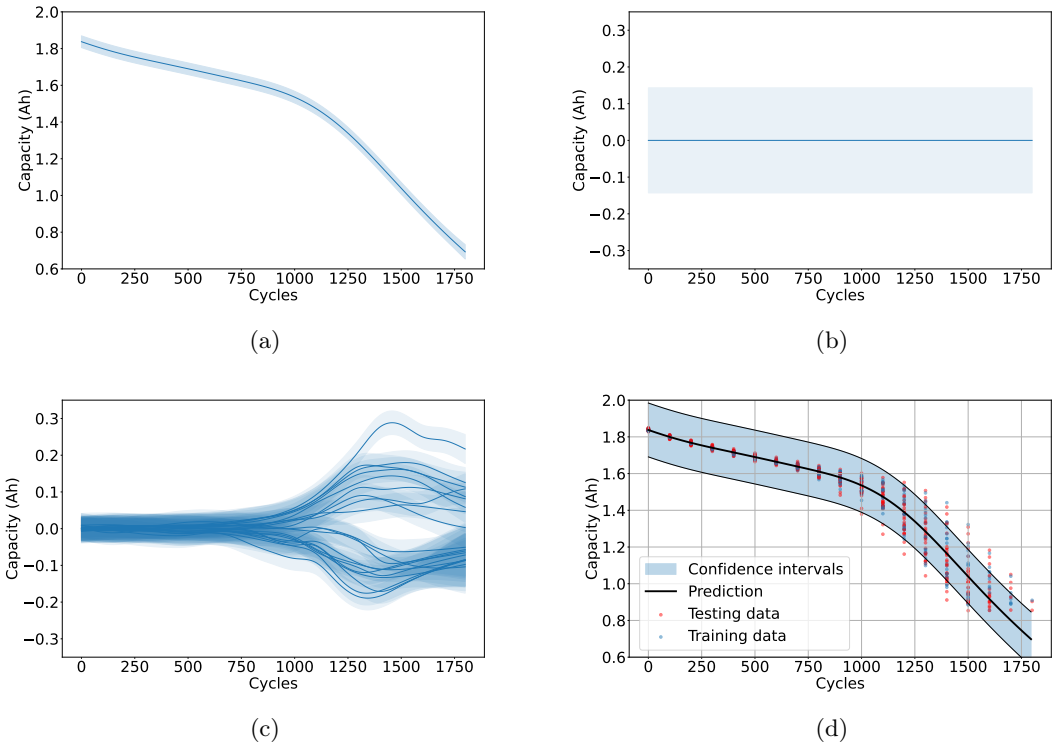


FIGURE 2.8 – Qualitative results with CGP (a) prediction of the trend μ (b) prediction of cell-to-cell standard deviation with CGP compared to its empirical estimation and its constant estimation for GPR (c) $y_{standard}$ prediction (d) complete capacity prediction.

We applied this model to the Aachen dataset. And Fig. 2.8 displays the predictions for each of its components with 95% confidence intervals. The plotted results used the Matérn 3/2 kernel for $y_{centered}$. We performed a similar experiment with a Matérn 5/2 kernel, but no notable difference was observed.

The average evolution plotted in Fig. 2.8a is decreasing with a reducing slope at the first cycles but with a significant acceleration around cycle 1000. The non-parametric fitting of the mean is interesting to detect these changes automatically. Associated confidence intervals have almost a constant width. This is due to the stationary kernel, regularly spread data points, and a long length scale.

We then plotted in Fig. 2.8b predictions for $y_{centered}$ for a new battery. As we explained in the

previous section, it has a constant zero mean and a constant standard deviation. The standard deviation is equal to 0.07 Ah which is higher than the one of the posterior law of μ approximately constant equal to 0.02 Ah and than the negligible random noise which has a standard deviation around 1.10^{-4} Ah. The relatively low variance of μ must be linked to the important number of batteries used for training. The negligible experimental noise is a specificity of this dataset which may be due to the authors' pre-processing of the data. This allows us to focus on cell-to-cell variations.

The comparison between Fig. 2.8b with Fig. 2.8c is quite revealing. The latter contains the predictions of $y_{centered}$ for batteries in the training set. We see a clear difference between the predicted behavior for a new battery and the actual behavior of observed batteries. For training data, all curves are close to zero at the first cycles but then start to move away at cycles 800/1000. This behavior is inconsistent with a constant variance and illustrates the increase of cell-to-cell variations.

Finally, complete predictions of capacity for a new battery can be seen in Fig. 2.8d. We added the training data in blue and the testing data in red. We see the already observed trend, which seems to fit well with the training and testing data. On the other hand, we see that the confidence intervals have an almost constant width and do not fit the data. They are too broad in the first cycles and probably too narrow in the last cycles.

The low quality of fit of the variance is a consequence of the stationarity assumption and its homoscedasticity consequence. This can be understood by going back to Eq. (2.46), providing the variance of the cell-to-cell variation term. For a new battery b^*, it is in fact only equal to the first term, the second term being null because of the independence with training data. Yet, the first term corresponds to the stationary prior imposing homoscedasticity. It demonstrates that the use of a stationary prior is not suited to learning the evolution of cell-to-cell variations.

2.3.2 A non-parametric modeling of the cell-to-cell variations with the Chained Gaussian processes

Proposed model

We see that there is a need to take into account the time evolution of the cell-to-cell variations. To highlight the constant variance assumption of the GPR model with a stationary kernel, we can rewrite it,

$$y(t,b) = \mu(t) + \sigma_0 y_{standard}(t,b) + \varepsilon \tag{2.47}$$

with $y_{standard}$ the scaled (constant unitary variance) version of $y_{centered}$. To take into account the evolution of the cell-to-cell variations, our goal is to learn the following model

$$y(t,b) = \mu(t) + \sigma_0(t) y_{standard}(t,b) + \varepsilon, \tag{2.48}$$

replacing the constant σ_0 by a function of time. This time we will set the homoscedastic

hypothesis on $y_{standard}$, which is no more restrictive since $y_{standard}$ has a constant variance by design.

The most direct choice to learn such a model would be to use a GPR with a nonstationary kernel [130]. An increasing variance can be obtained thanks to linear or polynomial kernels and Aitio *et al* [131] proposed the use of a Wiener kernel in a similar setting. However, these kernels still impose restrictive hypotheses on uncertainties with generally a polynomial evolution. We do not have any precise prior idea of the σ_0 function and we would like it to be estimated as precisely as possible. Also, we choose to model it nonparametrically without constraint on its shape, through the use of a Gaussian process prior, as proposed in [132]. Such a model cannot be learned with the Gaussian process regression framework, since it requires a nonlinear combination of Gaussian processes. That is why we introduced the Chained Gaussian process framework which can handle models depending non-linearly on several Gaussian processes prior.

Explicitly, Adams *et al.* [132] proposed to write $\sigma_0(t) = P(\eta(t))$, where P is a deterministic function to ensure positivity of the standard deviation, exponential or softplus function for example, and we set a GP prior on η. In this model, the parameters are the functions μ, η and $y_{standard}$. Considering a finite number of locations we thus have the following likelihood

$$p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\eta}, \mathbf{y}_{standard}) = \prod_{i=1}^{n} \mathcal{N}(\boldsymbol{\mu}_i + P(\boldsymbol{\eta}_i)\mathbf{y}_{standard,i}, \sigma_\varepsilon^2), \tag{2.49}$$

where $\boldsymbol{\mu} = \mu(X)$, $\boldsymbol{\eta} = \eta(X)$, $\mathbf{y}_{standard} = y_{standard}(X)$, neglecting the battery indicator for μ and η. Since this likelihood is separable, the CGP framework can be applied as presented above.

Regarding the priors, we keep the similar prior than for the GPR model for μ, and $y_{standard}$ is modeled similarly to $y_{centered}$ but with a variance hyperparameter fixed to 1 by construction. For the new component η we choose a constant mean function m_η to be estimated and we suppose that this term is smooth with k_η a SE kernel.

Qualitative results

We can now apply the CGP model to the Aachen dataset and compare the performances with the one obtained for the GPR model. The previous section illustrated the limits of a homoscedastic model, we can now observe the improvement with a heteroscedastic model. We displayed the results for a Matérn 3/2 kernel for $y_{standard}$ and an exponential transform. Replacing the kernel with a Matérn 5/2 or using a softplus transform does not change the results.

Prediction of the global trend is plotted in Fig. 2.9a. It can hardly be distinguished from the previous case. But now it seems that associated uncertainties are much smaller, almost negligible. In Fig. 2.9b we plotted predictions for $\sigma_0(t)$ in black. We displayed associated confidence intervals but they are so narrow that we can hardly observe them on the graph. For comparison, we added the constant standard deviation σ_0 predicted in the previous section, and the empirical standard deviation at each time computed on the complete dataset (all batteries and every 5 cycles). Empirical standard deviation also contains variability due to experimental noise, but since we concluded previously that it was almost negligible, it will not affect the interpretation.

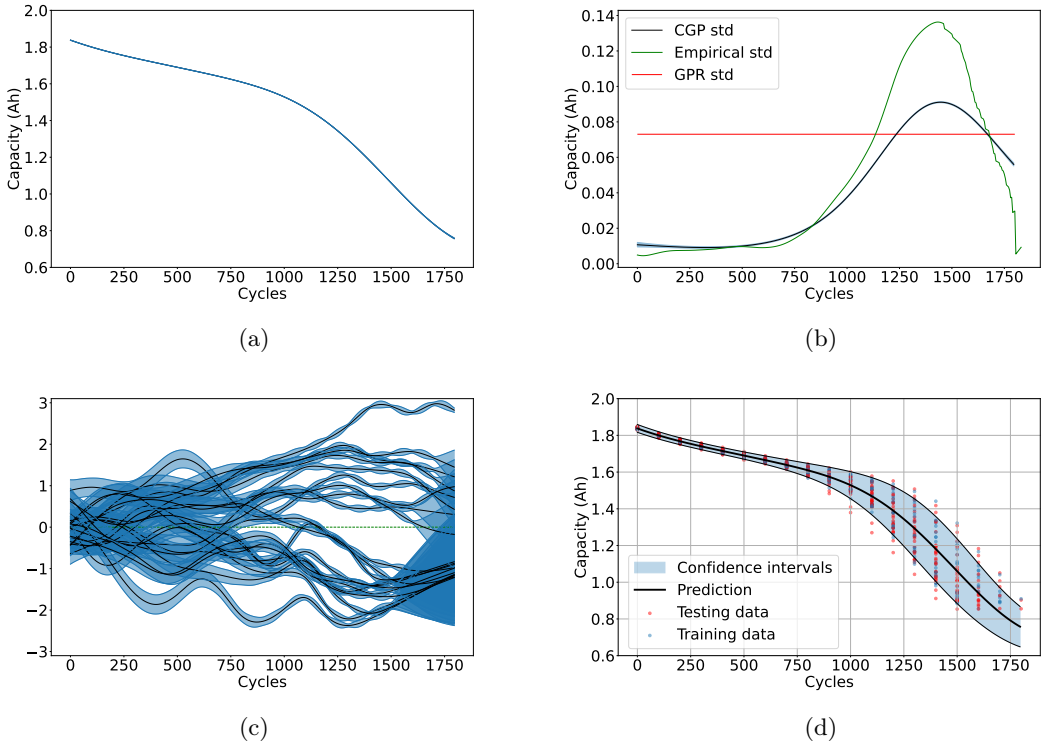


FIGURE 2.9 – Qualitative results with CGP (a) prediction of the trend μ (b) prediction of cell-to-cell standard deviation with CGP compared to its empirical estimation and its constant estimation for GPR (c) $y_{standard}$ prediction (d) complete capacity prediction.

For the 700 first cycles, the empirical standard deviation is low and almost constant. The CGP fits well with the empirical standard deviation, whereas as we saw previously, the GPR largely overestimates it. Then around cycle 700, the empirical standard deviation starts to increase quickly. We see a similar behavior for the CGP model even if it does not increase as fast. After cycle 1100 the GPR standard deviation underestimates the empirical value. The CGP underestimates it as well but the error is smaller.

Surprisingly, there is a last phase where the empirical variance starts to decrease. This is due to a data truncation [133], we only have access to capacity measurements above a 50% capacity loss, see Fig. 2.9d. After cycle 1500 the empirical variance is only computed on a subset of batteries, the one with the slowest aging, which explains the variance decrease. The CGP model is sensitive to this effect and predicts a decrease in the variance whereas an increase could be expected. Improving the robustness of the CGP model will be part of future work. More generally, it is a common practice in experiments to stop after a fixed SoH threshold is reached. We highlight that this data truncation can cause bias in trend and variance estimation if it is not considered in the model.

In Fig. 2.9c, we added predictions for $y_{standard}$ for training batteries. Compared to the previous posterior law of $y_{centered}$ component for the GPR model, their behavior is more coherent with a stationarity assumption.

Finally, complete predictions of the capacity can also be seen in Fig. 2.9d. The predicted trend is similar and fits still well with training and testing data. The major change is in confidence intervals. They are much thinner in the first cycles than in the GPR case and increase with time fitting well with the variability observed in the data. So we clearly observe the improvement compared to the previous homoscedastic model 2.8d.

Quantitative results

We complete this comparison with a more quantitative analysis of the two models. We want to compare the performances regarding mean and uncertainty prediction, so two indicators are considered.

At first, to quantify the adjustment of the mean, we choose the mean absolute error (MAE),

$$\text{MAE} = \frac{1}{n^*} \sum_{i=1}^{n^*} |\mathbf{y}_i^* - \widehat{\mathbf{y}}_i^*| \tag{2.50}$$

with $\mathbf{y}^*$ the vector of testing outputs not used to train the model, and $\widehat{\mathbf{y}}^*$ the associated predictions, both of size n^*. We preferred MAE with an absolute value of the differences rather than MSE with a square of differences. A few data points can strongly affect the square in the MSE, whereas the absolute value is more robust. Robustness is essential in this experimental context, where the most prominent differences can be due to some experimental errors.

Then to validate uncertainty quantification, we use the negative log predictive density (NLPD) [134]

$$\text{NLPD} = -\frac{1}{n^*} \sum_{i=1}^{n^*} \log p(\mathbf{y}_i^* | \mathbf{x}_i^*) \tag{2.51}$$

with $p(\mathbf{y}_i^* | \mathbf{x}_i^*)$ the predicted density of $\mathbf{y}_i^*$ evaluated at input $\mathbf{x}_i^*$. This metric is commonly used in the GP community. As MAE, it is sensitive to errors in mean prediction, but at the same time, it penalizes under or overestimated uncertainties. We want it to be as small as possible.

In the previous section, we used 20 batteries for training the model, this ensured some robustness in the results, but this is not realistic for most R&D applications. Indeed, in practice, only a few batteries are tested at one experimental condition, often 2 or 3. So in this section, we use smaller training sets. Different training sizes are considered (2, 4, 6, 8, 10, and 20) to have some ideas on how many batteries are necessary for satisfactory performance (see [135] for a more complete discussion on the sample size required). For a small number of training batteries, the learned model can vary significantly with the choice of the training set. To quantify this variability, we repeated each experiment one hundred times with randomly selected training batteries. Each time, the fitted model was then tested, and the two performance indicators computed, on the remaining batteries.

We computed these indicators for six models, two GPRs, and four CGPs. In both cases, we tried a Matérn 3/2 and Matérn 5/2 version for k_0, and for CGP, we compared two different P

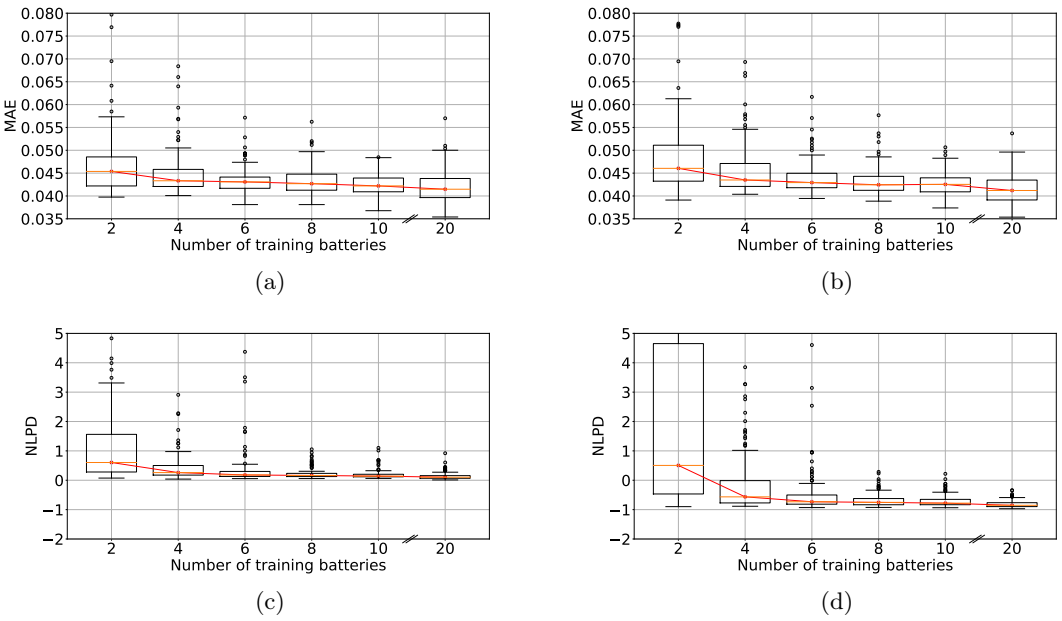


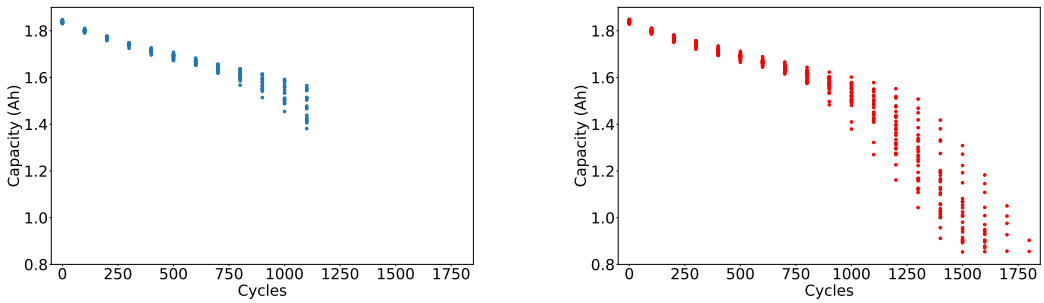
FIGURE 2.10 – Quantitative results as a function of the number of training batteries with MAE performance above and NLPD performance bellow. The left collumn diplays results for GPR and the right collumn for CGP.

transforms, the exponential function, and the softplus. Some differences can be observed because of the choice of kernel and transform, however these differences are small compared to the one between a GPR and a CGP models. So, in the following, we will focus on comparing the GPR and CGP models, displaying results for the GPR model with Matérn 3/2 kernel and the CGP model with exponential positive transform and Matérn 3/2 kernel.

Fig. 2.10 shows the evolution of MAE and NLPD indicators. Concerning mean prediction, there is no major difference between the GPR and CGP models. In both cases, the error decreases with the number of training batteries and so does the variability linked to the choice of the training batteries. However, interestingly, we observe that the MAE decreases significantly between 2 and 4 batteries but then stabilizes quickly. MAE continues to decrease with the number of batteries but much more slowly. So, if one is only interested in mean prediction, it is interesting to notice that good performances can already be obtained with 4 batteries.

Concerning uncertainties, using NLPD, we now see an important difference between GPR and CGP. For the GPR model, it decreases with the number of batteries staying above zero. For CGP, we first observe a high dispersion of NLPD when using two batteries, followed by a clear improvement as soon as four batteries are used for training. It reaches a value significantly lower than in the GPR case, close to -1. The poor uncertainty quantification when using only two batteries could be expected because of the amplitude of the cell-to-cell variations and the difficulty of learning a variance function with only two observations. However, it improved quickly and gave good performance even with few batteries.

This experiment confirmed that using a heteroscedastic model improves the predicted uncer-



(a) (b)

FIGURE 2.11 – Example of a forecasting task with (a) Training data (b) Testing data

tainties while keeping similar performances for mean prediction. Interestingly, we observed that satisfactory performances could be obtained even with a low number of batteries, say 4. Finally, we recommend testing different kernels and positive transforms. Some supplementary gains could be obtained.

2.3.3 A physics-informed Chained Gaussian process for forecasting

Until now we limited predictions to observed cycles. However, the forecasting of never observed cycles would also be of high interest for manufacturers. An essential bottleneck for them is the cost of aging tests, limiting the number of tested elements and the insight on the degradation process. One of the main reasons for testing costs is their duration. End-of-life generally occurs after months or even years of testing. Testing time is also a constraint in itself since, in a competitive sector, manufacturers need to characterize a new battery design as soon as possible. Moreover, even if the 20% capacity loss is a standard end-of-life condition in the automotive industry, modeling of lower SoH levels could be required. In some applications, batteries could be used up to 40 or 50% of capacity loss, or this modeling could be necessary for second-life applications [136], requiring even longer testing time. Thus, shortening testing time is an essential need for manufacturers.

As an illustration, let us consider the following use case : using the Aachen dataset we suppose that only half of the batteries have been tested and that experiments had to be stopped after $T_{max} = 1100$ cycles. That corresponds to the observations in Fig. 2.11a, which are the data available to train a model characterizing health degradation. However, we would like the learned model to generalize well to new batteries (cycled later) see Fig. 2.11b, for observed cycles from 0 to 1100 but also for following cycles between cycles 1100 and 1800. Here 700 cycles could be spared, resulting in an appreciable time gain. So in this section, the task is similar to the one of the previous sections but we will focus our interest on the performance in the forecast area.

Difficulties of Chained Gaussian processes model regarding forecasting

At first, we tackled this use case using a CGP model as presented in the previous section. Fig. 2.12 displays the complete predictions and then shows more details on μ and its first and

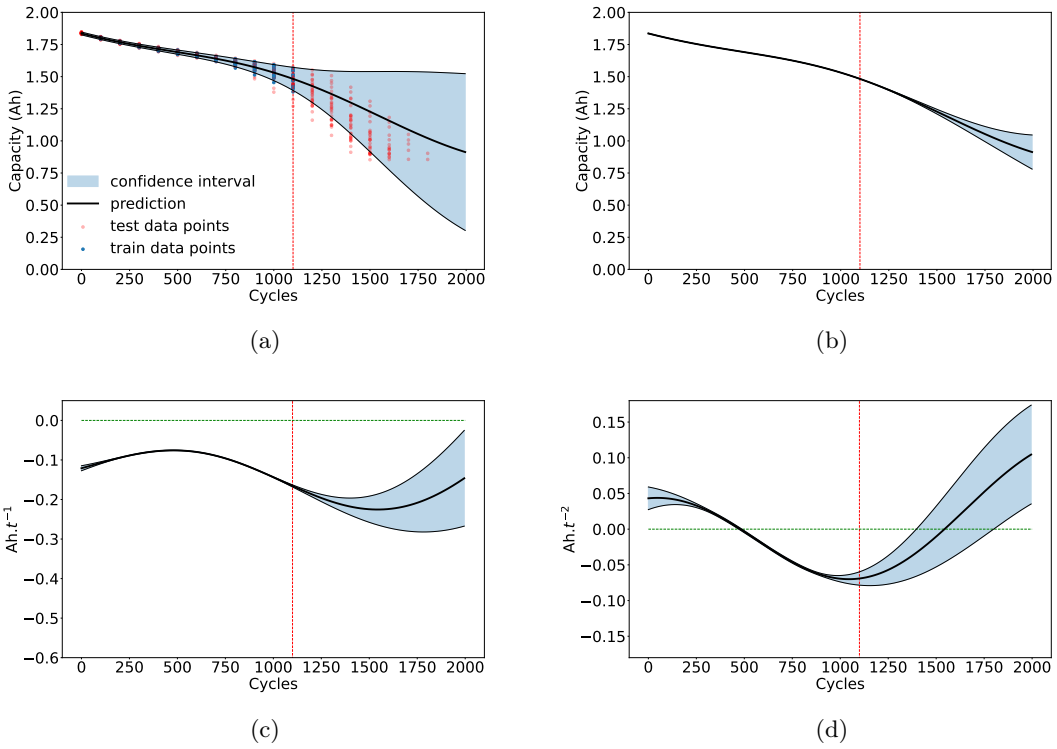


FIGURE 2.12 – Qualitative experiment with CGP model without constraints (a) capacity prediction (b) μ prediction (c) $\frac{d\mu}{dt}$ prediction (d) $\frac{d^2\mu}{dt^2}$ prediction.

second derivatives. We added a dashed red line to highlight the limit of the observed data and a green dashed line to check the sign of derivatives.

Considering capacity prediction in Fig. 2.12a, we observe a qualitatively satisfactory fit to training data. As observed in our previous work, the non-parametric nature of GPs allows us to fit a complex mean function accurately, fitting the testing points. Besides, the increase of the cell-to-cell variation is handled with most of the testing points inside the confidence intervals.

However, forecasting performance varies with the predictive range. Indeed, it seems accurate for the first two hundred future cycles, but then it degrades, leading to strongly overestimated predictions. The predicted mean is higher than the highest capacity measurements. However, we remark that most of the points stay in predicted confidence intervals. Predicted uncertainties increase significantly, and confidence intervals include both extreme cases of accelerated degradation and of a capacity staying equal to the last observed value.

We now look more in-depth at the trend as Fig. 2.12b displays its prediction. Its mean is precisely the predicted trend of the previous plot. We can observe a significant acceleration after cycle 700, which continues in short-range forecasting but progressively moderates. Associated epistemic uncertainties are almost negligible in interpolation but then increase with distance to training data contributing to total uncertainties in capacity prediction. This can be more clearly seen by looking at the derivatives of μ in Fig. 2.12c and 2.12d. Considering the first

derivative, it remains negative for all predicted cycles. However, it increases around cycle 1500 the second derivative becoming positive. CGP model accurately learns an acceleration of the capacity degradation after cycle 500 but does not maintain it in forecasting, which leads to an overestimation of the mean.

This behavior is typical of a GP having a kernel decreasing toward zero with the distance between data points. Extrapolating at a distance of the order of the length scale, the correlation to training data is close to zero, and predictions converge to the prior mean. In our use case, the prior mean of μ is the empirical mean of the data, which is 1.65 Ah. So in long-range forecasting, the predicted mean will stop decreasing and increase toward this value. Indeed we can see in Fig. 2.12d that the second derivative occurs just after the last observed data points. Qualitatively, this is a conservative behavior that in many cases is interesting, avoiding divergences in the prediction, but which is a bottleneck in case of real acceleration.

Prior knowledge on the degradation

The proposed CGP model is a data-driven approach, mainly physic-agnostic. It is able to fit complex models with few assumptions, reaching high accuracy on training data, but its performance drops when only a few data are available, as in this forecasting context. An interesting idea to balance this difficulty is to use a compromise between purely data-driven and model-based approaches, including some physical knowledge within the model. This idea is currently intensively studied in the machine learning community, known as physics-informed machine learning [137, 138], and is rising in battery applications [139].

These approaches generally constrain predictions, more or less strictly, to respect physical modeling of the process. However, the degradation of LIBs is due to coupled nonlinear aging mechanisms and general physical modeling of full SoH degradation is not available. Thus we propose here to consider more qualitative prior knowledge on the degradation, with the derivatives of the degradation trend.

Indeed, in most settings, the trend is known to be monotonic, decreasing for capacity and increasing for internal resistance, the sign of the first derivative staying constant. Knowing this behavior, we will expect predictions to respect it and associate constraints on the first derivative of the predicted trend could be interesting, as figured in [140]. Then, focusing on capacity, as we saw for the Aachen dataset an acceleration of the capacity fade can be observed. This phenomenon is general [141] and can be observed in many settings [142, 143, 144]. Thus, one can expect that after some point the mean degradation curve will stay concave with a negative second derivative.

It is a common practice, to decompose the degradation into phases. Such decompositions are useful from both a physical and a mathematical perspective. From a physical perspective, phases allow the decomposition of the degradation trend according to the dominating degradation mechanism. From a mathematical viewpoint, it allows having areas with some mathematical properties fixed. On the Aachen dataset, considered here, while introducing the dataset, the authors presented a decomposition in two phases with a slow degradation and a fast degradation phase [47, 73], see the decomposition in red in Fig. 2.13. Using the detailed analysis of [145], they associated the first phase with a loss of lithium inventory and the second with a loss of active material loss.

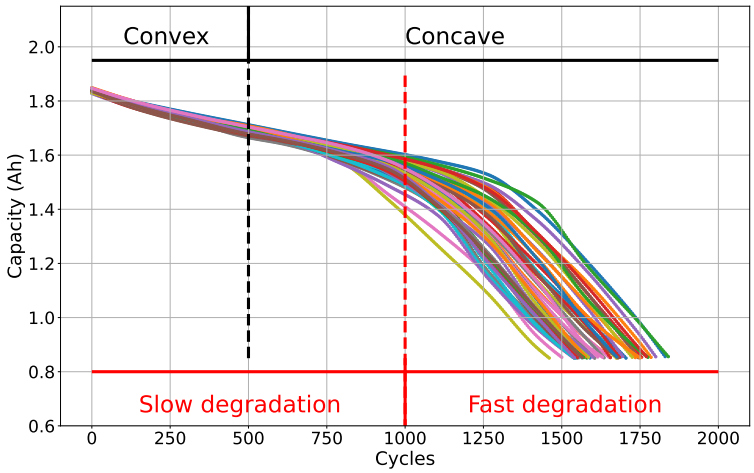


FIGURE 2.13 – Decomposition of the degradation trend in two fixed sign curvature areas

While keeping the same physical interpretation on this dataset, we propose in this work to change the mathematical criterion separating the phases. Instead of separating them according to the value of the slope, the first derivative, we will consider the sign of the second derivative, such that the sign of the curvature will be constant within each phase. As shown in Fig. 2.13 in black, the Aachen dataset can be decomposed in a convex part with a decelerating degradation and then in a concave part with the acceleration of the degradation. These phases are separated by an inflection point, where curvature changes with a second derivative equal to zero. This criterion is also used in [141], speaking of sublinear and superlinear degradation instead of convex and concave phases. As highlighted by the review, these behaviors are observed in many settings. The general physical interpretation is mainly a solid-electrolyte interphase (SEI) growth for the convex part, while many causes have been identified for the concave part, such as the well-studied lithium-plating.

Fast and slow degradation phases correspond to a quite subjective criterion. On the other hand, in our proposed division, each phase is of a fixed second derivative sign, and can directly be linked to possible constraints. Depending on the area, we may want to set positivity or negativity constraints on the second derivative. In the forecasting task, we will mainly focus on the accelerated degradation phase, since it is the last observed phase and because of important possible errors in modeling due to the important variation of the slope. Moreover, it is important to notice that the concave phase starts earlier than the fast degradation phase. Literature on SoH degradation often concentrates on the knee-point (position of maximum curvature) [141], however, the inflection point is an interesting alternative regarding the forecasting task. In the Aachen University dataset, the inflection point occurs before the knee point, before a 10% capacity fade, and such behavior is also observed in [146] (see Fig.6).

Even if it is less well-known we could also consider prior knowledge on σ_0 the cell-to-cell variation. As we can see in Fig. 2.11b, SOH is often identical for a batch of batteries but progressively differentiates, corresponding to an increase of σ_0, so we may expect it to increase at

each time, a phenomenon also observable in [93], [72], or [94] for example. Its second derivative constraint may also be considered. Finally $y_{standard}$ corresponds to standardized fluctuations of y. Depending on the battery, it may increase or decrease or even be non-monotonic. Also, for this component, no prior knowledge is available.

Inclusion of constraints within the CGP framework

We discussed the available physical prior knowledge on the derivative, that maybe useful to improve the forecasting. Now we can explain how to include it in the model. In the literature, monotonicity and convexity constraints have been considered in many settings. They are rather standard with model-based approaches, being directly imposed by the choice of the model [147]. Concerning data-driven approaches, Gamma process is an often-used statistical method to model monotonic degradations [148]. However, such parametric methods impose strong restrictions, with possibly limited quality of fit. Setting derivative constraints on artificial intelligence methods is less frequent but a few studies proposed to impose monotonicity constraints on degradation trend, with B-splines [149], neural-networks [150] or Gaussian processes (GP) [140]. To the best of our knowledge, convexity constraints on data-driven approaches have not been considered yet in battery applications.

In order to keep the precise cell-to-cell variation modeling we will try here to directly set constraints on the previous CGP model. So we focused on physics based methods proposed for GP-based approaches. Regarding Gaussian process regression, a classical choice to improve forecasting is to include a prior mean function inspired by physics modeling [81, 80]. However, such an approach loses most of the interest of non-parametric modeling, forecasting relying essentially on the chosen prior mean. As explained previously, we cannot neither consider more complex techniques such as [151], requiring precise physical modeling of the phenomenon. So we focus on methods imposing directly the expected derivative constraints. Such approaches have been reviewed in [152], considering also boundary and boundedness constraints. Among the available methods, we considered the strategy proposed by Riihimäki and Vehtari [112], as it is compatible with the CGP framework. We will present how the CGP framework can be extended to include these constraints. Finally, we highlight the recent and independent work of Bai et al. [140], which imposed a decreasing constraint on their GPR model of the capacity fade, using an alternative framework [153], showing the interest of setting constraints.

To include derivative constraints, Riihimaki and Vehtari [112] proposed to add virtual points translating the derivative prior knowledge. Considering an univariate case, with $f' = \frac{df}{dt}$, we suppose that we have prior knowledge on the monotonicity thus on the sign of f'. We want to set derivative constraints at locations, n^v locations concatenated in X^v. At these positions virtual observations $\mathbf{z}^v$ taking values in $\{0, 1\}$ are added, equal to 0 if f should be decreasing, to 1 if it should be increasing.

To link them to the derivative, we suppose that

$$\mathbf{z}_i^v | \mathbf{f}'_i \sim \mathcal{B}(s(\mathbf{f}'_i)), \tag{2.52}$$

where $\mathcal{B}$ is the Bernoulli distribution, $\mathbf{f}' = f'(X^v)$ and s is a sigmoid function for a mapping from $\mathbb{R}$ to $[0, 1]$, logit or the inverse of the normal cumulative distribution function. Thus if f' is

positive with a high probability at some location, a virtual variable at the same location has a high probability of being equal to 1 and conversely. We used the v exponent to highlight that this does not correspond to real observations ; they are artificial data.

Riihimaki *et al.*'s proposal was initially planned for the GPR framework. Later Tolvanen [154] included it in the sparse variational framework, which uses a similar inference than CGP but sticks on a Gaussian likelihood. We propose here to include it in the CGP framework. Regarding the model the virtual variables can similarly be added, the constraint being set on the d-th component f_d of the model. With the addition of the virtual observations, the model has an extended likelihood $p(\mathbf{y}, \mathbf{z}^v|\mathbf{f}_1, \cdots, \mathbf{f}_C, \mathbf{f}'_d)$ which can also be factorized as follows

$$p(\mathbf{y}, \mathbf{z}^v|\mathbf{f}_1, \cdots, \mathbf{f}_C, \mathbf{f}'_d) = \prod_{i=1}^{n} p(\mathbf{y}_i|\mathbf{f}_{1,i}, \cdots, \mathbf{f}_{C,i}) \prod_{j=1}^{n^v} p(\mathbf{z}^v_j|\mathbf{f}'_{d,j}). \tag{2.53}$$

Now $\boldsymbol{\theta}$ corresponds to the variable of previous section, completed with the new derivatives, $(\mathbf{f}_c, \mathbf{f}'_d)_{c\in[\![1,C]\!]}$. We propose to adapt the previous variational distribution such that

$$q(\boldsymbol{\theta}) = p(\mathbf{f}'_d|\mathbf{f}_d) \prod_{c=1}^{C} q(\mathbf{f}_c), \tag{2.54}$$

keeping similar Gaussian $q(\mathbf{f}_c)$ with the same variational hyperparameters.

With analogous computations than for the CGP model, a new ELBO can be derived

$$\text{ELBO}_{\text{DCGP}} = \text{ELBO}_{\text{CGP}} + \sum_{j=1}^{n^v} \int \log p(\mathbf{z}^v_j|\mathbf{f}'_{d,j}) q(\mathbf{f}'_{d,j}) d\mathbf{f}'_{d,j} \tag{2.55}$$

with $q(\mathbf{f}'_{d,j}) = \int p(\mathbf{f}'_{d,j}|\mathbf{f}_{d,j}) q(\mathbf{f}_{d,j}) d\mathbf{f}_{d,j}$ with an exact analytic expression, using Gaussian properties and (2.8) and (2.9). We observe that the new ELBO corresponds to the classical ELBO of the CGP framework corrected by an additional term depending on the constraint, which can be seen as a regularization term.

The proposed derivation imposes a constraint on the first derivative of one of the components, to impose monotonicity. Convexity can be imposed similarly, replacing f'_d by f''_d. In fact, using the formalism of [153], any linear constraint with the shape $\mathcal{L}f \leq 0$ or $\mathcal{L}f \geq 0$ can be used in this framework ; the law of $\mathcal{L}f$ being still Gaussian with a known expression. This includes the important case of boundedness, using a shifted version of f. By adding the appropriate correction term, several constraints can be included at the same time, in different areas, on different components, and on the first or second derivative. Regarding the cell-to-cell variation, we should note that a constraint cannot be directly applied on σ_0 but it can be set on η.

This method is local, depending on the number and the position of the virtual points. We highlight that there exists no guarantees that the constraints will be fully respected. The second term favors solutions respecting the constraint, the enforcement increasing with the number of virtual points. In the forecasting application, we will only set constraints on future cycles, however,

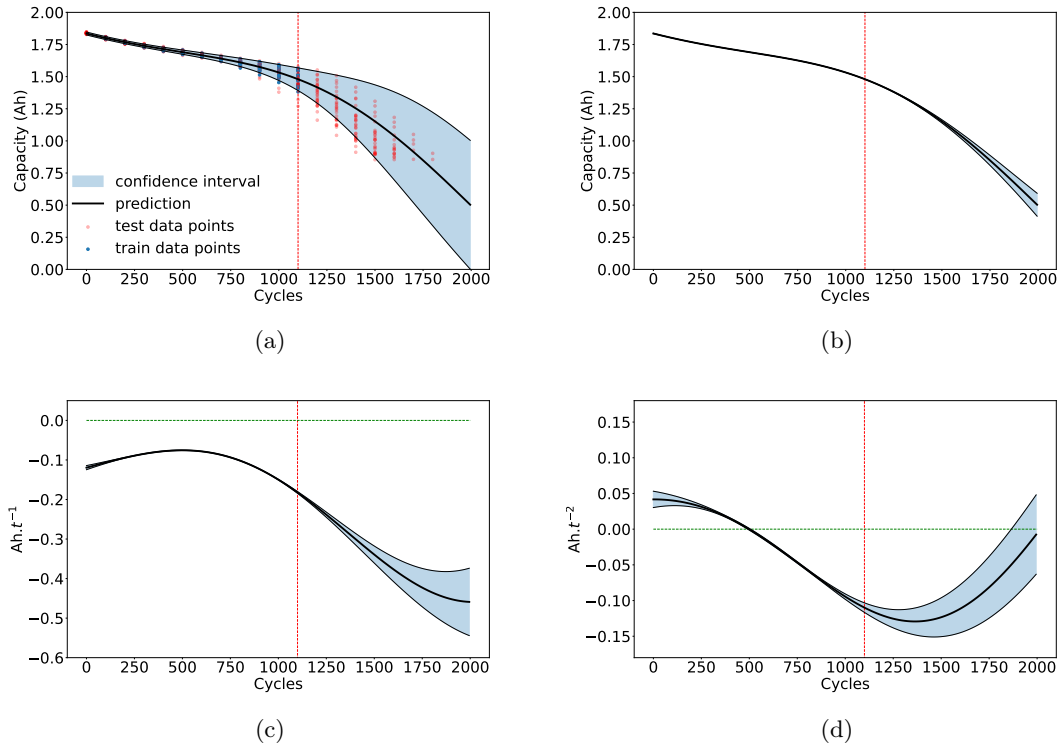


FIGURE 2.14 – Qualitative experiment with CGP model with derivative information (a) Complete capacity prediction (b) μ prediction (c) $\frac{d\mu}{dt}$ prediction (d) $\frac{d^2\mu}{dt^2}$ prediction.

in more complex multivariate cases, selecting the virtual point position, may be difficult. As explained by the authors [112], an iterative procedure may be necessary to ensure constraints, adding virtual points progressively. Different strategies may be used, placing virtual points on a grid or preferably in places where constraints are not respected.

We also remark that this method is not entirely local. Indeed, as noted by the authors, the addition of monotonicity constraints tends to increase the smoothness of predictions. In practice, length-scale hyperparameters are often higher, which can affect predictions not only at virtual value's position but everywhere. Looking at ELBO_{DCGP}, Eq. (2.55), there is no difference between actual and virtual data. For a fixed n, increasing the number of virtual points n^v will reduce the influence of actual data during optimization. Thus we recommend being as parsimonious as possible, using derivative constraint only when the unconstrained method produced incoherent results.

Qualitative validation

We can now test this improved model on the use case presented above, trying to improve the results obtained with the CGP. In that case, even if we do not have access to testing data, looking only at predictions of our model, we see that the last observed cycles are inside a concave phase.

We may think that the previous predictions displayed in Fig. 2.12a are too conservative, seeming incoherent with the expected behavior such that this phase should continue further. Thus we can use the introduced methodology to impose the negativeness of the second derivative after cycle 500 and until cycle 2000. For that, we have to choose the position of associated inducing variables. The constraint is respected in observed cycles, so we only set points in the forecasting area between cycles 1100 and 2000 area. We placed virtual points regularly, one point every 25 cycles. We also have prior knowledge of the cell-to-cell variation σ_0, but predictions already revealed the expected behavior, so we did not use additional constraints. Results are displayed in Fig. 2.14.

Looking at complete predictions of the new model in Fig. 2.14a we first see that fit quality stays similar on training cycles, the main changes being in the forecasting area. There, mean prediction is much more satisfactory than in the case without constraints. 200 cycles after the last observed data, the predicted mean overestimates less testing data. Uncertainty modeling is also improved with much narrower confidence intervals still including most testing points. Now even the upper bound of the confidence interval is decreasing.

Looking at predictions for μ in Fig. 2.14b, we see that its mean decreases to lower values than in the previous case. Interestingly, associated uncertainties are much smaller, and the upper bound is also decreasing. This partly explains why final uncertainties on the capacity are reduced.

With derivative predictions in Fig. 2.14c and 2.14d, we see that the first derivative of μ is negative and stays decreasing until cycle 2000, the limit of the constrained area, the second derivative remaining negative. So this example illustrates well that we can impose constraints on derivatives, and that it can improve the modeling. It also highlights the local nature of the method, and the constraint being imposed in the neighborhood of virtual points.

As an intermediary step, not displayed, we imposed constraints on the first derivative of μ. Compared to the initial case without constraint, results were improved, with a weaker overestimation and a second derivative becoming negative a bit later. Qualitatively the results remained incoherent with prior knowledge of an accelerating degradation. Indeed prior knowledge of accelerated degradation is more informative than just the decrease in SoH. However, an advantage of first derivative constraints is that it does not require observing the inflection point, being usable for shorter tests.

We also observed, that when applied to the complete dataset as in previous section, the derivative constraints, allow to correct the bias caused by data truncation. Only the batteries with the slowest degradation was observed, leading artificially to a decrease of the cell-to-cell variability prediction. This unrealistic prediction was no more observed once derivative constraints on the trend was set.

Quantitative validation

Finally, we propose a more quantitative validation of the forecasting in the accelerated degradation phase. We compare the performances with and without the constraints and their evolution with the forecasting range it can be used. To quantify model performance, we will use, as in the previous section, mean absolute error (MAE), to quantify the fit to the mean, and negative log predictive distribution (NLPD) for the uncertainties.

We check the evolution of these indicators as a function of the predictive range, from 100 to 800 cycles. To avoid bias in computed indicators, we only consider cycles up to 1500. That means that concurrent models are trained on a training batch of batteries observed from cycle 0 to 1500 minus the predictive range, and then they are tested on the testing batch on cycles 1500 minus the predictive range to 1500. Thus, a predictive range of 800 cycles corresponds to observing data up to cycle 700 just after the inflection point. Compared models are a standard GPR model, the CGP model described in our previous work, so without constraint, and finally its updated version with a second derivative constraint (denoted DCGP). The GPR uses a squared exponential kernel and constant prior mean equal to the empirical mean of the data. For DCGP, as in the previous section, we set virtual points in the forecasting area every 25 cycles.

To quantify the variability due to the choice of training batch, we repeat each experiment 20 times, randomly selecting half of the batteries for training and half for testing each time. Finally, to provide a baseline for comparison, we also learned a CGP model on all cycles (constraint not being necessary for only interpolation). This allows comparing forecasting performance to the ideal case where we would have taken the time to finish the experiments.

We start by analyzing performances for mean prediction with plots in the left columns of Fig. 2.15. At first, one may be surprised by the evolution of MAE for the case of the complete model in red, with decreasing MAE. However, in that case, all cycles are observed ; the only challenge is to generalize to new batteries. Coming back to Fig. 2.11a we remember that the cell-to-cell variation is increasing. Thus predictions between cycles 1300 to 1500 will generally be worse than predictions between 1100 and 1300. Removing the effect of the forecasting range, we test on a more and more simple setting on average.

This remark has to be kept in mind to understand the performance evolution of the three concurrent models, in blue. GPR and CGP models, in Fig. 2.15a and 2.15c have similar behavior with an error increasing with the forecasting range. This similarity is coherent with our previous study, where we saw that mean predictions were similar for a GPR and CGP model. On the other hand, DCGP in Fig. 2.15e has a performance much closer to the reference using all data even for a high forecasting range ; they are both decreasing. In that case, we should study the relative error of both models, which is increasing.

Then, regarding NLPD on the right side to also consider uncertainties. We observe and justified by the same argument that the indicator for the complete model (in blue) is decreasing with the forecasting range. This time the opposition is more between GPR on the one hand, in Fig. 2.15b, and CGP and DCGP on the other, in Fig. 2.15d and 2.15f. NLPD is generally worst in the GPR case because of its restriction on uncertainties modeling, which was discussed in our previous work. In particular, the GPR model cannot anticipate the cell-to-cell variation increase in this setting. Then, for a forecasting range under 500 cycles, CGP and DCGP face a comparable increase in NLPD, with slightly better results for DCGP. As we could see with the qualitative example, mean prediction is biased, but associated confidence intervals remain coherent with testing data leading to not completely incoherent results. However, after a forecasting range of 500, NLPD increases suddenly for the CGP model. It reaches its limits and becomes unusable. This limit is reached later for the DCGP model, at a range of 800 cycles.

This experiment confirmed that including derivative information significantly reduced the bias in mean forecasting, especially in this complex phase with a strong acceleration of the degradation. Moreover, this method kept the accurate uncertainties modeling of our previous model, being able to forecast them as well. Performances of the forecast remained close to the one using all data

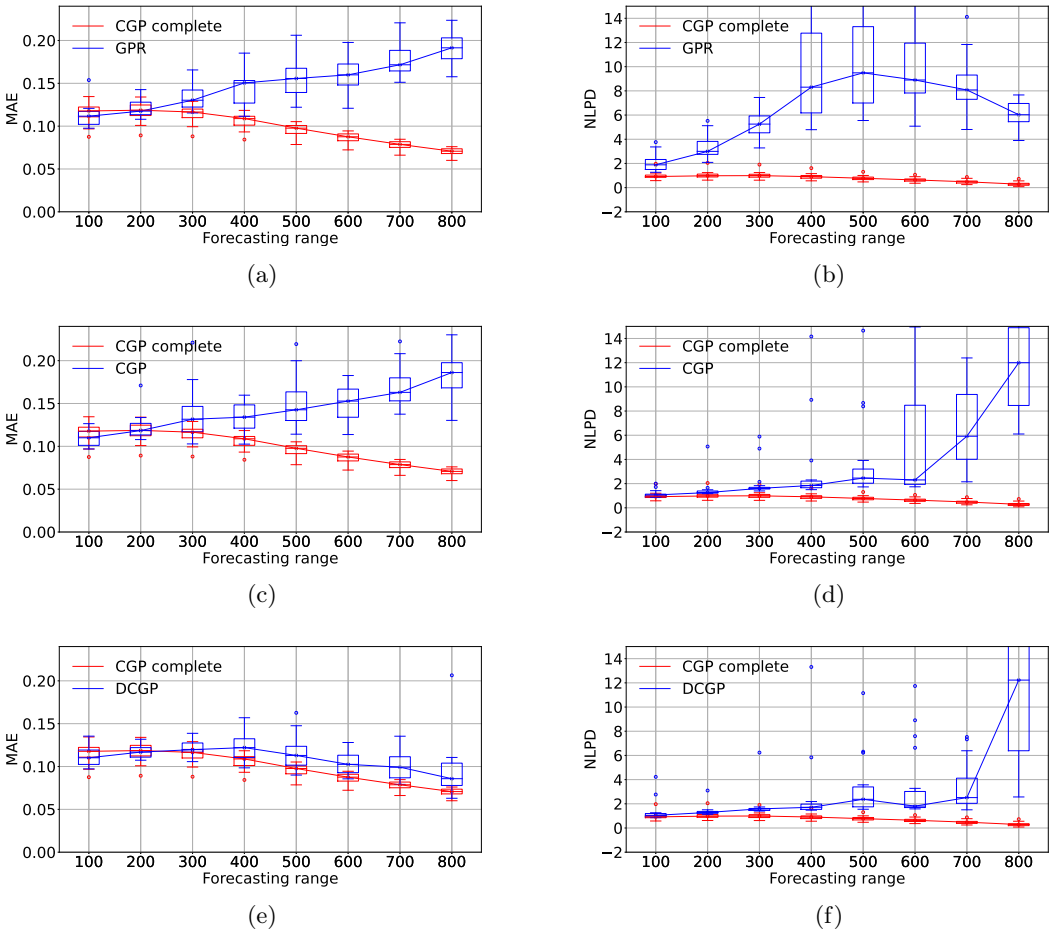


FIGURE 2.15 – Quantitative results as a function of the forecasting range, with MAE performance on the left column and NLPD performance on the right. The upper row displays results of GPR, middle row of standard CGP and bottom row of CGP with derivative information. In each case performance is compared with the one of a standard CGP model using all data.

even after several hundred cycles of forecasting. This opens the possibility to reduce significantly the testing time, relying only on the forecasted values.

2.4 Discussion

In this chapter we proposed several Gaussian process-based approaches to model the health degradation at a unique experimental condition. Our proposal tackled three main challenges :

— The interpretability and decomposition between uncertainty sources, provided by a dedicated kernel design decomposing predictions into several meaningful components

— The heteroscedasticity of the cell-to-cell variations, addressed through the use of the CGP framework which allows the use of customized likelihoods combining, possibly non linearly, several Gaussian process prior, each estimating a different functional parameter
— The forecasting of unobserved cycles, improved by the inclusion physical prior knowledge, as monotonicity and convexity constraints

The combination of these three ingredients allows a precise uncertainty quantification, taking into account its different sources and estimating its time evolution. The non-parametric nature of GP allows a complex and precise fitting on the data, and the physical constraints compensates the lack of data in extrapolation.

Concerning the uncertainty quantification, the proposed model focused on the cell-to-cell variations, an important source of aleatoric uncertainty. Further work on other uncertainties should be done. In particular, since in practice few batteries are tested, attention should be set on epistemic uncertainty. An idea could be to include uncertainties on hyperparameters. This can be done with a fully Bayesian approach with priors on mean function and kernel parameters. However, estimating posterior laws would require Monte Carlo Markov Chain (MCMC) simulations that would increase the computational cost, particularly in a multi-latent setting with several Gaussian processes.

The CGP framework allows an improved uncertainty modeling, but it leads to an increase in the computational burden. Some work could be done to reduce this cost. We could reduce the number of inducing variables of the CGP model ; however, this may result in a loss of performance. We could also improve the optimization process by trying other optimizers and parameterizations of variational hyperparameters, as argued in [155]. Finally, other recent improvements such as variational Fourier features [156] may be considered.

Regarding the inclusion of derivative constraints, we first highlight that the presented extension of CGP is general. It is not specific to our particular capacity model, not even to battery applications. It allows using the abilities of the CGP framework, with any likelihood depending on several non-parametric functions, and to set constraints on each of these components. Not only can monotonicity be imposed, but also acceleration with the second derivative. Several constraints can be imposed simultaneously on different components or different orders. GPs depending on several explanatory variables, could also be considered by applying constraints against different directions. In practice, it may also be useful to handle the experimental noise. The amplitude of this noise often leads to non-monotonic predictions. This problem may be amplified when some data are missing creating important gaps between two observations. In that setting, monotonicity constraints can be particularly interesting.

Before setting constraints, the first step is to determine what are the observed phases and what prior knowledge is available on each of them. For this we recommend asking for expert opinion, especially for curvature constraints. We must be aware that wrongly imposing a constraint may be counter-effective. This is particularly important when we focus on prediction in an area without data, as in the present work. For example, we could imagine that capacity degradation stops accelerating. In that case, predictions obtained without constraint, with wide confidence intervals, may be more accurate than the surely decreasing prediction. In fact, it is the case for lower capacity values as we can see in [73] where a third phase, convex, occurs. Moreover, even if the phases are well known, the proposed method can be used within the last observed phase. It allows to extend it, however, it is not able to anticipate the arrival of a new phase. If only the first convex part is observed we will not be able to anticipate future acceleration. If we do not have

sure prior knowledge that the last phase will continue until the end of life (either because we know that there will be another phase somewhere or because we do not have any idea), we have to assume (or not) the risk of making a wrong decision. And the longer the constraint space is, the higher the risk. In any case, we highlight that monotonicity constraints remain almost always applicable.

We also highlight that we focused on the forecasting in the concave phase, because the convex part was quite short in this dataset, and because the concave part represents a real difficulty in prediction. If the acceleration is not anticipated, high prediction errors can occur. This is particularly true with GP-based methods which tend to be too conservative. The first convex phase is generally easier to handle corresponding to a slow degradation phase, more suited to a conservative behavior. However, monotonicity and curvature constraints could be imposed similarly in convex phases (imposing a positive second derivative) if necessary. More generally, every experimental setting may not display the same two phases, degradation without a first convex phase or with only a convex phase as in Fig. 1 [157] can similarly be observed. However, we are still able to decompose the degradation in concave and convex phases, and corresponding constraints can also be imposed. So, the proposed method is not limited to the decomposition used in this paper.

All the methods applied in this chapter considered a population of batteries, meaning any batteries with given characteristics and tested at a particular experimental condition. So it could be applied similarly to other experimental conditions, at another temperature, depth of discharge or charging/discharging current. The present study considered a fixed experimental condition setting but it could also be used for more realistic experimental patterns. Consider, for example, the time-varying data introduced [158] (section 5.4), where the discharging current is proportional to a typical driving speed profile. Yet, the application of several experimental patterns simultaneously, possibly leading to interpolation or extrapolation of experimental conditions, remains a difficult task. This is because of the limitation in the number of tested conditions, and the high variability in degradation with different accelerations.

Additionally, we applied our model in a cycling context with capacity modeling. However, our model is not specific to this setting. Other SoH indicators, as the internal resistance, could be used instead. Furthermore, the storage application of batteries could be considered as well with only a calendar evolution. This can be handled directly by replacing the cycle number with calendar time. It will be interesting in the future to test our model in various contexts. Finally, even if the focus of this thesis is on the characterization of the degradation, all the elements presented here could directly be used for RUL prediction problem, by making prediction for batteries inside the training batch, also taking advantage of the prior derivative knowledge.

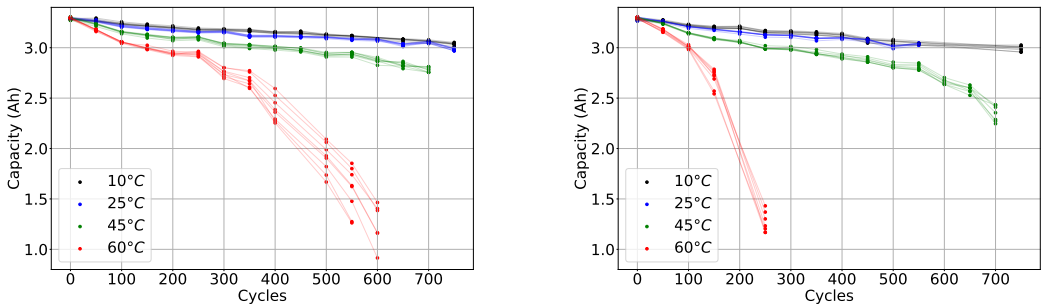
Chapter 3

Battery health degradation prediction at unobserved conditions with Optimal transport

Contents

3.1 Introduction

The previous chapter modeled the state-of-health evolution at a unique experimental condition. This condition can be seen as a reference condition, useful to evaluate the performance of a battery model and eventually permit comparison between battery models. However, this evaluation is often insufficient since performances are highly dependent on the experimental conditions as we can see in the Maryland dataset provided by Diao et al. [72], presented in 1.2.2, and displayed in Fig 3.1. For this dataset we observe a major influence of the temperature on the degradation, a second-order influence of the current charge cut-off, and the interaction between these two factors.



(a) (b)

FIGURE 3.1 – Capacity degradations from the Maryland dataset for a discharge rate of 0.7 C and a charge current cut-off of (a) 0.2 C or (b) 0.025 C .

In general, many other factors such as the depth-of-discharge, the charge and discharge rate, can have an influence. These factors must be taken into account since, a battery type may be more sensible than another in some conditions, performing better in certain cases but worse in others.

A difficulty in practice is that testing several experimental conditions represents an important cost for manufacturers. Notably, for each temperature a dedicated controlled room needs to be used. Thus, few experimental conditions are tested in practice, and the performance characterization using only observed data experiments is incomplete. Also, as described in Sec.1.3.1, we need a model predicting the performance of the battery degradation for unobserved experimental factors, for example for the Maryland dataset at 52 or 35 ° C and a CCC of 0.0125C. Several experimental factors will be considered, but we will focus on the temperature effect, the preeminent degrading factor. As previously, these predictions should not be limited to a simple mean prediction. A precise uncertainty quantification must be proposed. Here, we will concentrate on the interpolation area with intermediary conditions, the harder task of extrapolation, e.g. predictions at 65 °C, will not be considered.

The task of modeling degradation curve in different experimental conditions is commonly tackled by model-based approaches either with empiric models [159, 72], semi-empiric models [160], mechanistic models [161] or equivalent circuit model [68]. However, as already discussed, one can argue that model-based approaches can face limitations and lack of versatility. Moreover, we would like to conserve the interesting results obtained in the previous chapter, notably the precise uncertainty quantification. Yet, Gaussian process-based and data-driven approaches in general are mostly limited to a unique experimental condition. With GPs, a few works [162, 131] tackled this task but with severe constraints on their models to permit the fit : limitation to a linear time degradation [162] or independence of the time and experimental factors effects [131]. Liu *et al.* [163] did not impose such constraints, however considering a storage application without strong acceleration of the degradation. Moreover, all these methods stuck to the Gaussian process regression framework which, remains limited to quantify uncertainties evolution. The main issue is that SoH is typically measured frequently in time allowing a precise data-driven time modeling, but because of the costs of experiments, this is not the case for experimental conditions. Only a few are observed, leading to observations too sparse for a high-quality data-driven fit.

To address the constraints posed by the data while maintaining the precise time evolution

modeling achieved earlier, we will describe a two-step methodology coupling the benefits of the data-driven and model-based approaches. The time degradation will correspond to the data-driven approach detailed in the previous chapter and we will present here the second step extending the result to several experimental conditions. This additional step follows a model-based approach allowing to mitigate the limited number of observed conditions by the inclusion of physical prior knowledge. In this step the inputs are the experimental factors and the outputs are the predictions of the first step, corresponding to probability distributions. So it will require a regression method able to deal with such outputs. Among many proposals, we will focus on optimal transport approaches [164, 165] providing a sound framework to deal with probability distributions.

Also in this chapter, we will start in Sec. 3.2 by justifying the selected approach, explaining the limits of a direct Gaussian-based regression to model experimental factors effect and proposing instead a two-steps approach. As it will be the basis to learn the experimental factors effect, we will take some time to present in Sec. 3.3 the general necessary background on Optimal transport theory. We will then see in Sec. 3.4 concretely how it can be used for prediction with the introduction of the conditional Wasserstein barycenter. Two types of this predictor relying on a model-based approach will be derived and tested on the Maryland dataset. Finally, the limitations and possible improvements of the method will be discussed in Sec. 3.5.

This chapter contains the results developed in the following article, published in the Journal of energy storage

— Larvaron, B., Clausel, M., Bertoncello, A., Benjamin, S., Oppenheim, G., & Bertin, C. (2024). Conditional Wasserstein barycenters to predict battery health degradation at unobserved experimental conditions. Journal of Energy Storage, 78, 110015.

3.2 Choice of the approach to model experimental factors effects

3.2.1 A first attempt : A direct approach

The model of the SoH time degradation at a fixed experimental condition presented in the previous chapter was the following

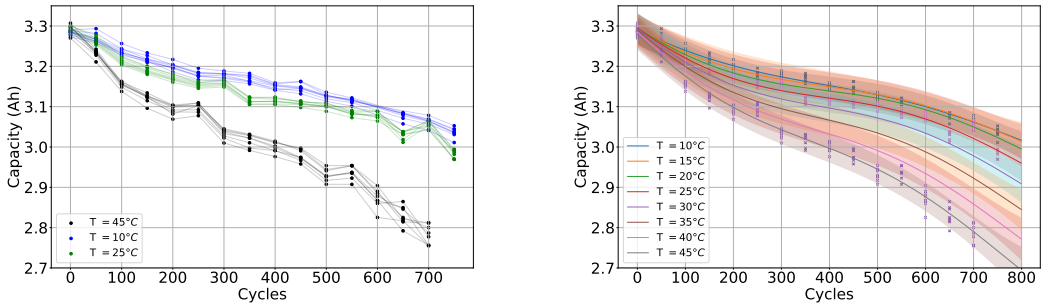
$$y(t,b) = \mu(t) + \sigma_0(t) y_{standard}(t,b) + \varepsilon, \tag{3.1}$$

with μ the trend of degradation, σ_0 the amplitude of the cell-to-cell variations [3], $y_{standard}$ the standardized deviation of each particular battery and ε an experimental white noise.

To take into account the influence of the experimental conditions resumed in a vector $\mathbf{c} \in \mathbb{R}^p$ the most natural approach is to consider a unique model function of all inputs, time, and experimental factors playing identical roles. Following this approach the previous model (3.1) can be updated in the following way

$$y(t, b, \mathbf{c}) = \mu(t, \mathbf{c}) + \sigma_0(t, \mathbf{c}) y_{standard}(t, b) + \varepsilon, \tag{3.2}$$

where μ and σ_0 depend on the conditions as characteristics of the population whereas $y_{standard}$ does not, corresponding only to the deviation of a particular battery caused by its tiny production differences. This kind of approach is present in the literature for example in [162, 131, 163] within the GPR framework.



(a) (b)

FIGURE 3.2 – (a) Degradation data from the Maryland dataset at the three lowest temperatures and (b) associated Gaussian-process interpolation

This update can simply be done by using a dedicated lengthscale for the time and each particular factor, as explained in Sec. 2.2.1. So we can tackle this larger problem at a low cost, staying within the CGP framework and keeping all the previous advantages. Liu *et al.* [166] ,remaining in the GPR framework, presented and validated the prediction of a complete degradation curve at an unobserved condition. We illustrate the result of such an approach on the Maryland dataset. Reducing the problem to three conditions at the lowest temperatures, Fig. 3.2a, we display in Fig. 3.2b the corresponding interpolation.

This approach can be further improved through the use of derivative constraints relative to the factors. In the previous chapter, see Sec. 2.3.3, we detailed the use of derivative constraints regarding time, but the presented method can similarly be applied in a multivariate context by considering partial derivatives. However, caution is needed as the application of derivative constraints becomes more complex in a multivariate setting. The method relies on the virtual points where we impose the derivative constraint and that we have to place in the input space. As the number of factors increases it becomes harder to assess that the constraint is respected everywhere.

This direct approach provides interesting results, however, we found out that it could only be applied in relatively simple cases. As in the illustration above, Liu et al. [163] considered a use case with similar degradation curves, without strong acceleration. The case that we consider with the Maryland dataset in Fig. 3.1 is more complex. Notably, we see that the degradation at 60 °C is much faster than at lower temperatures. In that case, the interpolation task becomes extremely complex and the direct approach faces its limits. We observed that such models are much less precise and robust at observed conditions than a model fitted specifically on one of these conditions. This can be explained by the fact that in the direct model case several hyperparameters

are shared for all experimental conditions whereas in the specialized time-degradation, they are precisely learned on each of them. Moreover, learning a direct model is a much more complex task than learning several models each at a particular condition. To obtain a model maintaining the performance obtained in the previous section we propose another approach.

3.2.2 Our proposal : A two-step approach

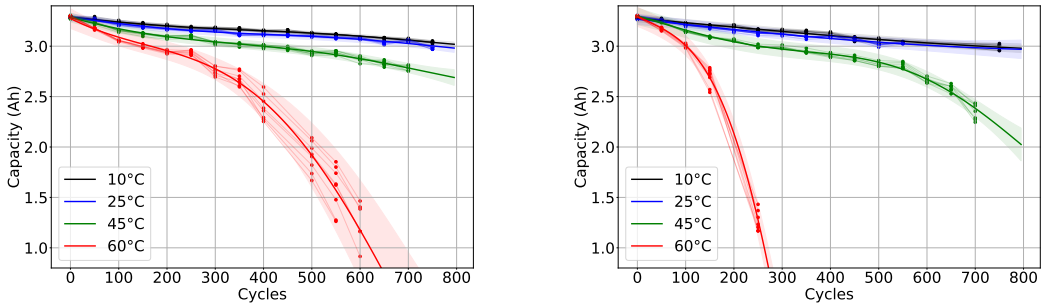


FIGURE 3.3 – Time degradation modelings for a discharge rate of 0.7 C and a charge current cut-off of (a) 0.2 C or (b) 0.025 C. The results are obtained by applying the chained Gaussian process model independently to each condition.

We argue that time and experimental conditions are variables of a different nature and have to be considered separately. Indeed, in experiments SoH is frequently measured whereas few experimental conditions can be tested in practice. Moreover, the important acceleration of the degradation with temperature increases this difficulty. Because of their differences, we should not apply the same modeling method to both groups of variables. Also, we propose instead a two-step approach splitting the time modeling and the effect of experimental factors. First, time degradation is done following the previous model (3.1) at each condition, then a second step models the effect of the factors. It uses the predictions of the first step as training data to build a second model.

More precisely, the first step consists of fitting independently, on each of the K conditions the model (3.1) using the data $((t_i, b_i), y_i)_{i=1}^{B_k}$. We obtain the fits presented in Fig 3.3 with the predicted means and associated 95% confidence intervals. As the fits result from a Gaussian processes-based method, they consist of multivariate Gaussian distributions, curves being approximated by their evaluation at a finite number of positions, as many as needed. For each condition $\mathbf{c}_k$, the result of the fit is denoted $\mathbf{z}_k \approx (t \mapsto y_{\mathbf{c}_k}(t, b^*))$, with $\mathbf{z}_k \sim \mathcal{N}(\mathbf{m}_k, \Sigma_k)$. $\mathbf{z}_k$ is of dimension d, the number of predicted times, so it belongs to $\mathcal{N}^d$ the space of Gaussian distributions of such dimension. All of the $\mathbf{z}_k$ are of the same dimension, meaning that they are all evaluated at the same particular times until t_{max}. In this setting t_{max} corresponds to cycle 800. We see that it is not mandatory to have observations until this limit, we can use instead some forecasting as presented in Section 2.3.3.

Now the effect of the experimental factors can be taken into account in the second step. This step uses the results of the first step as input data, $(\mathbf{c}_k, \mathbf{z}_k)_{k=1}^{K}$, with the factors $\mathbf{c}_k$ as inputs and the fits $\mathbf{z}_k$ as output. It builds a model that for any new condition $\mathbf{c}^*$ can predict another Gaussian

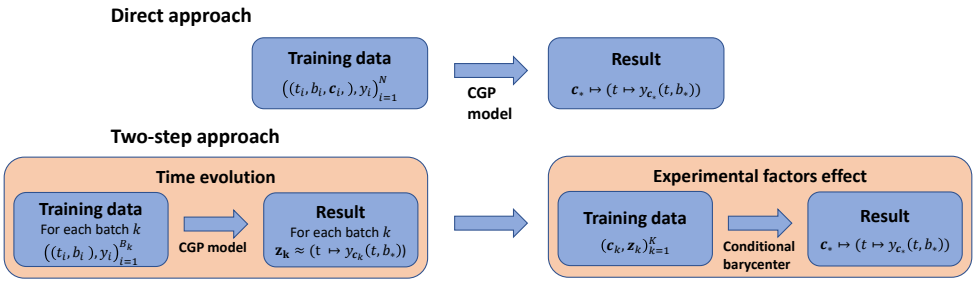


FIGURE 3.4 – Summary of the direct and two-step approaches

distribution, $\mathbf{z}^*$, corresponding to a degradation curve with uncertainties. The difficulty of this step relies on the complex nature of the $\mathbf{z}_k$. Denoting f a function that maps an experimental condition to a Gaussian distribution, $f : \mathcal{C} \to \mathcal{N}^d$, the developments below will propose such a f providing suited predictions. Importantly, to keep the performances of the individual time fitting we will concentrate on interpolating f functions, such that $f(\mathbf{c}_k) = \mathbf{z}_k$. Interpolating models are often criticized because they learn too complex functions fitting noise experiments. However, here the $\mathbf{z}_k$ are learned as an aggregation of several battery degradations, the effect of experimental noise is thus reduced, and it becomes interesting to fit them as closely as possible. A summary of both direct and two-step approaches can be found in Fig. 3.4.

Remark In fact, in our case, an additional difficulty occurs. Our model uses the chained Gaussian processes framework [95], with a nonlinear combination of Gaussian processes components and the result is not exactly a Gaussian distribution. As an approximation, we will use the described methodology independently on each of the components, obtaining for each one a prediction at the new location that we will then merge to get the final prediction. Nevertheless, the described approach is directly applicable while using a standard Gaussian processes regression.

This two-step approach requires a predicting method over complex outputs, probability distributions in our case. Two kinds of methodologies exist to handle such problems. The most often used nowadays consists of using an embedding step, mapping complex inputs toward standard vectorial objects, and then applying standard statistical methods. This approach has been intensively used with images, text, or graph applications [167, 168]. However, it has the downside of losing, to some extent, the structure of the objects. So we will prefer another approach consisting in extending the statistical methods to be able to handle these complex data [169]. Many such methods have been proposed in the literature [170, 171, 172, 15], in this work, we will focus on methods provided by the optimal transport theory. This is a natural choice since our outputs are probability distributions and optimal transport provides a strong theoretical foundation on these objects [164]. Sec. 3.3 will first provide the necessary background on optimal transport and Sec. 3.4 will introduce the proposed predictors to model experimental factor effect.

3.3 Background on optimal transport

Historically, the optimal transport problem is traced back to the work of the french mathematician Gaspard Monge [173], starting from the applied problem of optimizing earth displacement for military applications. This was then rediscovered in Russia for economic applications notably by Kantorovitch [174] who subsequently received the Nobel prize in Economics. Kantorovitch proposed a rigorous formulation of the problem, generalizing the initial problem of Monge. A numerical solution to his problem was then proposed by Dantzig [175]. It used the linear programming framework, and it stayed for a long time as a standard application of such methods. Later, in the 1990s, mathematicians explored many developments of this theory notably Brenier [176, 177] in link with works in fluids dynamics. These results were reviewed in the two books of Villani [178, 164] and more recently by Santambrogio for applied mathematicians [179]. Concurrently, optimal transport gained success in applications, initially in for image processing with the use of the earth's mover distance [180]. Nowadays, its use spread to many other domains in machine learning and statistics [165], such as in domain adaptation [181, 51], generative models with Wasserstein GAN [182] or style transfer [183].

In this section we will start in Sec. 3.3.1 with Optimal transport fundamentals, allowing to map one distribution to another and to compare them. Then moving toward our goal, we will see in Sec. 3.3.2 extensions of the original optimal problem allowing to interpolate distributions. This will provide the basic element for our goal, providing a complete description of all the intermediary states between the two models.

3.3.1 Fundamental concepts of optimal transports

Monge problem

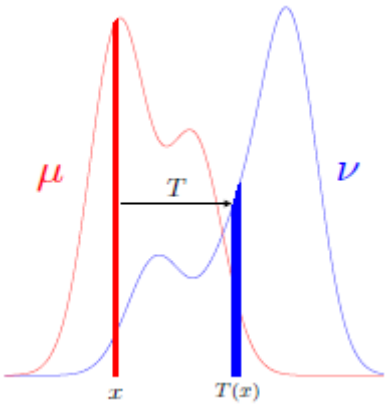


FIGURE 3.5 – Illustration of Monge Map from μ to ν, from [14]

Qualitatively, the fundamental problem of optimal transport introduced by Monge [173] considers the optimal displacement of an amount of earth to another location as illustrated in Fig. 3.5. The two amounts of earth are called μ and ν, and the displacement is defined by the function T. This function associates each small amount of earth of μ at location $\mathbf{x}$ to a location

$T(\mathbf{x})$ in ν. All the earth at a particular location is moved to a unique location. Moreover, once that T is applied at each location, ν should be recovered exactly. The important idea is that we are looking for the optimal transport T map which minimizes the global effort of moving an amount of earth to the other.

Mathematically, μ and ν are seen as probability distributions defined respectively on $\mathcal{X}$ and $\mathcal{Y}$ and $T : \mathcal{X} \to \mathcal{Y}$. The effort minimization problem corresponds to

$$\min_{T:T_{\#}\mu=\nu} \int_{\mathcal{X}} c(\mathbf{x}, T(\mathbf{x}))d\mu(\mathbf{x}) \tag{3.3}$$

In this formulation $c(\mathbf{x}, T(\mathbf{x}))d\mu(\mathbf{x})$ corresponds to the marginal effort at location $\mathbf{x}$, which is proportional to a cost $c(\mathbf{x}, T(\mathbf{x}))$ of moving from $\mathbf{x}$ to $T(\mathbf{x})$ and to the amount of mass $d\mu(\mathbf{x})$. In Monge's original formulation, this cost corresponded to the Euclidean distance, but a more general cost can be used. This marginal effort is then integrated over the whole space $\mathcal{X}$ to provide a complete displacement cost. Finally $T_{\#}\mu = \nu$ imposes, that once transported by T, μ must be equal to ν. $\#$ is the push-forward operator giving the image measure of μ by T, such that $T_{\#}\mu(A) = \mu(T^{-1}(A))$

Kantorovitch problem

The above Monge problem faces several mathematical limitations. By definition, this problem is asymmetric regarding μ and ν, and uniqueness and existence are not guaranteed in general. The problem tackled by Kantorovitch [174], is a relaxation of Monge's problem addressing this problem. Qualitatively, the Kantorovitch problem modify the fact that the amount of earth at a location $\mathbf{x}$ can only be moved to a unique location. Instead, the earth can be spread along all the space.

Mathematically it means that the marginal mass at a position $\mathbf{x} \in \mathcal{X}$ is spread on $\mathcal{Y}$ according to a probability distribution. In global, instead of looking for a map T, we are now globally looking for a probability distribution π^* on the Cartesian product $\mathcal{X} \times \mathcal{Y}$ which should be solution of

$$\min_{\pi \in \mathcal{U}(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(\mathbf{x}, \mathbf{y})d\pi(\mathbf{x}, \mathbf{y}) \tag{3.4}$$

where

$$\mathcal{U}(\mu, \nu) = \left\{\pi \in \mathcal{M}_+^1(\mathcal{X} \times \mathcal{Y}) : P_{\mathcal{X}\#}\pi = \mu \text{ and } P_{\mathcal{Y}\#}\pi = \nu\right\}. \tag{3.5}$$

This last line imposes that each π is a probability distribution with marginals μ and ν, it is a transport between these two distributions. The optimization (3.12) can be interpreted similarly to the Monge problem. The marginal effort is this time $c(\mathbf{x}, \mathbf{y})d\pi(\mathbf{x}, \mathbf{y})$ with still $c(\mathbf{x}, \mathbf{y})$ the unitary cost of moving from $\mathbf{x}$ to $\mathbf{y}$ but this time all the mass at $\mathbf{x}$ is not transported to a single point. So $d\pi(\mathbf{x}, \mathbf{y})$ corresponds to the mass transported in particular to $\mathbf{y}$. Integrating over $\mathbf{y}$ provides the effort of moving the mass at $\mathbf{x}$ and the second integration over $\mathbf{x}$ provides the total effort.

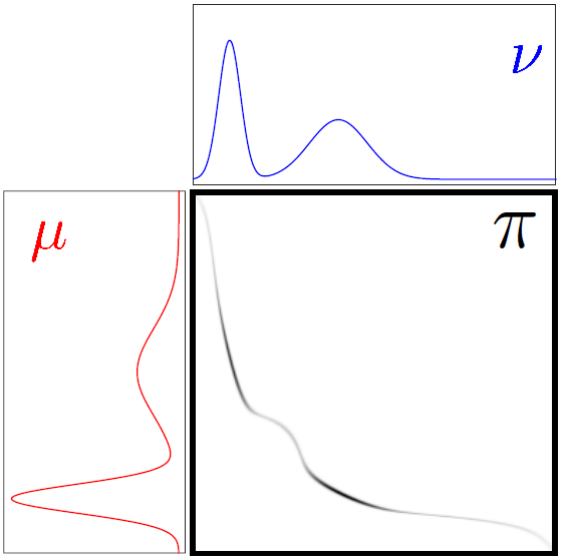


FIGURE 3.6 – Illustration of the solution of the Kantorovitch problem with the optimal transport plan, from [11]

This optimal distribution π^* is called the (optimal) transport plan and figure 3.6 illustrates the solution.

In the standard case where $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ and $c(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^2$. In that case, a famous result referred to as the Brenier theorem [184, 185] proves the existence and the unicity of the optimal transport plan. Moreover, the solution is characterized by providing an equivalence between the Monge and Kantorovitch problem. Indeed the optimal transport plan π is equal to $(Id, T)_{\#}\mu$, that is why, in Fig. 3.6 the mass of probability lies on a line, the line $(\mathbf{x}, T(\mathbf{x}))$. This theorem also details the properties of the Monge map which is the gradient of a convex function.

Wasserstein distance

In the previous paragraph, we focused on the optimal transport plan, but the value of the least effort is also of high interest in itself. Assuming that $\mathcal{X} = \mathcal{Y}$ and that the cost c is a distance denoted d at a power p, the minimal effort provides a rigorous distance between distributions

$$\mathcal{W}_p(\mu, \nu) = \left(\min_{\pi \in \mathcal{U}(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} d(\mathbf{x}, \mathbf{x}')^p d\pi(\mathbf{x}, \mathbf{x}') \right)^{1/p} \tag{3.6}$$

In practice, mainly two particular cases of this distance are considered. The case $p = 1$ was extensively utilized in the image processing community in the 2000s, referred as the earth mover distance [180]. Nowadays the 2-Wasserstein distance, case $p = 2$, is the most commonly used, because of the numerical advantages in its computation.

The displacement definition of the Wasserstein distance leads to a physically more satisfying

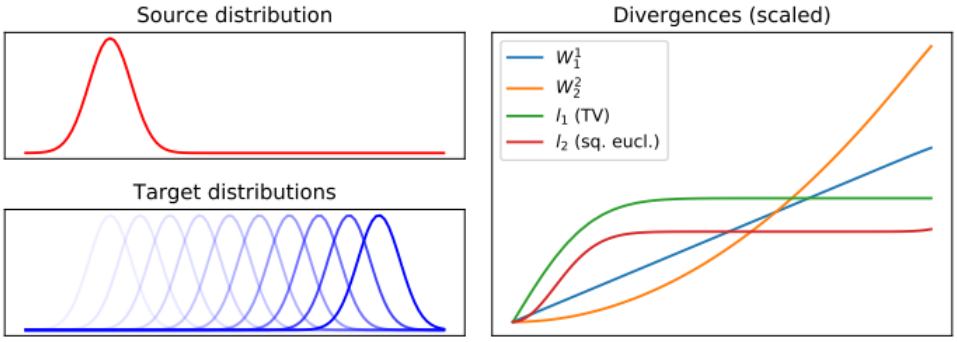


FIGURE 3.7 – Comparison of the Wasserstein distance with the other standard distance over distributions l_1 and l_2, from POT [186] documentation

behavior compared to other classical distances l_1 and l_2 distances, as illustrated in Fig 3.7. Moving progressively one distribution from the other, for the two latter cases the distance remains constant since both supports are disjoint, whereas it keeps increasing for the Wasserstein distance. This effect has also motivated the machine learning community to use it as a more convenient loss function avoiding vanishing gradient, for example in Wasserstein GAN [182].

Particular case of Gaussian distributions

In general, solving the optimal transport problem can be challenging. However explicit or numeric solutions can be found in important particular cases. A solution in the case of discrete distributions can be computed numerically and benefits from computational acceleration [187]. An explicit solution is available in the univariate case, allowing applications on density data. Finally, a result of high interest for us is that the Gaussian case benefits from explicit solutions. Given

$$\mu \sim \mathcal{N}(\mathbf{m}_\mu, K_\mu) \quad \nu \sim \mathcal{N}(\mathbf{m}_\nu, K_\nu), \tag{3.7}$$

The Monge map corresponds to

$$T : \mathbf{x} \mapsto \mathbf{m}_\nu + A(\mathbf{x} - \mathbf{m}_\mu) \tag{3.8}$$

where

$$A = K_\mu^{-1/2}(K_\mu^{1/2} K_\nu K_\mu^{1/2}) K_\mu^{-1/2} \tag{3.9}$$

Similarly, the Wasserstein distance can be computed in close form corresponding to

$$\mathcal{W}_2^2(\mu, \nu) = \|\mathbf{m}_\mu - \mathbf{m}_\nu\|^2 + tr\left(K_\mu + K_\nu - 2\left(K_\mu^{1/2} K_\nu K_\mu^{1/2}\right)^{1/2}\right) \tag{3.10}$$

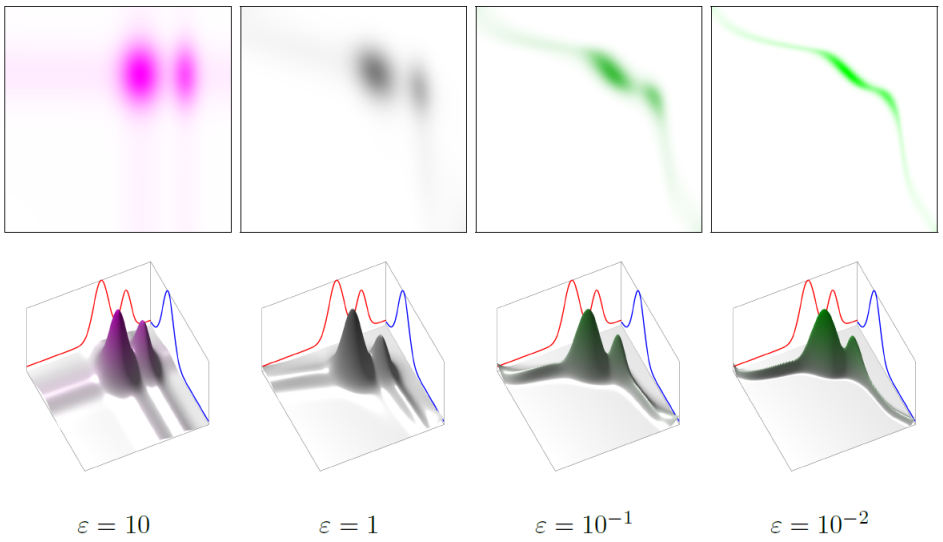


FIGURE 3.8 – Influence of ε on the optimized transport plan, from [11]

This corresponds to the Wasserstein distance between multivariate Gaussian distributions. Mallasto *et al.* [188] treated the case of the Wasserstein distance in the infinite dimension case, with Gaussian processes. They demonstrated that as the number of dimensions converged to infinity, the distance in the multivariate case converged toward the infinite dimension distance. So we can approximate the distance between Gaussian processes by the distance between multivariate distributions.

Extensions of the classical optimal transport problem

The Monge and Kantorovitch problems represent the standard optimal transport problem. Many variations of these problems have been considered in the literature to overcome some of its limits. We present here shortly the most famous ones.

Probably the most well-known variation of the optimal transport problem is the addition of a regularization, transforming problem (3.12) into

$$\min_{\pi \in \mathcal{U}(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(\mathbf{x},\mathbf{y}) d\pi(\mathbf{x},\mathbf{y}) + \varepsilon\gamma(\pi) \tag{3.11}$$

with γ the regularization term, the function of the transport plan, and ε the hyperparameter quantifying the importance of the regularization. The most well-known regularization is the entropic one proposed by Cuturi [187] in case of discrete measures which reduces significantly the computation time. Fig 3.8 illustrates the influence of ε on the optimal transport plan. For small values of ε we recover the solution to the Kantorovitch problem, equivalent to the Monge solution, with approximately all mass on a line. But then as ε increases, the mass is progressively spread around this line leading to a smoother solution.

This idea opened the way to computational optimal transport [165] with many new applications in machine learning. Recently, several works proposed similar regularization in the particular case of Gaussian measures [189, 149, 190]. Another popular approach to reduced computation time in the case of multivariate distributions is the sliced optimal transport [191]. This approach benefits from the ease of computation in the univariate case. It projects repetitively the measures on unidimensional subspaces, approximating the complete objective.

An important constraint of optimal transport is that it is mostly limited to measures on similar spaces, the limitation being caused by the need to compute a cost function. In particular, this issue occurs when we need to compare distributions of different dimensions. A proposed solution was to consider the Gromov variant of optimal transport [192]. In the general setting of two distributions $\alpha_{\mathcal{X}}$ and $\alpha_{\mathcal{Y}}$ respectively defined on the spaces $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$, $d_{\mathcal{X}}$ and $d_{\mathcal{Y}}$ being the respective distances, the Gromov-Wasserstein problem takes the form

$$\min_{\pi \in \mathcal{U}(\mu,\nu)} \int_{\mathcal{X}^2 \times \mathcal{Y}^2} |d_{\mathcal{X}}(\mathbf{x}, \mathbf{x}') - d_{\mathcal{Y}}(\mathbf{y}, \mathbf{y}')|^2 d\pi(\mathbf{x}, \mathbf{y}) d\pi(\mathbf{x}', \mathbf{y}') \tag{3.12}$$

This was required, for example, for the application of optimal transport to graph data of different sizes [193] and recent developments were also provided in the Gaussian case [194].

Finally, another important variation of the optimal transport problem is its unbalanced version [195] where marginals have different total mass. This variant permits the creation of mass, which may be interesting for example when distributions represent objects of different volumes, or in case of noisy data improving the robustness of the transport. A recent review of different versions of this problem can be found in [196] with particular developments in the Gaussian case [149].

3.3.2 Interpolation of distributions with optimal transport

The standard optimal transport problems consider a mapping, deterministic in the Monge case, and probabilistic for the Kantorovitch one, describing where each amount of mass of a distribution should be placed in the second one. However, considering our use case, we do not want a mapping, but an interpolation between each of the observed conditions with detailed information on all the intermediary states. Coming back to the earth-moving analogy we are not only interested in the optimal allocation of each amount of mass from one distribution to another, but we are also interested in their complete trajectory. This section will consider this issue and explain how it has been generalized from two to multiple distributions.

Interpolation between two distributions

The continuous version of the Kantorovitch problem has been formalized by Benamou and Brenier [197]. Replacing the notation of μ and ν by α_0 and α_1 with the idea that we are looking for a full displacement of α_t, t ranging from 0 to 1, each α_t being a probability distribution. At each time the distribution is advected following a vector field denoted v_t. The idea is to compute the path $(\alpha_t)_{t=0}^1$ of minimal length, which leads to the reformulation of the optimal transport problem

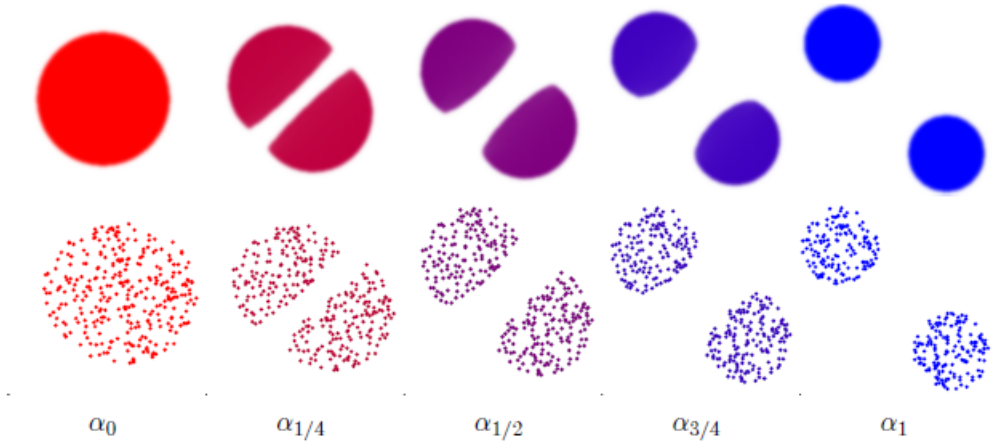


FIGURE 3.9 – Displacement between two measures in the continuous and discrete case, from [11]

$$\mathcal{W}_2^2(\alpha_0, \alpha_1) = \min_{(\alpha_t, v_t)_t} \int_0^1 \int_{\mathbb{R}^d} \|v_t(\mathbf{x})\|^2 d\alpha_t(\mathbf{x}) dt \tag{3.13}$$

under a constraint of mass conservation

$$\frac{\partial \alpha_t}{\partial t} + \operatorname{div}(\alpha_t v_t) = 0. \tag{3.14}$$

An illustration of this optimal path is displayed in Fig. 3.9 for both a continuous and a discrete case.

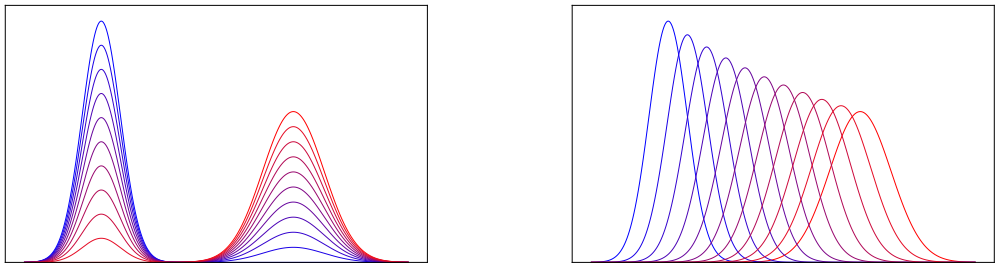
When the Monge problem has a solution T the previous displacement problem is equivalent to McCann's interpolation [12], which has an explicit formulation,

$$\alpha_t = ((1-t)Id + tT)_{\#}\alpha_0. \tag{3.15}$$

In particular, we saw that the Monge problem had an explicit solution in the Gaussian case. Thus the McCann's interpolation of two distributions can be computed in close form. For $\alpha_0 \sim \mathcal{N}(\mathbf{m}_0, K_0)$ and $\alpha_1 \sim \mathcal{N}(\mathbf{m}_1, K_1)$, α_t is Gaussian for any t, $\alpha_t \sim \mathcal{N}(\mathbf{m}_t, K_t)$ with

$$\mathbf{m}_t = (1-t)\mathbf{m}_0 + t\mathbf{m}_1, \qquad K_t = ((1-t)I + tA)K_0((1-t)I + tA) \tag{3.16}$$

Mathematically, this path corresponds to a geodesic between the two distributions. More qualitatively, the interpolation provided by the optimal transport is considered more physically sounded than more classical interpolations. In Fig. 3.10 we compare the interpolation of two 1D Gaussian distributions proposed by a standard Euclidean interpolation versus an the optimal transport distribution. We observe that the Euclidean interpolation corresponds to a vertical



(a) (b)

FIGURE 3.10 – Interpolation path between two univariate Gaussian distribution considering (a) the Euclidean distance, (b) the Wasserstein distance

displacement against an horizontal displacement for optimal transport. The optimal transport method is considered more coherent physically since it corresponds to a real displacement of the distribution whereas in the Euclidean case some mass can appear suddenly. Similarly, to the standard optimal transport problem, variants of the displacement can be considered with a regularization term [189] or with a Gromov version in case of different dimensions [194].

Interpolation between several distributions with Wasserstein barycenter

McCann's interpolation provides a useful method to interpolate between two distributions, still, it can seem limited considering our aim. Indeed it is only a unidirectional interpolation, whereas we would be interested in an interpolation between several distributions with several experimental factors. Moreover, it only exploits the knowledge of two distributions whereas we have several observed model in our case, meaning that it cannot exploit all the available information. Thus, we need a generalization of the above method, interpolating between several distributions. A popular way to solve this problem using optimal transport is the use of Wasserstein barycenters.

Barycenters are well-known in the Euclidean case where they correspond to a weighted mean of some vectors. Considering vectors $(\mathbf{x}_1, \ldots, \mathbf{x}_m)$ and associate weights, also called barycentric coordinates, $(\lambda_1, \ldots, \lambda_m)$, the barycenter $\bar{\mathbf{x}}$ is equal to

$$\bar{\mathbf{x}} = \sum_{j=1}^{m} \lambda_j \mathbf{x}_j \tag{3.17}$$

An interesting property of barycenters is that considering a fixed family of vectors and making barycentric coordinates vary one can obtain any points from the convex hull of these vectors. So we say that barycenters define an interpolating surface between vectors. In the special case of two distributions, the convex hull is simply the segment between the two points.

This notion of barycenter can extend to more complex objects. Indeed in the Euclidean case, a barycenter can equivalently be seen as the solution to the following problem

$$\bar{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{R}^n}{\arg\min} \sum_{j=1}^{m} \lambda_j \|\mathbf{x}_j - \mathbf{x}\|^2. \tag{3.18}$$

The idea to generalize the barycenters to more complex objects is to replace in this formula the Euclidean distance by the appropriate distance of the interest space. So, barycenters can eventually be generalized to any metric space, and it was introduced in the case of distributions using the Wasserstein distance in [191] from an applied viewpoint and [13] from a mathematical perspective. Considering m distributions α_j over $\mathbb{R}^d$ their Wasserstein barycenter is thus the solution to the following problem

$$\bar{\alpha} = \underset{\alpha\in\mathcal{P}(\mathbb{R}^d)}{\arg\min} \sum_{j=1}^{m} \lambda_j \mathcal{W}_2^2(\alpha_j, \alpha), \tag{3.19}$$

with $\mathcal{P}(\mathbb{R}^d)$ the set of probability distributions over $\mathbb{R}^d$.

In the case of m Gaussian distributions $\alpha_j \sim \mathcal{N}(\mathbf{m}_j, K_j)$, the barycenter with coordinates λ_j, is also a Gaussian distribution $\bar{\alpha} \sim \mathcal{N}(\bar{\mathbf{m}}, \bar{K})$, with

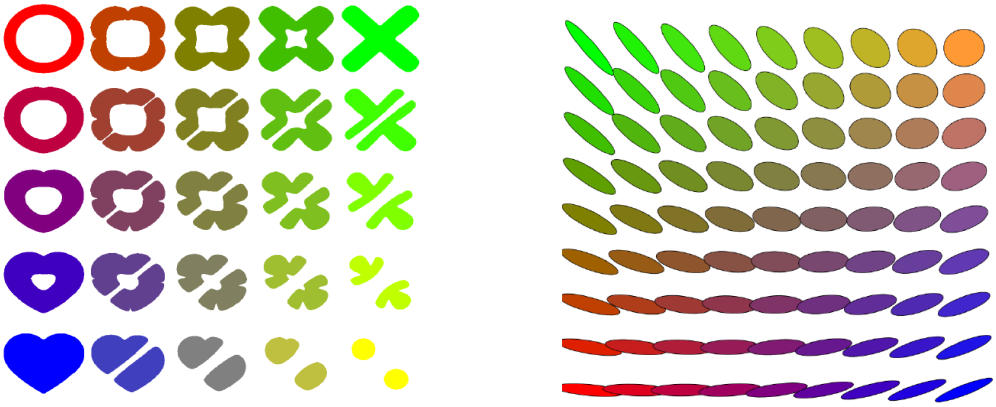
$$\bar{\mathbf{m}} = \sum_{j=1}^{m} \lambda_j \mathbf{m}_j \quad \text{and} \quad \bar{K} = \sum_{j=1}^{m} \lambda_j \left(\bar{K}^{\frac{1}{2}} K_j \bar{K}^{\frac{1}{2}}\right)^{\frac{1}{2}}. \tag{3.20}$$

The covariance matrix cannot be directly computed from this formula ($\bar{K}$ being on the left and right part of the equation). One should use the iterative algorithm presented by Álvarez-Esteban *et al.* [198]. Starting from an initialization $\bar{K}_{(0)}$ the iteration

$$\bar{K}_{(s+1)} = \bar{K}_{(s)}^{-1/2} \left(\sum_{j=1}^{m} \lambda_j \left(\bar{K}_{(s)}^{\frac{1}{2}} K_i \bar{K}_{(s)}^{\frac{1}{2}}\right)^{\frac{1}{2}}\right)^2 \bar{K}_{(s)}^{-1/2} \tag{3.21}$$

applied up to convergence.

Analogously to the Euclidean case, Wasserstein barycenters provide an interpolating surface between distributions. Fig. 3.11a shows the interpolation between four images at the corners seen as discrete probability and similarly Fig. 3.11b displays them for Gaussian distributions. Agueh and Carlier [13] demonstrated that in the case of two distributions, the obtained path corresponds exactly the McCann's interpolation. It can be observed in previous figures by looking at the sides linking two corners. So the Wasserstein barycenters benefit from the same physical interest as the McCann's interpolation. They gained an important success in application through the development of faster optimization methods benefiting from regularization methods [199].



(a) (b)

FIGURE 3.11 – Barycentric interpolation surface between (a) four discrete distributions seen as images (b) four Gaussian distributions, from [11]

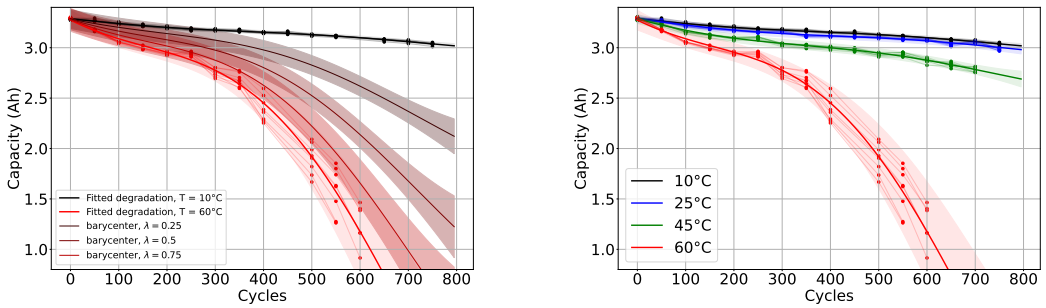
3.4 Conditional Wasserstein barycenters to model the condition effect

The previous section introduced the main elements of optimal transport theory, notably the optimal transport map, the Wasserstein distance, and the corresponding barycenters. Now we will present how these elements can be used to model the experimental factors effects and provide a regression method over distribution outputs. We will motivate the general idea of a conditional Wasserstein barycenter [200, 201], a flexible way to generalize the standard regression. To mitigate the small number of observed experimental conditions we will focus its use on a model-based approach. Additionally, to keep the performances previously obtained we will also restrict to interpolating methods. Two such models will be derived, one using the ideas of structured regression, adapting classical machine learning ideas, and another using the Fréchet regression, generalizing the standard linear regression.

3.4.1 Conditional Wasserstein barycenter

The interpolation methods presented in the previous section can be applied in our use case. For example, we can apply the Gaussian process-based methods described in the previous chapter independently at two temperatures say 10 and 60°C. This provides two Gaussian distributions α_0 and α_1 which can thus be interpolated using McCann's interpolation. The result is displayed in Fig. 3.12a providing us with a detailed interpolation of the two degradation curves with uncertainties.

The continuous formulation of the optimal transport produces an interesting interpolation between two or more distributions, still, it does not really answer to our goal. We want to predict a degradation model at an unobserved condition. In our case, we would like a prediction at 50°C. Yet McCann's interpolation only provides a path indexed by an abstract value between 0 and 1, that cannot be related directly to a temperature. The same remark remains true for barycenters



(a) (b)

FIGURE 3.12 – Comparison between (a) the barycentric interpolating path between the two extreme temperatures and (b) the actual fitted models at each temperatures

with several abstract indexes.

Still, this interpolation seems useful. Qualitatively, the provided path seems to pass closely to the models fitted at intermediary conditions. Visually, this can be seen, by comparing the interpolating path with the fitted curves in Fig. 3.12b. We see that the degradation at 25°C is close to the one at 10°C so it could be well represented by a barycenter with λ close to 0, whereas the degradation at 45°C may be more similar to a barycenter with λ close to 0.25.

This idea can be formulated more rigorously with the concept of barycentric coordinate regression, used by Boneel *et al.* [202]. It defines the idea of closeness by computing the minimal Wasserstein distance between the considered distribution and the barycentric surface (or McCann's interpolation path in case of only two distributions). Thus they considered the following optimization problem

$$\underset{\boldsymbol{\lambda}\in\Sigma_S}{\arg\min}\, \mathcal{W}_2^2(\bar{\alpha}_{\boldsymbol{\lambda}}, p) \tag{3.22}$$

with p the distribution that we want to approximate, $\alpha_{\boldsymbol{\lambda}}$ the barycenter of available distributions with coordinates $\boldsymbol{\lambda}$, and Σ_S the simplex of barycentric coordinates summing to one. In fact, this optimization corresponds to the projection of a distribution on the barycentric interpolating surface.

Using this method, we consider the projection of the degradation at 25 °C and 45°C on the interpolating path between the degradation at 10 °C and 60 °C. The interpolation path depends only on one real index λ so solving problem (3.22) consists in finding the optimal t value. The results of the projection are compared with the true distribution in Fig 3.13a for experiments at 25 °C and Fig 3.13b for experiments at 45 °C. Only two distributions are used to build the interpolating path but the projection already provides interesting results. By using the interpolation surface between several distributions the research space would be broader and closer results could be obtained.

This example illustrates that the surface generated by barycenters can potentially lead to an interesting prediction under the condition that the barycentric coordinates are well-selected. This

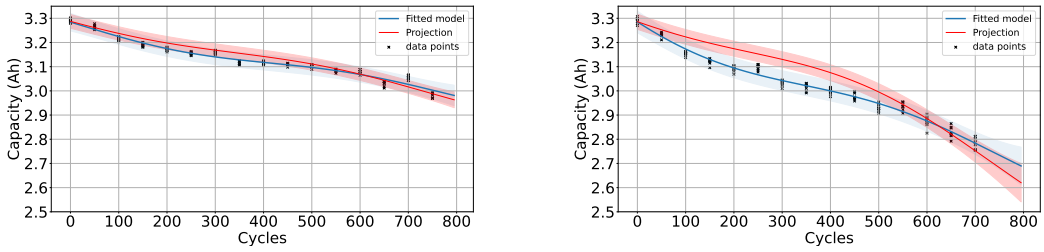


FIGURE 3.13 – Prediction versus true data and associated fitted model at (a) 25 °C and (b) 45 °C.

motivates the introduction of the conditional Wassertein barycenter or equivalently conditional Fréchet mean [201, 200]. Using $(\mathbf{z}_j)_{j=1}^m \subset (\mathbf{z}_k)_{k=1}^K$ it takes the form

$$f(\mathbf{c}) = \arg\min_{\mathbf{z}} \sum_{j=1}^{m} \lambda_j(\mathbf{c}) \mathcal{W}_2^2(\mathbf{z}_j, \mathbf{z}), \tag{3.23}$$

a barycenter of the $\mathbf{z}_k$ with coordinates depending on the experimental factor, the term "conditional" referring to this dependence. Note that this barycenter is computed between m distributions with m not necessarily equal to K. The value of m will depend on the particular method, the more distributions included in the barycenter, the more flexible the prediction, but with a potentially higher risk of overfitting (for $m = n$ all degradations can be interpolated exactly).

Written in this way we clearly see that the considered problem is a regression problem. The method considers explicitly the space of experimental factors. The conditional barycenter should be understood as a generalization of the conditional expectation used for regression on standard outputs, as the barycenter is a generalization of the expectation. We explained previously that our focus was on interpolating methods, predicting the observed distributions at observed conditions. But instead of considering a simple interpolation between distributions as in the previous section, with McCann method or barycenters, we will consider an interpolating regressor, interpolating the distributions while taking into account their position in the space of experimental factors.

Conditional barycenters provide a general framework for predictions on complex outputs, depending on the choice of the coordinates functions λ_j. The chosen model for these coordinates can be either model-based or data-driven. Here, because of the small number of observed conditions, we will use a model-based approach, specifying the shape of the λ_j functions. This will permit more robust predictions through the inclusion of physical knowledge on the effect of the factors compensating for the lack of data.

3.4.2 Structured regression model

Our first proposal of conditional Wasserstein barycenter relies on the structured regression framework. Here, coordinates are seen as parametric functions that need to be learned thanks to the training data set [201]. The λ_j are seen as parametric functions $\lambda_{j,\theta}(\mathbf{c})$, depending on a parameter θ leading to the following estimator

$$f_\theta(\mathbf{c}) = \arg\min_{\mathbf{z}} \sum_{j=1}^{m} \lambda_{j,\theta}(\mathbf{c}) \mathcal{W}_2^2(\mathbf{z}_j, \mathbf{z}). \tag{3.24}$$

Any model can be used to learn the λ_j functions, for example, neural networks were used in [201]. However, here, the number of observations is too limited so we will prefer a model-based method, explicitly modeling the weights with a physical interpretation.

We consider the case of the temperature as the unique experimental factor ($\mathbf{c} = T$). In the previous section, we considered the interpolation between the two extreme conditions (the one at 10°C, denoted $\mathbf{z}_1$, and the one at 60°C, denoted $\mathbf{z}_2$). Since it is only a barycenter between two elements it takes the following form

$$f_\theta(T) = \arg\min_{\mathbf{z}} \left(1 - \lambda_\theta(T)\right) \mathcal{W}_2^2(\mathbf{z}_1, \mathbf{z}) + \lambda_\theta(T) \mathcal{W}_2^2(\mathbf{z}_2, \mathbf{z}). \tag{3.25}$$

Looking back to the interpolating path displayed in Fig. 3.12a, $\lambda_\theta(T)$ can be interpreted as the level of degradation, when it is equal to 0 we are at the level of degradation observed at 10°C, and when it is equal to 1, we are at the level observed at 60°C. With this interpretation, we can propose a specific physical model based on its evolution with temperature. So chose the following empirical model

$$\lambda_{\boldsymbol{\theta}}(T) = a_0 - a_1 \exp(a_2 T), \tag{3.26}$$

corresponding to the model proposed by Diao et al. [72] for a fixed time, with in that case $\boldsymbol{\theta} = (a_0, a_1, a_2)$. It allows notably a monotonic and accelerating degradation with temperature.

The predictor is now specified but we still need to estimate $\boldsymbol{\theta}$. To this end, we use the structured regression approach, originally propoased by Tsochantaridi *et al* [203]. Considering a general setting with $X \in \mathcal{X}$ as input, $Y \in \mathcal{Y}$ as output and $\mathcal{L} : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ as loss function, learning the parameter $\boldsymbol{\theta}$ corresponds to finding $\boldsymbol{\theta}^*$ such that the expected risk, the average error of our predictor,

$$\mathcal{R}(\boldsymbol{\theta}) = \mathbb{E}\left[\mathcal{L}(f_{\boldsymbol{\theta}}(X), Y)\right], \tag{3.27}$$

is minimized. Here, the expectation is taken over the joint distribution of (X, Y). However, this distribution is unknown, so the expected risk cannot be computed. Instead, given a dataset $(X_i, Y_i)_{i=1}^n$, its empirical counterpart is considered

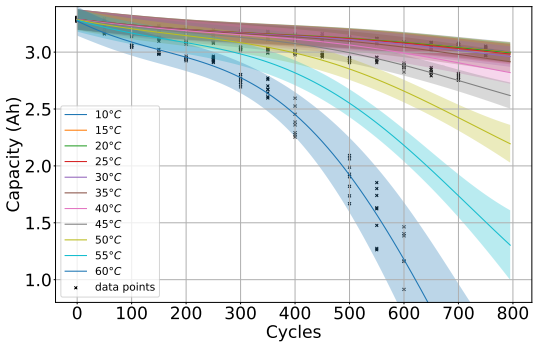


FIGURE 3.14 – Structured regression predictions

$$\hat{\mathcal{R}}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}(f_{\boldsymbol{\theta}}(X_i), Y_i). \tag{3.28}$$

This one can be computed and minimized. In the usual regression case $\mathcal{X} = \mathbb{R}^p$, $\mathcal{Y} = \mathbb{R}$ and $\mathcal{L}$ is a squared error or a negative log-likelihood. However, it is not limited to this setting. In our case, $\mathcal{Y} = \mathcal{N}^d$ so we can use the 2-Wasserstein distance as loss function, leading to an empirical risk of

$$\hat{\mathcal{R}}(\boldsymbol{\theta}) = \frac{1}{K}\sum_{k=1}^{K} \mathcal{W}_2^2(f_{\boldsymbol{\theta}}(\mathbf{c}_k), \mathbf{z}_k). \tag{3.29}$$

We optimized this method thanks to standard automatic differentiation methods, but alternatively, the gradient could be computed explicitly, possibly in a stochastic version [204].

We applied this method on the four degradations of Fig 3.1a, corresponding to four temperatures and with a unique CCC of 0.2 C and a unique discharge rate of 0.7C. The predictions are displayed in Fig. 3.14. We observed that predictions at 10 and 60°C are almost equal to the time degradation fitted on the data since they were used to build the interpolating path and are reached exactly respectively for a λ coordinate of 0 and 1. Qualitatively, predictions seem interesting, monotonicity constraints either in time or in temperature are respected and acceleration of the degradation is also taken into account. However, the fit is still limited, the predicted curves do not interpolate at 25 and 45°C the originally fitted models. An important limitation of this method is that it uses the two extreme distributions to create the interpolating path, whereas other data are only used to fit the coordinates function. We were able to set an empirical model because coordinate only depend on one value λ that can be interpreted. Setting a model on several coordinate parameter would be a much more complex task. This limits the interpolation to a unique experimental factor and does not exploit all the available knowledge. A barycenter computed on all the observed conditions should reach a better accuracy.

3.4.3 Fréchet regression model

To overcome these limitations, we propose the use of another framework, the Fréchet regression [1][200], which proposes another way to select the barycenter coordinates. This time the method is based on an analogy with the standard linear regression, but it will still permit a similar physical prior knowledge inclusion. Recalling that the prediction of the linear regression can be written as an Euclidean barycenter with known weights, the authors proposed to use for prediction a barycenter with the same weights. Thus, in this case, we have a close form of the barycentric coordinates.

Linear Fréchet regression

Explicitly, let us assume, for now, that the complex outputs $\mathbf{z}$ are in fact simple real outputs $z \in \mathbb{R}$. Considering p experimental factors $\mathbf{c} = (c_1, \dots, c_p)$, the standard linear regression takes the form,

$$z = \beta_0 + \beta_1 c_1 + \cdots + \beta_p c_p \tag{3.30}$$

Using the data $(\mathbf{c}_k, z_k)_{k=1}^K$, an estimation $\hat{\beta}$ of $\beta = (\beta_0, \cdots, \beta_p)$ can be obtained as solution of the mean square problem, leading to the prediction $z^* = \hat{\beta}_0 + \hat{\beta}_1 c_1^* + \cdots + \hat{\beta}_p c_p^*$. Importantly, these predictions can be rewritten as a linear combination of the observations

$$z^* = \sum_{k=1}^{K} \lambda_k(\mathbf{c}^*) z_k \tag{3.31}$$

with weights

$$\lambda_k(\mathbf{c}^*) = 1 + (\mathbf{c}_k - \bar{\mathbf{c}})^T \hat{\Sigma}^{-1} (\mathbf{c}^* - \bar{\mathbf{c}}), \tag{3.32}$$

where $\bar{\mathbf{c}} = K^{-1} \sum_{k=1}^K \mathbf{c}_k$ and $\hat{\Sigma} = K^{-1} \sum_{k=1}^K (\mathbf{c}_k - \bar{\mathbf{c}})(\mathbf{c}_k - \bar{\mathbf{c}})$.

Since, for a fixed input c, weights sum to 1, this weighted mean corresponds well to an Euclidean barycenter with corresponding coordinates. In a general context the authors proposed to use these weights to compute a conditional barycenter. Later, Fan et al. [205] focused on the particular case of multivariate distributions with the Wasserstein barycenter. To make predictions in the second step we will thus compute the corresponding conditional Wasserstein barycenter on the complex outputs $\mathbf{z}_k \in \mathcal{N}^d$ instead of $z_k \in \mathbb{R}$.

This method provides us with a well-defined predictor, yet, we may think that a linear method can be too restrictive, notably to model the acceleration of degradation with the temperature. Indeed we can see in Fig. 3.15 the linear Fréchet regression, applied to the same four conditions

1. In their article the authors introduced both a local and global version of the Fréchet regression. Here we will deal with the global version which corresponds to a model-based approach

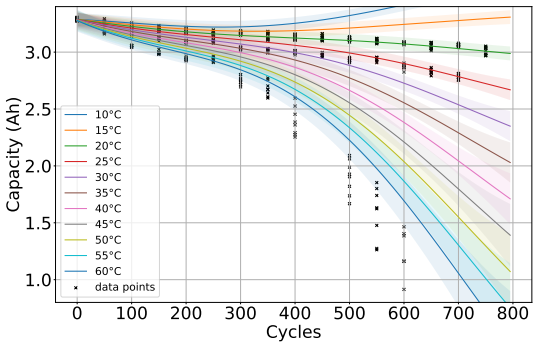


FIGURE 3.15 – Linear Fréchet regression predictions

than in the previous section, provides a severe underfitting of the data. Focusing on the mean, we saw on Eq. (3.20) that for Gaussian distributions the barycenter mean is simply a weighted average of the mean vectors. As a result, Fréchet regression does a linear regression on the distribution means. For a particular time t, corresponding to a component of the mean vector $\bar{\mathbf{m}}$, the SoH evolution with temperature will only be a straight line. So the predicted capacity evolution with temperature is linear. The capacity loss between, 40 °C and 60 °C is the same as the one between 10 and 30 °C, which is clearly not the case in practice.

Polynomial Fréchet regression

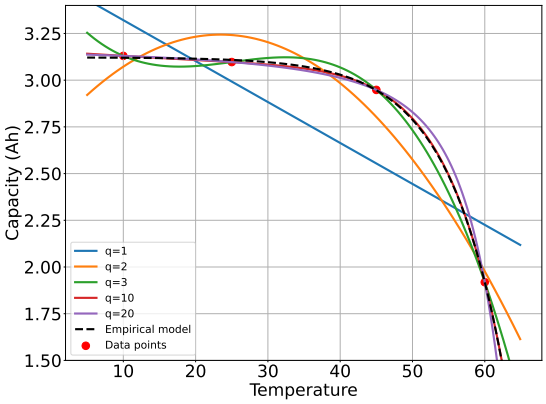


FIGURE 3.16 – Comparison of polynomial regressions of different degrees q with the empirical model on the temperature degradation

In standard linear regression, this kind of limitation can be overcome by using a polynomial regression model. In the case of a unique experimental factor c, the polynomial regression corresponds to

$$z = \beta_0 + \beta_1 c + \cdots + \beta_q c^q, \tag{3.33}$$

which is just a particular case of linear regression with $\mathbf{c} = (c, \cdots, c^q)$ as inputs. So we can directly obtain a polynomial version of the Fréchet regression using the weight formula (3.32) with the extended inputs.

In fact, this polynomial fit provides a similar behavior to the empirical model (3.26) proposed in the previous paragraph. As illustrated in Fig. 3.16 we observed that, regarding the mean of the trend component, polynomial regression is well-suited to fit the temperature evolution. Increasing the polynomial degree, we progressively improve the fit to the data up to interpolation for $q = K - 1$. Furthermore, continuing to increase the polynomial degree can lead to additional benefits. The data are still interpolated but smoother monotone results can be obtained. In fact, for a well-selected degree, we can obtain an approximation of the empirical model (3.26) of the previous section, respecting monotonicity and acceleration. Additionally, compared to the previous approach this method presents the advantage of interpolating all the $\mathbf{z}_k$ not only the one at extreme conditions, improving the global fitting.

This method depends on the polynomial degree p. It is selected to provide the closest result to the empirical model. A more standard choice would have been to select it by a cross-validation method yet, in this setting, we do not have enough data. The difference with the previous model is that we had only one empirical model describing the full evolution, through λ, whereas here we will have for the mean, an independent model for each time. This allows a more complex model for the mean. The interest of this method is clear for the mean and we similarly apply it to the covariance matrix, which remains coherent since we observed an accelerating increase of the variance with temperature.

Considering the previous developments the needed model is generally over-parametrized (the number of parameters is higher than K). This leads to a computational difficulty since in that case, the covariance matrix, $\hat{\Sigma}$, is no more invertible. To avoid this issue we propose to replace the coordinate formula by

$$\lambda_i(\mathbf{c}^*) = 1 + (\mathbf{c}_i - \bar{\mathbf{c}})^T \hat{\Sigma}^+ (\mathbf{c}^* - \bar{\mathbf{c}}), \tag{3.34}$$

with $^+$ the Moore Penrose pseudo-inverse [206]. In case of real outputs, we can show that the predictions of the over-parametrized linear regression are linear combinations of the outputs with these weights (more details provided in the Appendix C). So as Fréchet regression extends the standard linear regression, we can use a generalization of the over-parametrized linear regression. Thus, we see that even if the linear regression is a constrained parametric framework it has an interesting flexibility.

We displayed, in Fig.3.17 the results obtained with the polynomial version of the Fréchet regression. As for the structured regression, the predictions respect the monotonicity and acceleration constraints. But, additionally, all data curves are now interpolated, which then leads to more coherent interpolation between them. The use of a barycenter of all the observed degradation provides well a more precise modeling.

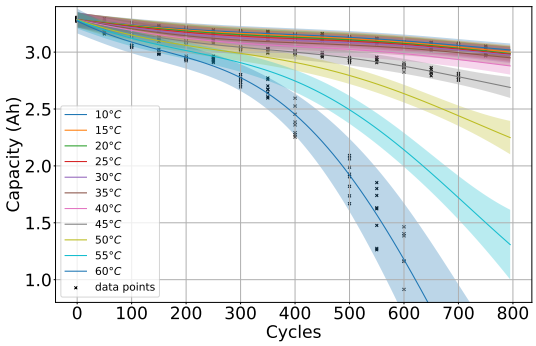
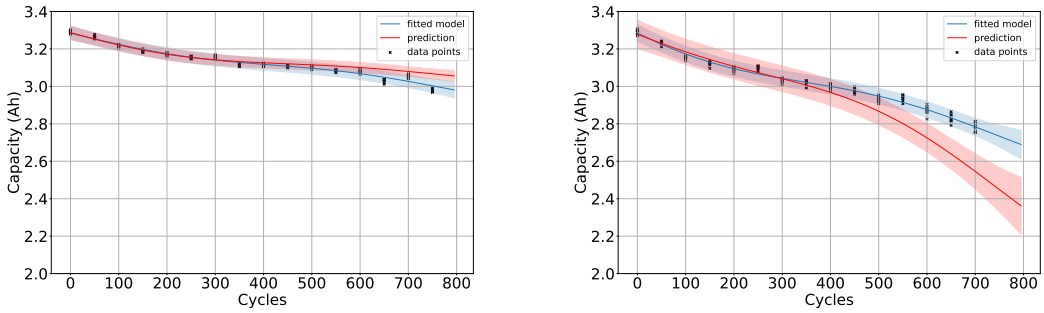


FIGURE 3.17 – Polynomial Fréchet regression predictions

Quantitative comparison between the structured regression and Fréchet regression

To validate more rigorously the predictive abilities of our models, we need to realize a train/test procedure. Regarding our objective, it consists of removing all observations from one of the experimental conditions, making predictions at this position, and then comparing with the true removed data. However, we will be strongly limited in this task because of the few amount of observed temperatures. A model learned on three temperatures will necessarily perform much less than a model with four temperatures, in the first case the maximum temperature gap is of 35°C against 20°C in the complete case. Also, we should not expect this validation to completely reflect the performances reached by a complete model. It will lead to under-estimated performances. However, it still allows us to assess the robustness of the model and provides an upper bound of the prediction error.



(a) (b)

FIGURE 3.18 – Prediction versus true data and associated fitted model at (a) 25 °C and (b) 45 °C.

Fig 3.18a and 3.18b permits visualizing these predictions, comparing the predicted degradation curves for the polynomial Fréchet model (in red), respectively at 25 and 45°C, with the true data and the fitted degradation curves (in blue). We understand that predictions are usually better at 25°C than at 45°C, since degradation curves at 10 and 45°C are closer one to another than

the 25°C and 60°C degradation curves. A similar remark can be made for initial cycles versus last cycles. Performance is usually higher for the first cycles where the degradation curves at the different conditions are similar. But then, predictions become more and more difficult for later cycles. Still, keeping in mind that only three temperatures are used for training the model, the results seem relatively close to the directly fitted models.

	Structured regression	Linear Fréchet	Polynomial Fréchet
25°C	0.091	0.193	0.081
45°C	0.207	0.522	0.207

TABLE 3.1 – MAE performance indicators

	Structured regression	Linear Fréchet	Polynomial Fréchet
25°C	28.745	10.733	19.802
45°C	63.31	50.491	19.223

TABLE 3.2 – NLPD performance indicators

As in the previous chapter, see 2.3.2, performance is quantified thanks to MAE for mean performance and NLPD for uncertainties. Values of the performance indicators, for the three models on both temperatures, are displayed in the tables 3.1 for MAE and 3.2 for NLPD. Regarding the mean, the best performances are obtained for the polynomial Fréchet regression but the results of the structured regression are not so far. On the other hand, linear Fréchet's mean performances are worse. Considering uncertainties with NLPD, the linear Fréchet regression provides surprisingly the best predictions at 25°C, but it is due to an overestimated variance which allows to include the training data inside the confidence intervals. Otherwise, the polynomial Fréchet regression keeps the best performances.

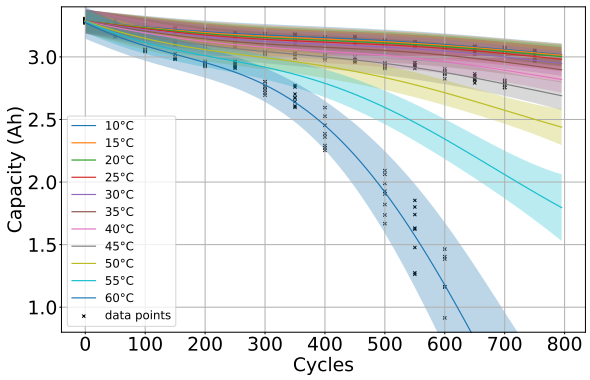
Polynomial Fréchet regression with several experimental factors

Considering now several experimental factors, say two, denoted c_1 and c_2. Another strong constraint of the linear model (3.30), is that it does not model any interaction between the experimental factors. Considering back the analogy with simple real outputs, z, this issue can similarly be handled by extending the inputs with the variable c_1c_2 in the input vector $\mathbf{c}$. This leads to the model

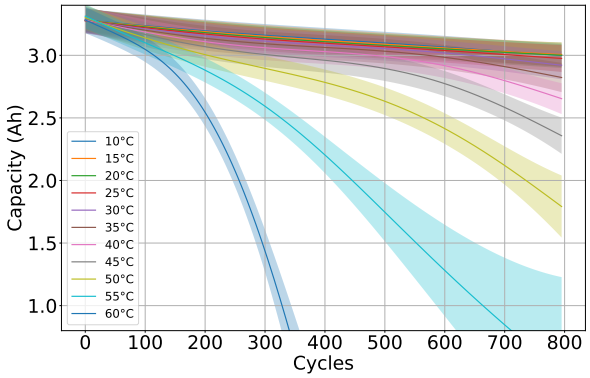
$$z = \beta_0 + \beta_1 c_1 + \beta_2 c_2 + \beta_{1,2} c_1 c_2. \tag{3.35}$$

where $\beta_1 c_1$ and $\beta_2 c_2$ correspond to the direct effects and $\beta_{1,2} c_1 c_2$ to the interaction effect. This interaction can be coupled with a polynomial evolution, considering the following model

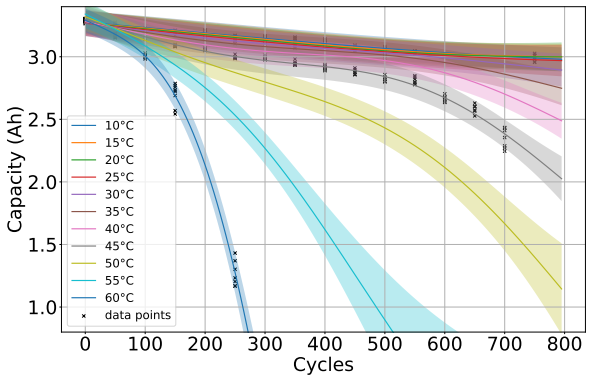
$$z = \beta_0 + \sum_{i}^{q_1} \beta_{1,i} c_1^i + \sum_{i}^{q_2} \beta_{2,i} c_2^i + \sum_{i,j}^{r_1,r_2} \beta_{i,j} c_1^i c_2^j \tag{3.36}$$



(a) CCC of 0.2C



(b) CCC of 0.1125C



(c) CCC of 0.025C

FIGURE 3.19 – Modeling of both the temperature and CCC effect with the polynomial Fréchet regression. The eight time degradations of Fig. 3.1 are interpolated.

As it is also a particular case of linear regression, the Fréchet regression can directly be adapted to include such interactions and polynomial effects. In (3.36), the direct effects are of degree q_1 and q_2, and the interaction between them are of orders r_1 and r_2. In general, there is no reason to have a different order for the direct effects than for the interactions, so we will use, $r_1 = q_1$ and $r_2 = q_2$.

With this model, in addition to the temperature, we can consider the effect of the current charge cut-off, the stopping condition of the charge. Also, the model now uses the eight degradation curves displayed in Fig. 3.3. On tests not displayed here, we first fitted a model without interactions but it led to a low-quality fit. So we used the Fréchet model with polynomial interactions as in (3.36). Since only two values of CCC are observed, we are limited to a linear degradation with this variable (order of 1). For the temperature, the polynomial degradation can still be used. Fig. 3.19 displays the prediction for a range of temperatures at the two observed CCC plus an intermediary unobserved value. The prediction still has the interesting property of interpolating all the observed conditions while respecting monotonicity concerning the factors.

However, in this case, some difficulties arise. Now, it is less clear what polynomial order should be used. Indeed, using comparison with the empirical model, both temperature degradations have a different optimal degree. We had to use a compromise. We must also notice that this task provided less robust results when removing one of the experimental conditions. Finally, the linear evolution with CCC cannot be validated. So we could not perform a quantitative validation in this case.

3.5 Discussion

Prediction of the health degradation curve is not the only possible application of Conditional Wasserstein barycenters. In many settings, a precise model can be built at a particular operating condition, but a direct model is unfeasible. Wasserstein barycenters are not limited to Gaussian distributions, they can also be computed for empirical distributions [199]. So for example the results of Monte Carlo methods predicting the remaining useful life distribution [143, 68] at different conditions could be then interpolated in a second step. Even more generally, replacing the Wasserstein distance with another distance, the method can be applied to many kinds of complex objects, functional data (e.g. OCV curves), correlation matrices...

Coming back to our application, several useful improvements should be considered. Starting with uncertainties, our approach quantifies precisely the uncertainties in the time degradation but a new epistemic uncertainty occurs in the second step, linked to prediction at an unobserved experimental condition. The conditional barycenter is a generalization of the conditional mean, so only of the mean prediction. An uncertainty quantification would require a distribution model over the space of complex objects which is not common, except in some particular cases [207]. Another possibility in practice can consist in adding some regularization [189, 187] as explained in Sec. 3.3.1, to increase uncertainties as predictions are further from observed data.

The Wasserstein distance in itself may face some limitations. As described in (3.6), it has to be computed between distributions on the same spaces. In practice, this imposes that the degradation curves fitted in the first step must have precisely the same time range. If at a condition, predictions can be done only up to cycle 400 whereas for another it can go up to cycle 800, it would result

in two distributions, one on $\mathbb{R}^{d_1}$ and another $\mathbb{R}^{d_2}$. A typical way to handle optimal transport for different dimension distributions is through the Gromov-Wasserstein distance [192], defined in Sec. 3.3.1, which can also be used to compute barycenters [194]. However, this distance only considers the uncertainty structure, with no influence on the mean. The mean being a key element in our case, other approaches should be considered. Another remark on the Wasserstein distance is that it may be limited in its way of handling the mean. In general, with a quadratic cost, the mean of the barycenter is the barycenter of the means. A consequence of this is that each dimension (each time) is handled independently, some information on the curve shape may be lost with such an approach.

Considering more precisely the Fréchet regression, we saw it as a generalization of the linear regression, and similarly, we extended polynomial and over-parametrized linear regression. Using the same argument, any linear smoother method, using a weighted mean of the outputs with weights only function of the input, can be generalized similarly. This includes many different statistical methods such as splines regression or kriging, a limitation being the hyper-parameters selection, but this can be done similarly to the polynomial degree selection. In addition, as an extension of the linear regression, some additional tools were proposed [200] such as adjusted R-square for model selection, which were not of interest to us, since we considered an over-parametrized case, but may be appealing in other cases.

Finally, two difficulties occurred while computing barycenters. In our setting, the degradation curves and their uncertainty are evaluated at an important number of locations. This can lead to singular covariance matrices since uncertainty evolution is limited compared to the evaluation resolution. In that case, the barycenter algorithm (3.21) cannot be used. In practice, this problem can be solved by adding a small penalty on matrices diagonal, but singularity has to be watched to avoid divergence. More theoretically, as explained in the recent work of Haasler and Frossard [208] existence and uniqueness of the barycenter of degenerate Gaussian distributions remains up to date an open problem.

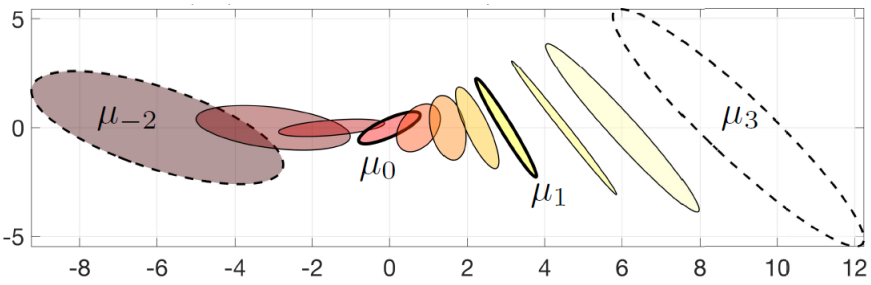


FIGURE 3.20 – Interpolation between two Gaussian distributions and extrapolation, from [14]

The second issue is linked to the sign of barycentric coordinates. In theory, barycenters are only characterized for positive weights [13], while Fréchet's regression can require negative weights, as can be seen (3.32). In practice as in [205], we simply applied the barycenters formula derived in the positive restricted case, which yields satisfactory results. However, there is a lack of a theoretically grounded consideration of barycenters with negative weights, which correspond to the extrapolation of barycentric surface. This problem becomes particularly important if one is interested in the extrapolation of experimental conditions. As illustrated in Fig. 3.20 from [14] McCann's interpolation formula can also be applied in extrapolation, yet geodicity is no longer ensured. It remains an open problem to generalize barycenters to extrapolation in a consistent

way, at least in particular cases, or in a loosed form.

Conclusion and perspectives

This thesis proposed an approach to model the battery health degradation curve with uncertainties, taking into account the influence of the experimental factors. For this, we tackled first this problem at a reference experimental condition, focusing on Gaussian process-based methods considering both time interpolation and forecasting. Then we included the dependence of the experimental factors thanks to the Wasserstein conditional barycenter provided by the optimal transport theory, building a model on the results of the Gaussian process methods.

A core topic in this thesis was uncertainty quantification. We discussed the different sources of uncertainty in experimental data, either aleatoric, linked to the natural variability of the experiment, or epistemic linked to the lack of data. The aleatoric uncertainty can come from many sources, variability of battery manufacturing or in the experimental process, and have different structures, linked to a battery or a batch, symmetric or not, or time-dependent. The epistemic uncertainty on the other hand is particularly important considering forecasting or experimental conditions interpolation, but also in contexts with few tested batteries.

We saw all the interest of Gaussian process-based methods for uncertainty quantification. The kernel design method provides a useful and relatively straightforward way to decompose the different sources of uncertainties. We encountered challenges in modeling time-dependent aleatoric uncertainties due to the stationarity assumption commonly employed in Gaussian process regression. As a response, the Chained Gaussian processes, exploiting the flexibility of variational inference methods, provided many new possibilities with custom likelihood and combinations of several Gaussian processes. It permitted us to model non-parametrically the time evolution of the cell-to-cell variations.

In our goal of modeling experimental factors effects, we saw the interest of optimal transport theory to handle the probabilistic results provided by Gaussian process-methods. This field provides a theoretically sound framework to compare distributions with the Wasserstein distance and to interpolate them through McCann interpolation and barycenters. It was particularly well-suited to the case of Gaussian distributions with closed-form solutions. We however highlighted that important open problems remain in that field concerning the interpolation of degenerated Gaussian distributions and the extrapolation problem.

Our general modeling approach evolved during this thesis. Given our statistical and machine learning background with no prior experience in the battery domain, we started from a purely data-driven approach. However, because of the limited quantity of battery experimental data, we progressively included physical prior knowledge. We saw the interest of including constraints within the Chained Gaussian process model to improve forecasting and we exploited physical

semi-empirical models to take into account the experimental factors.

The mixing of both data-driven and model-based methods is promising. It enables to couple the flexibility of the former with the robustness of the latter, particularly when data is rare. As in many other application fields, physics-informed machine learning methods are gaining popularity for battery applications. Our modeling could probably be improved by a more complete interaction of the two approaches and more precise physic prior knowledge, considering also mechanistic of equivalent circuit approaches. This highlights the necessity of developing multi-disciplinary teams, as was the case in this thesis, to exploit all the available knowledge.

The modeling of experimental factors effects introduced the other important topic of building models on structured objects. Even if complex data embedding meets currently important successes in image or natural language processing, adapting directly the model structure to the data is a promising alternative. In particular, well-known mathematical structures can benefit from existing associated powerful tools as optimal transport for probability distributions. In this context, the conditional barycenter provides an elegant and versatile framework to generalize regression in any metric space. It is a promising way to aggregate several modeling results.

In our opinion, the current state of research is well-developed regarding degradation modeling at a reference experimental condition, even if the forecasting still requires some concerns. Yet, modelings taking into account the experimental factors effect are much less developed, especially considering data-driven approaches. This is mainly due to the current limitations regarding the number of tested experimental conditions.

In our case, we considered a few experimental factors with fixed conditions. Much progress needs to be made to include more experimental factors and more complex conditions. Experimental datasets with time-varying currents are progressively developed, but they are limited to some contexts and there are still few datasets with time-varying temperatures. Thus we are still far from bridging the gap between experimental and on field data.

There is even more progress to be made regarding the modeling of battery characteristics. Most datasets use a single battery type, and the combination of different datasets is not simple since the experimental conditions are also different. The experimental tests in the public dataset are often performed on commercial batteries whose characteristics are not publicly shared which also limits the comparison. Finally, even if characteristics such as the electrode sizes are quantitative, many characteristics such as the chemistry or the design are qualitative, which can raise some modeling difficulties.

Several related issues were mentioned and would need further development. An important concern for Lithium-ion batteries is the safety issue and the prevision of accidents with thermal runaway. The development of individual prediction techniques using BMS data is also of high commercial interest, providing services to optimize battery lifetime and maintenance. This problem is important at different scales, we considered the module in this thesis, but practical interest also relies on battery pack performance or even in BESS containers performance. It requires dedicated methods, using online learning to provide quick updates each time measurements are made.

Appendix

A CC-CV charging profile

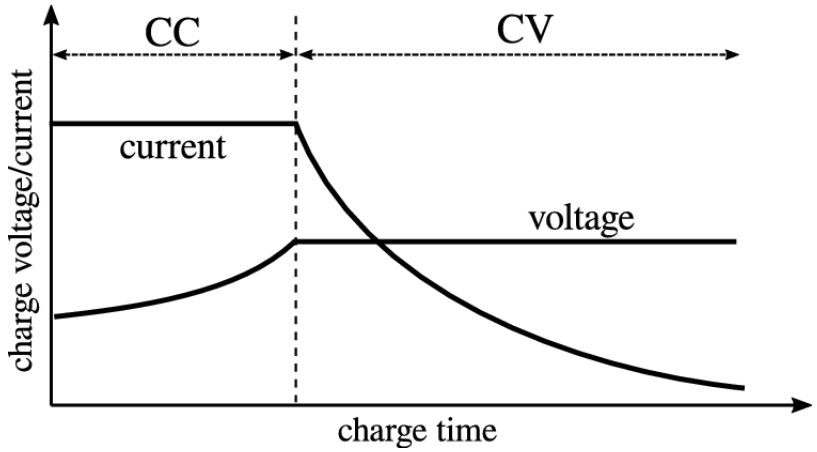


FIGURE 1 – Sketch of a CCCV charge, from [209]

The charging profiles used for the Aachen and Maryland datasets follow a CC-CV policy. This charging profile is widely used, notably for electrical vehicles [210]. The procedure is sketched in Fig. 1. It contains first the Constant Current (CC) phase, where current is imposed and voltage increases progressively up to an upper limit for the voltage. Then in the Constant Voltage (CV) phase, the voltage is maintained to this value. The current decreases until it reaches a current cut-off. A similar policy can be used in discharge with a negative current and decreasing voltage in the CC phase and a current increasing toward 0 in the CV phase.

B Practical implementation details of the Gaussian processes-based modelings

Gaussian process methods and particularly the chained Gaussian process framework, are quiet computationally intensive. In this work we used the Python package GPflow [211]. It is specialized in Gaussian processes methods relying on variational inference. Moreover, relying on

Tensorflow [212] as back-end it is optimized to take advantage of GPU to speed up computations.

Before optimization of the hyperparameters, we normalized the data. It means that each variable, except the battery indicator which is qualitative, is centered and reduced (subtraction of the empirical mean and division by the empirical standard deviation). Log marginal likelihood and ELBO are non-convex functions. Using gradient-based methods, there are no guarantees of convergence to a global minimum. So it is recommended to run optimization with different initializations. So, for each model, we randomly initialized 10 times the hyperparameters. Then we checked the convergence of each hyperparameter thanks to Tensorboard monitoring. Finally, if convergence was observed we selected the model with the lowest final loss function.

Each optimization was done using an AMD EPYC CPU and an NVIDIA A100 GPU on exclusive usage. Since different initializations are independent they can be run trivially in parallel with different bashes. For all models we used the classical L-BFGS-B algorithm [213, 214]. [155] demonstrated that natural gradient optimization outperformed other optimizers to optimize variational hyperparameters. However in our own experience, when optimizing both variational and model hyperparameters L-BFGS-B performed well. It is also easier to use, without the need to specify a policy for the learning rate.

The computation time of the CGP model is much longer. For the qualitative analysis, 800 observations were used and one initialization took around 1 minute for GPR model but almost 1 hour for CGP.

C Overparametrized Fréchet regression

In 3.4.3 we presented the Fréchet regression to model the experimental factor effect. To improve the quality of fit, we proposed an overparametrized version of the Fréchet regression with weights 3.34. Let us explain in more details this proposal.

Starting from the standard linear regression, we consider an output $z \in \mathbb{R}$ depending of inputs $\mathbf{c} \in \mathbb{R}^p$, we consider the equivalent linear regression model

$$z = \beta_0 + \boldsymbol{\beta}_1^T (\mathbf{c} - \bar{\mathbf{c}}) \tag{1}$$

with $\bar{\mathbf{c}}$ the mean vector of each input variables, and denoting p the dimension of $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1)$.

In the standard case ($p \leq n$) we estimate β_0 and $\boldsymbol{\beta}_1$ by minimizing the mean square problem

$$\frac{1}{K} \sum_{k=1}^{K} \left(z_k - (\beta_0 + \boldsymbol{\beta}_1^T (\mathbf{c}_k - \bar{\mathbf{c}}) \right)^2 \tag{2}$$

which provides the solution

$$\hat{\beta}_0 = \bar{z} \tag{3}$$

$$\hat{\boldsymbol{\beta}}_1 = \left(\tilde{C}^T\tilde{C}\right)^{-1}\tilde{C}^T\mathbf{z} \tag{4}$$

with $\tilde{C}$ the input matrix with each of its j columns centered by its mean $\tilde{C} = \left(\bar{\mathbf{c}}_{i,j} - \bar{\mathbf{c}}_j\right)_{i,j}$, and $\mathbf{z} = (z_1, \cdots, z_K)$. Using this estimation a prediction for a new input $\mathbf{c}_*$ corresponds to $z_* = \hat{\beta}_0 + \mathbf{c}_*^T\hat{\boldsymbol{\beta}}_1$ which can be written as a linear combination of the output $z_* = \sum_{k=1}^K \lambda_k(\mathbf{c}_*)z_k$ with

$$\lambda_k(\mathbf{c}_*) = 1 + (\mathbf{c}_k - \bar{\mathbf{c}})^T\hat{\Sigma}^{-1}(\mathbf{c}_* - \bar{\mathbf{c}}), \tag{5}$$

denoting $\hat{\Sigma} = \tilde{C}^T\tilde{C}$, the empirical covariance matrix.

In the over-parametrized case ($p > n$), the mean square problem has an infinity of solutions. Also, a common choice [215] is to choose the one of minimal norm. In fact, since we still want our predictions to be centered, we will keep the constraint $\hat{\beta}_0 = \bar{z}$ and find $\boldsymbol{\beta}_1$ minimizing

$$\frac{1}{K}\sum_{k=1}^K \left(z_k - (\bar{z} + \boldsymbol{\beta}_1^T(\mathbf{c}_k - \bar{\mathbf{c}})\right)^2 \quad \text{s.t.} \quad \|\boldsymbol{\beta}_1\| \text{ is minimized} \tag{6}$$

which leads to the solution

$$\hat{\beta}_0 = \bar{z} \tag{7}$$

$$\hat{\boldsymbol{\beta}}_1 = \left(\tilde{C}^T\tilde{C}\right)^{+}\tilde{C}^T\mathbf{z}, \tag{8}$$

similarly, a prediction for a new input $\mathbf{c}_*$ corresponds to $z_* = \hat{\beta}_0 + \mathbf{c}_*^T\hat{\boldsymbol{\beta}}_1$ which can be written as a linear combination of the output $z_* = \sum_{k=1}^K \lambda_k(\mathbf{c}_*)z_k$ with

$$\lambda_k(\mathbf{c}_*) = 1 + (\mathbf{c}_k - \bar{\mathbf{c}})^T\hat{\Sigma}^{+}(\mathbf{c}_* - \bar{\mathbf{c}}) \tag{9}$$

Bibliography

[1] International Energy Agency. *Global EV Outlook 2023 : Catching up with Climate Ambitions*. en. Global EV Outlook. OECD, avr. 2023. ISBN : 978-92-64-85692-9. DOI : 10.1787/cbe724e8-en. URL : https://www.oecd-ilibrary.org/energy/global-ev-outlook-2023_cbe724e8-en (visité le 24/10/2023).

[2] Christoph R. Birkl et al. "Degradation diagnostics for lithium ion cells". en. In : *Journal of Power Sources* 341 (fév. 2017), p. 373-386. ISSN : 03787753. DOI : 10.1016/j.jpowsour.2016.12.011. URL : https://linkinghub.elsevier.com/retrieve/pii/S0378775316316998 (visité le 17/06/2022).

[3] David Beck et al. "Inhomogeneities and cell-to-cell variations in lithium-ion batteries, a review". In : *Energies* 14.11 (2021), p. 3276. DOI : https://doi.org/10.3390/en14113276.

[4] Xiaosong Hu et al. "Battery Lifetime Prognostics". en. In : *Joule* 4.2 (fév. 2020), p. 310-346. ISSN : 25424351. DOI : 10.1016/j.joule.2019.11.018. URL : https://linkinghub.elsevier.com/retrieve/pii/S2542435119305859 (visité le 16/09/2021).

[5] Carl Edward Rasmussen et Christopher K. I. Williams. *Gaussian processes for machine learning*. MIT press, 2006. URL : http://www.gaussianprocess.org/gpml/chapters/RW.pdf.

[6] Robert R. Richardson, Michael A. Osborne et David A. Howey. "Gaussian process regression for forecasting battery state of health". en. In : *Journal of Power Sources* 357 (juill. 2017), p. 209-219. ISSN : 03787753. DOI : 10.1016/j.jpowsour.2017.05.004. URL : https://linkinghub.elsevier.com/retrieve/pii/S0378775317306250 (visité le 16/09/2021).

[7] Christopher Williams et Carl Rasmussen. "Gaussian Processes for Regression". In : *Advances in Neural Information Processing Systems*. T. 8. MIT Press, 1995. URL : https://proceedings.neurips.cc/paper_files/paper/1995/hash/7cce53cf90577442771720a370c3c723-Abstract.html (visité le 21/09/2023).

[8] Alan D Saul et al. "Chained Gaussian Processes". en. In : (2016), p. 10.

[9] Jaakko Riihimäki. "Advances in Approximate Bayesian Inference for Gaussian Process Models". en. In : (2011), p. 62.

[10] Cédric Villani. *Optimal transport, old and new*. 2008.

[11] Gabriel Peyré et Marco Cuturi. *Computational Optimal Transport*. en. arXiv : 1803.00567. Mars 2020. URL : http://arxiv.org/abs/1803.00567 (visité le 16/09/2021).

[12] Robert J. McCann. "A Convexity Principle for Interacting Gases". In : *Advances in Mathematics* 128.1 (1997), p. 153-179. DOI : https://doi.org/10.1006/aima.1997.1634.

[13] Martial Agueh et Guillaume Carlier. "Barycenters in the Wasserstein space". In : *SIAM Journal on Mathematical Analysis* 43.2 (2011), p. 904-924. doi : https://doi.org/10.1137/100805741.

[14] Boris Muzellec. "Leveraging Regularization, Projections and Elliptical Distributions in Optimal Transport". en. Thèse de doct. 2020.

[15] Alexander Petersen et Hans-Georg Müller. "Fréchet regression for random objects with Euclidean predictors". en. In : *The Annals of Statistics* 47.2 (avr. 2019). issn : 0090-5364. doi : 10.1214/17-AOS1624. url : https://projecteuclid.org/journals/annals-of-statistics/volume-47/issue-2/Fr%c3%a9chet-regression-for-random-objects-with-Euclidean-predictors/10.1214/17-AOS1624.full (visité le 14/11/2022).

[16] Katherine Calvin et al. *IPCC, 2023 : Climate Change 2023 : Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland.* en. Rapp. tech. Edition : First. Intergovernmental Panel on Climate Change (IPCC), juill. 2023. doi : 10.59327/IPCC/AR6-9789291691647. url : https://www.ipcc.ch/report/ar6/syr/ (visité le 24/10/2023).

[17] Ipcc. *Global Warming of 1.5°C : IPCC Special Report on Impacts of Global Warming of 1.5°C above Pre-industrial Levels in Context of Strengthening Response to Climate Change, Sustainable Development, and Efforts to Eradicate Poverty.* en. 1re éd. Cambridge University Press, juin 2022. isbn : 978-1-00-915794-0 978-1-00-915795-7. doi : 10.1017/9781009157940. url : https://www.cambridge.org/core/product/identifier/9781009157940/type/book (visité le 24/10/2023).

[18] Adrian König et al. "An Overview of Parameter and Cost for Battery Electric Vehicles". en. In : *World Electric Vehicle Journal* 12.1 (fév. 2021), p. 21. issn : 2032-6653. doi : 10.3390/wevj12010021. url : https://www.mdpi.com/2032-6653/12/1/21 (visité le 25/10/2023).

[19] *The role of battery energy storage systems in renewable power.* url : https://www.openaccessgovernment.org/battery-energy-storage-systems-renewable-power/162666/.

[20] Bruce Dunn, Haresh Kamath et Jean-Marie Tarascon. "Electrical Energy Storage for the Grid : A Battery of Choices". en. In : *Science* 334.6058 (nov. 2011), p. 928-935. issn : 0036-8075, 1095-9203. doi : 10.1126/science.1212741. url : https://www.science.org/doi/10.1126/science.1212741 (visité le 17/06/2022).

[21] Tianmei Chen et al. "Applications of Lithium-Ion Batteries in Grid-Scale Energy Storage Systems". en. In : *Transactions of Tianjin University* 26.3 (juin 2020), p. 208-217. issn : 1006-4982, 1995-8196. doi : 10.1007/s12209-020-00236-w. url : http://link.springer.com/10.1007/s12209-020-00236-w (visité le 12/07/2023).

[22] K. Zaghib, A. Mauger et C.M. Julien. "Rechargeable lithium batteries for energy storage in smart grids". en. In : *Rechargeable Lithium Batteries.* Elsevier, 2015, p. 319-351. isbn : 978-1-78242-090-3. doi : 10.1016/B978-1-78242-090-3.00012-2. url : https://linkinghub.elsevier.com/retrieve/pii/B9781782420903000122 (visité le 27/06/2022).

[23] *As solar capacity grows, duck curves are getting deeper in California.* https://www.eia.gov/todayinenergy/detail.php?id=56880#.

[24] Nouha Dkhili et al. "A survey of modelling and smart management tools for power grids with prolific distributed generation". en. In : *Sustainable Energy, Grids and Networks* 21 (mars 2020), p. 100284. ISSN : 23524677. DOI : 10.1016/j.segan.2019.100284. URL : https://linkinghub.elsevier.com/retrieve/pii/S235246771930092X (visité le 16/01/2024).

[25] Sai Sudharshan Ravi et Muhammad Aziz. "Utilization of electric vehicles for vehicle-to-grid services : Progress and perspectives". In : *Energies* 15.2 (2022), p. 589.

[26] Pierre Blondel. "Estimation de l'état interne d'une batterie lithium-ion à l'aide d'un modèle électrochimique". fr. Thèse de doct. 2019.

[27] Pedro V.H. Seger, Pierre-Xavier Thivel et Delphine Riu. "A second life Li-ion battery ageing model with uncertainties : From cell to pack analysis". en. In : *Journal of Power Sources* 541 (sept. 2022), p. 231663. ISSN : 03787753. DOI : 10.1016/j.jpowsour.2022.231663. URL : https://linkinghub.elsevier.com/retrieve/pii/S0378775322006619 (visité le 04/01/2024).

[28] Jean-Marie Tarascon. "Développement durable : environnement, énergie et société". In : *L'annuaire du Collège de France. Cours et travaux* 111 (2012), p. 769-792.

[29] Yuanli Ding et al. "Automotive Li-ion batteries : current status and future perspectives". In : *Electrochemical Energy Reviews* 2.1 (2019), p. 1-28. DOI : http://dx.doi.org/10.1007/s41918-018-0022-z.

[30] Mario Marinaro et al. "Bringing forward the development of battery cells for automotive applications : Perspective of R&D activities in China, Japan, the EU and the USA". In : *Journal of Power Sources* 459 (2020), p. 228073. DOI : http://dx.doi.org/10.1016/j.jpowsour.2020.228073.

[31] David Linden et Thomas B Reddy. *Linden's handbook of batteries*. en. OCLC : 904009687. New York, NY : McGraw-Hill, 2011. ISBN : 978-0-07-162421-3 978-1-282-91428-5 978-0-07-162419-0.

[32] MM Thackeray et al. "Lithium insertion into manganese spinels". In : *Materials Research Bulletin* 18.4 (1983), p. 461-472. DOI : http://dx.doi.org/10.1016/0025-5408(83)90138-1.

[33] Ruben Brunetaud et al. "Méthodes non-intrusives d'identification des mécanismes de dégradation des batteries lithium-ion pour la prévision de leur durée de vie". Thèse de doct. Nov. 2022.

[34] John B. Goodenough et Kyu-Sung Park. "The Li-Ion Rechargeable Battery : A Perspective". en. In : *Journal of the American Chemical Society* 135.4 (jan. 2013), p. 1167-1176. ISSN : 0002-7863, 1520-5126. DOI : 10.1021/ja3091438. URL : https://pubs.acs.org/doi/10.1021/ja3091438 (visité le 19/10/2023).

[35] Borah. "On battery materials and methods". In : (2020).

[36] Florian Schipper et Doron Aurbach. "A brief review : Past, present and future of lithium ion batteries". In : *Russian Journal of Electrochemistry* 52.12 (2016), p. 1095-1121. DOI : http://dx.doi.org/10.1134/S1023193516120120.

[37] Yong Guo et al. "Solid-state lithium batteries : Safety and prospects". In : *eScience* 2.2 (2022), p. 138-163. ISSN : 2667-1417. DOI : https://doi.org/10.1016/j.esci.2022.02.008.

[38] Yeru LIANG et al. "A review of rechargeable batteries for portable electronic devices". en. In : *InfoMat* 1.1 (2019). _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1002/inf2.12000, p. 6-32. ISSN : 2567-3165. DOI : 10.1002/inf2.12000. URL : https://onlinelibrary.wiley.com/doi/abs/10.1002/inf2.12000 (visité le 20/10/2023).

[39] Christian SCHLASZA et al. "Review on the aging mechanisms in Li-ion batteries for electric vehicles based on the FMEA method". en. In : *2014 IEEE Transportation Electrification Conference and Expo (ITEC)*. Dearborn, MI : IEEE, juin 2014, p. 1-6. ISBN : 978-1-4799-2262-8. DOI : 10.1109/ITEC.2014.6861811. URL : http://ieeexplore.ieee.org/document/6861811/ (visité le 26/10/2023).

[40] Celina MIKOLAJCZAK et al. *Lithium-ion batteries hazard and use assessment*. Springer Science & Business Media, 2012.

[41] M. BROUSSELY et al. "Main aging mechanisms in Li ion batteries". en. In : *Journal of Power Sources* 146.1-2 (août 2005), p. 90-96. ISSN : 03787753. DOI : 10.1016/j.jpowsour.2005.03.172. URL : https://linkinghub.elsevier.com/retrieve/pii/S0378775305005082 (visité le 26/10/2023).

[42] Kai GOEBEL et al. "Prognostics in Battery Health Management". en. In : *IEEE Instrumentation & Measurement Magazine* 11.4 (août 2008), p. 33-40. ISSN : 1094-6969. DOI : 10.1109/MIM.2008.4579269. URL : http://ieeexplore.ieee.org/document/4579269/ (visité le 16/09/2021).

[43] Gonçalo dos REIS et al. "Lithium-ion battery data and where to find it". en. In : *Energy and AI* 5 (sept. 2021), p. 100081. ISSN : 26665468. DOI : 10.1016/j.egyai.2021.100081. URL : https://linkinghub.elsevier.com/retrieve/pii/S2666546821000355 (visité le 21/09/2021).

[44] Antti AITIO et David A. HOWEY. "Predicting battery end of life from solar off-grid system field data using machine learning". en. In : *Joule* 5.12 (déc. 2021), p. 3204-3220. ISSN : 25424351. DOI : 10.1016/j.joule.2021.11.006. URL : https://linkinghub.elsevier.com/retrieve/pii/S2542435121005328 (visité le 09/01/2024).

[45] Weihan LI. *battery degradation trajectory prediction*. https://git.rwth-aachen.de/isea/battery-degradation-trajectory-prediction/-/tree/master. 2021.

[46] Shahid A HASIB et al. "A Comprehensive Review of Available Battery Datasets, RUL Prediction Approaches, and Advanced Battery Management". en. In : 9 (2021).

[47] Thorsten BAUMHÖFER et al. "Production caused variation in capacity aging trend and correlation to initial cell performance". In : *Journal of Power Sources* 247 (2014), p. 332-338. DOI : https://doi.org/10.1016/j.jpowsour.2013.08.108.

[48] Weihan LI et al. "One-shot battery degradation trajectory prediction with deep learning". In : *Journal of Power Sources* (2021), p. 230024. DOI : http://dx.doi.org/10.1016/j.jpowsour.2021.230024.

[49] Weihan LI et al. "Forecasting battery capacity and power degradation with multi-task learning". en. In : *Energy Storage Materials* 53 (déc. 2022), p. 453-466. ISSN : 24058297. DOI : 10.1016/j.ensm.2022.09.013. URL : https://linkinghub.elsevier.com/retrieve/pii/S2405829722004998 (visité le 08/01/2024).

[50] Kristen A SEVERSON et al. "Data-driven prediction of battery cycle life before capacity degradation". In : *Nature Energy* 4.5 (2019), p. 383-391. DOI : http://dx.doi.org/10.1038/s41560-019-0356-8.

[51] Nicolas Courty et al. "Joint distribution optimal transportation for domain adaptation". In : *Advances in neural information processing systems* 30 (2017). DOI : https://proceedings.neurips.cc/paper/2017/hash/0070d23b06b1486a538c0eaa45dd167a-Abstract.html.

[52] Eyke Hüllermeier et Willem Waegeman. "Aleatoric and Epistemic Uncertainty in Machine Learning : An Introduction to Concepts and Methods". en. In : *Machine Learning* 110.3 (mars 2021). arXiv :1910.09457 [cs, stat], p. 457-506. ISSN : 0885-6125, 1573-0565. DOI : 10.1007/s10994-021-05946-3. URL : http://arxiv.org/abs/1910.09457 (visité le 10/04/2023).

[53] Stephen C Hora. "Aleatory and epistemic uncertainty in probability elicitation with an example from hazardous waste management". In : *Reliability Engineering & System Safety* 54.2-3 (1996), p. 217-223. DOI : http://dx.doi.org/10.1016/S0951-8320(96)00077-4.

[54] Eyke Hüllermeier et Willem Waegeman. "Aleatoric and epistemic uncertainty in machine learning : An introduction to concepts and methods". In : *Machine Learning* 110 (2021), p. 457-506.

[55] Bhaskar Saha et Kai Goebel. "Uncertainty Management for Diagnostics and Prognostics of Batteries using Bayesian Techniques". en. In : *2008 IEEE Aerospace Conference*. ISSN : 1095-323X. Big Sky, MT, USA : IEEE, mars 2008, p. 1-8. ISBN : 978-1-4244-1487-1 978-1-4244-1488-8. DOI : 10.1109/AERO.2008.4526631. URL : http://ieeexplore.ieee.org/document/4526631/ (visité le 11/01/2024).

[56] Xiaosong Hu et al. "Battery Lifetime Prognostics". In : *Joule* 4.2 (2020), p. 310-346. ISSN : 2542-4351. DOI : https://doi.org/10.1016/j.joule.2019.11.018.

[57] Rudolph Emil Kalman. "A new approach to linear filtering and prediction problems". In : (1960).

[58] M Sanjeev Arulampalam et al. "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking". en. In : *IEEE TRANSACTIONS ON SIGNAL PROCESSING* 50.2 (2002).

[59] A Pradeep Lall, B Hao Zhang et C Rahul Lall. "PHM of state-of-charge for flexible power sources in wearable electronics with EKF". In : *2018 IEEE International Reliability Physics Symposium (IRPS)*. IEEE. 2018, P-SR.

[60] Eric Walker, Sean Rayman et Ralph E White. "Comparison of a particle filter and other state estimation methods for prognostics of lithium-ion batteries". In : *Journal of Power Sources* 287 (2015), p. 1-12.

[61] Limin Li, Zhongsheng Wang et Hongkai Jiang. "Storage battery remaining useful life prognosis using improved unscented particle filter". In : *Proceedings of the Institution of Mechanical Engineers, Part O : Journal of Risk and Reliability* 229.1 (2015), p. 52-61.

[62] Wei He et al. "A physics-based electrochemical model for lithium-ion battery state-of-charge estimation solved by an optimised projection-based method and moving-window filtering". In : *Energies* 11.8 (2018), p. 2120.

[63] P Ramadass et al. "Development of first principles capacity fade model for Li-ion cells". In : *Journal of the Electrochemical Society* 151.2 (2004), A196. DOI : http://dx.doi.org/10.1149/1.1634273.

[64] Venkatasailanathan Ramadesigan et al. "Parameter estimation and capacity fade analysis of lithium-ion batteries using reformulated models". In : *Journal of the electrochemical society* 158.9 (2011), A1048. DOI : http://dx.doi.org/10.1149/1.3609926.

[65] F Brosa Planella et al. "A continuum of physics-based lithium-ion battery models reviewed". en. In : *Progress in Energy* 4.4 (oct. 2022), p. 042003. issn : 2516-1083. doi : 10.1088/2516-1083/ac7d31. url : https://iopscience.iop.org/article/10.1088/2516-1083/ac7d31 (visité le 13/07/2023).

[66] Marc Doyle, Thomas F Fuller et John Newman. "Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell". In : *Journal of the Electrochemical society* 140.6 (1993), p. 1526.

[67] Uwe Westerhoff et al. "Analysis of Lithium-Ion Battery Models Based on Electrochemical Impedance Spectroscopy". en. In : *Energy Technology* 4.12 (2016). _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1002/ente.201600154, p. 1620-1630. issn : 2194-4296. doi : 10.1002/ente.201600154. url : https://onlinelibrary.wiley.com/doi/abs/10.1002/ente.201600154 (visité le 20/10/2023).

[68] Bhaskar Saha et al. "An integrated approach to battery health monitoring using Bayesian regression and state estimation". In : *2007 IEEE autotestcon*. Ieee. 2007, p. 646-653. doi : http://dx.doi.org/10.1109/AUTEST.2007.4374280.

[69] Jie Liu, Wilson Wang et Fai Ma. "A regularized auxiliary particle filtering approach for system state estimation and battery life prediction". In : *Smart Materials and Structures* 20.7 (2011), p. 075021. doi : http://dx.doi.org/10.1088/0964-1726/20/7/075021.

[70] Dong Wang, Qiang Miao et Michael Pecht. "Prognostics of lithium-ion batteries based on relevance vectors and a conditional three-parameter capacity degradation model". In : *Journal of Power Sources* 239 (2013), p. 253-264.

[71] Pham Luu Trung Duong et Nagarajan Raghavan. "Heuristic Kalman optimized particle filter for remaining useful life prediction of lithium-ion battery". In : *Microelectronics Reliability* 81 (2018), p. 232-243.

[72] Weiping Diao, Saurabh Saxena et Michael Pecht. "Accelerated cycle life testing and capacity degradation modeling of LiCoO2-graphite cells". In : *Journal of Power Sources* 435 (2019), p. 226830. doi : http://dx.doi.org/10.1016/j.jpowsour.2019.226830.

[73] Marcus Johnen et al. "Modeling long-term capacity degradation of lithium-ion batteries". In : *Journal of Energy Storage* 34 (2021), p. 102011. doi : https://doi.org/10.1016/j.est.2020.102011.

[74] Thomas Waldmann et al. "Temperature dependent ageing mechanisms in Lithium-ion batteries–A Post-Mortem study". In : *Journal of power sources* 262 (2014), p. 129-135.

[75] Selina SY Ng, Yinjiao Xing et Kwok L Tsui. "A naive Bayes model for robust remaining useful life prediction of lithium-ion battery". In : *Applied Energy* 118 (2014), p. 114-123.

[76] Adnan Nuhic et al. "Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods". In : *Journal of power sources* 239 (2013), p. 680-688. doi : http://dx.doi.org/10.1016/j.jpowsour.2012.11.146.

[77] Shuai Wang et al. "Prognostics of lithium-ion batteries based on battery performance analysis and flexible support vector regression". In : *Energies* 7.10 (2014), p. 6492-6508. doi : http://dx.doi.org/10.3390/en7106492.

[78] Wen-An Yang et al. "A hybrid prognostic approach for remaining useful life prediction of lithium-ion batteries". In : *Shock and Vibration* 2016 (2016).

[79] Dong Wang, Qiang Miao et Michael Pecht. "Prognostics of lithium-ion batteries based on relevance vectors and a conditional three-parameter capacity degradation model". In : *Journal of Power Sources* 239 (2013), p. 253-264. issn : 0378-7753. doi : https://doi.org/10.1016/j.jpowsour.2013.03.129.

[80] Robert R Richardson, Michael A Osborne et David A Howey. "Gaussian process regression for forecasting battery state of health". In : *Journal of Power Sources* 357 (2017), p. 209-219. doi : http://dx.doi.org/10.1016/j.jpowsour.2017.05.004.

[81] Datong Liu et al. "Prognostics for state of health estimation of lithium-ion batteries based on combination Gaussian process functional regression". In : *Microelectronics Reliability* 53.6 (2013), p. 832-839. doi : http://dx.doi.org/10.1016/j.microrel.2013.03.010.

[82] Thirumalai Parthiban, R Ravi et N Kalaiselvi. "Exploration of artificial neural network [ANN] to predict the electrochemical characteristics of lithium-ion cells". In : *Electrochimica Acta* 53.4 (2007), p. 1877-1882. doi : http://dx.doi.org/10.1016/j.electacta.2007.08.049.

[83] Mohammad Rezvani et al. "A comparative analysis of techniques for electric vehicle battery prognostics and health management (PHM)". In : *SAE Technical Paper* 191 (2011), p. 1-9. doi : http://dx.doi.org/10.4271/2011-01-2247.

[84] Akram Eddahech et al. "Behavior and state-of-health monitoring of Li-ion batteries using impedance spectroscopy and recurrent neural networks". In : *International Journal of Electrical Power & Energy Systems* 42.1 (2012), p. 487-494.

[85] Weihan Li et al. "One-shot battery degradation trajectory prediction with deep learning". en. In : *Journal of Power Sources* 506 (sept. 2021), p. 230024. issn : 0378-7753. doi : 10.1016/j.jpowsour.2021.230024. url : https://www.sciencedirect.com/science/article/pii/S0378775321005528 (visité le 28/09/2021).

[86] Bhaskar Saha, Kai Goebel et Jon Christophersen. "Comparison of prognostic algorithms for estimating remaining useful life of batteries". In : *Transactions of the Institute of Measurement and Control* 31.3-4 (juin 2009). Publisher : SAGE Publications Ltd STM, p. 293-308. issn : 0142-3312. doi : 10.1177/0142331208092030. url : https://doi.org/10.1177/0142331208092030 (visité le 12/01/2024).

[87] Nabil Laayouj et Hicham Jamouli. "Lithium-ion battery degradation assessment and remaining useful life estimation in hybrid electric vehicle". In : *Renewable Energy and Sustainable Development* 2.1 (2016), p. 37-44. doi : http://dx.doi.org/10.21622/RESD.2016.02.1.037.

[88] Weijun Gu et al. "A new method of accelerated life testing based on the Grey System Theory for a model-based lithium-ion battery life evaluation system". In : *Journal of Power Sources* 267 (2014), p. 366-379. doi : http://dx.doi.org/10.1016/j.jpowsour.2014.05.103.

[89] Dong Zhou et al. "On-line remaining useful life prediction of lithium-ion batteries based on the optimized gray model GM (1, 1)". In : *batteries* 3.3 (2017), p. 21. doi : http://dx.doi.org/10.3390/batteries3030021.

[90] Dong Wang et al. "Nonlinear-drifted Brownian motion with multiple hidden states for remaining useful life prediction of rechargeable batteries". In : *Mechanical Systems and Signal Processing* 93 (2017), p. 531-544.

[91] Jing Feng, Paul Kvam et Yanzhen Tang. "Remaining useful lifetime prediction based on the damage-marker bivariate degradation model : A case study on lithium-ion batteries used in electric vehicles". In : *Engineering Failure Analysis* 70 (2016), p. 323-342.

[92] Katharina Rumpf et al. "Influence of Cell-to-Cell Variations on the Inhomogeneity of Lithium-Ion Battery Modules". en. In : *Journal of The Electrochemical Society* 165.11 (2018), A2587-A2607. issn : 0013-4651, 1945-7111. doi : 10.1149/2.0111811jes. url : https://iopscience.iop.org/article/10.1149/2.0111811jes (visité le 04/01/2023).

[93] Stephen J Harris, David J Harris et Chen Li. "Failure statistics for commercial lithium ion batteries : A study of 24 pouch cells". In : *Journal of Power Sources* 342 (2017), p. 589-597. doi : https://doi.org/10.1016/j.jpowsour.2016.12.083.

[94] Yuliya Preger et al. "Degradation of commercial lithium-ion cells as a function of chemistry and cycling conditions". In : *Journal of The Electrochemical Society* 167.12 (2020), p. 120532. doi : http://dx.doi.org/10.1149/1945-7111/abae37.

[95] Alan D Saul et al. "Chained gaussian processes". In : *Artificial Intelligence and Statistics*. PMLR. 2016, p. 1431-1440. url : https://proceedings.mlr.press/v51/saul16.html.

[96] N a Wiener. *Extrapolation, Interpolation and Smoothing of Stationary Time Series.* MIT Press,Cambridge, Mass., 1949.

[97] Andrey Kolmogoroff. "Interpolation und extrapolation von stationaren zufalligen folgen". In : *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya* 5.1 (1941), p. 3-14.

[98] Daniel G Krige. "A statistical approach to some basic mine valuation problems on the Witwatersrand". In : *Journal of the Southern African Institute of Mining and Metallurgy* 52.6 (1951), p. 119-139.

[99] B. Matern. *Spatial Variation—Stochastic Models and Their Application to Some Problems in Forest Surveys and Other Sampling Investigations. Meddelanden från Statens Skogsforskningsinstitut.* 49, No.5. Alm¨anna F¨orlaget, Stockholm. Second edition (1986), Springer-Verlag, Berlin.., 1960.

[100] L.S Gandin. *Ob"ektivnyi analiz meteorologicheskikh polei. Gidrometeologicheskoe Izdatel'stvo, Leningrad. Translation (1965) : Objective Analysis of Meteorological Fields. Israel Program for Scientific Translations, Jerusalem.* Objective Analysis of Meteorological fields. Israel Program for Scientific Translations, Jerusalem., 1963.

[101] G Matheron. *La the´orie des variables re´gionalise´es et ses applications. Cahiers du Centre de Morphologie Mathe´matique de Fontainebleau, Fasc. 5, Ecole des Mines de Paris. Translation (1971) : The Theory of Regionalized Variables and Its Applications.* John Wiley & Sons, New York.Reprinted (1993), 1970.

[102] N. Cressie. *Statistics for Spatial Data.* John Wiley & Sons, New York.Reprinted (1993), 1991.

[103] Jean-Paul Chiles et Pierre Delfiner. *Geostatistics : modeling spatial uncertainty.* T. 497. John Wiley & Sons, 2009.

[104] Christopher Williams et Carl Rasmussen. "Gaussian processes for regression". In : *Advances in neural information processing systems* 8 (1995).

[105] L.S Gandin. *Bayesian Learning for Neural Networks.* Springer, New York. Lecture Notes in Statistics 118, 1996.

[106] Anthony O'Hagan. "Curve fitting and optimal design for prediction". In : *Journal of the Royal Statistical Society : Series B (Methodological)* 40.1 (1996), p. 1-24.

[107] Christopher KI Williams et David Barber. "Bayesian classification with Gaussian processes". In : *IEEE Transactions on pattern analysis and machine intelligence* 20.12 (1998), p. 1342-1351.

[108] Radford M Neal. "Monte Carlo implementation of Gaussian process models for Bayesian regression and classification". In : *arXiv preprint physics/9701026* (1997).

[109] Peter J Diggle, Jonathan A Tawn et Rana A Moyeed. "Model-based geostatistics". In : *Journal of the Royal Statistical Society Series C : Applied Statistics* 47.3 (1998), p. 299-350.

[110] Haitao Liu et al. "When Gaussian Process Meets Big Data : A Review of Scalable GPs". en. In : *arXiv :1807.01065 [cs, stat]* (avr. 2019). arXiv : 1807.01065. URL : http://arxiv.org/abs/1807.01065 (visité le 16/09/2021).

[111] Andreas Damianou et Neil D Lawrence. "Deep gaussian processes". In : *Artificial intelligence and statistics*. PMLR. 2013, p. 207-215.

[112] Jaakko Riihimäki et Aki Vehtari. "Gaussian processes with monotonicity information". In : *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop et Conference Proceedings. 2010, p. 645-652.

[113] Duvenaud. "Automatic model construction with Gaussian processes". Thèse de doct. 2014. URL : https://core.ac.uk/reader/77407350 (visité le 15/09/2023).

[114] David Duvenaud. "The Kernel Cookbook : Advice on Covariance functions (2014)". In : *URL https ://www. cs. toronto. edu/˜ duvenaud/cookbook* ().

[115] Trevor Hastie et Tibshirani R. "Generalized Additive Models". In : *Monographs on Statistics and Applied Probability* 43 (1990). DOI : http://dx.doi.org/10.1214/ss/1177013604.

[116] A Gelman et al. *Bayesian data analysis third edition*. Chapman et Hall/CRC, 2013.

[117] James Hensman, Nicolò Fusi et Neil D. Lawrence. "Gaussian Processes for Big Data". In : *Uncertainty in Artificial Intelligence*. Sous la dir. d'Ann Nicholson et Padhraic Smyth. T. 29. AUAI Press, 2013.

[118] Håvard Rue, Sara Martino et Nicolas Chopin. "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations". en. In : *Journal of the Royal Statistical Society : Series B (Statistical Methodology)* 71.2 (2009). _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2008.00700.x, p. 319-392. ISSN : 1467-9868. DOI : 10.1111/j.1467-9868.2008.00700.x. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2008.00700.x (visité le 13/11/2023).

[119] James Hensman et Zoubin Ghahramani. "Scalable Variational Gaussian Process Classification". en. In : (2016), p. 10.

[120] James Hensman, Nicolo Fusi et Neil D Lawrence. "Gaussian Processes for Big Data". en. In : (2013), p. 9.

[121] Andreas Damianou et Neil D. Lawrence. "Deep Gaussian Processes". en. In : *Artificial Intelligence and Statistics*. ISSN : 1938-7228. PMLR, avr. 2013, p. 207-215. URL : https://proceedings.mlr.press/v31/damianou13a.html (visité le 17/09/2021).

[122] Carl Doersch. *Tutorial on Variational Autoencoders*. en. arXiv :1606.05908 [cs, stat]. Jan. 2021. URL : http://arxiv.org/abs/1606.05908 (visité le 26/09/2022).

[123] David M. Blei, Alp Kucukelbir et Jon D. McAuliffe. "Variational Inference : A Review for Statisticians". In : *Journal of the American Statistical Association* 112.518 (avr. 2017), p. 859-877. ISSN : 1537-274X. DOI : 10.1080/01621459.2017.1285773.

[124] Christopher M. Bishop. *Patern recognition and machine learning*. Springer New York, 2006.

[125] Matthew D Hoffman et al. "Stochastic variational inference." In : *Journal of Machine Learning Research* 14.5 (2013).

[126] Hugh Salimbeni, Stefanos Eleftheriadis et James Hensman. "Natural Gradients in Practice : Non-Conjugate Variational Inference in Gaussian Process Models". en. In : (2018), p. 9.

[127] Datong Liu et al. "Prognostics for state of health estimation of lithium-ion batteries based on combination Gaussian process functional regression". en. In : *Microelectronics Reliability* 53.6 (juin 2013), p. 832-839. issn : 0026-2714. doi : 10.1016/j.microrel.2013.03.010. url : https://www.sciencedirect.com/science/article/pii/S0026271413000747 (visité le 28/09/2021).

[128] Lingling Li et al. "Remaining useful life prediction for lithium-ion batteries based on Gaussian processes mixture". In : *PloS one* 11.9 (2016), e0163004. doi : http://dx.doi.org/10.1371/journal.pone.0163004.

[129] M. Myllymäki, A. Särkkä et Aki Vehtari. "Hierarchical second-order analysis of replicated spatial point patterns with non-spatial covariates". In : *spatial statistics* 8 (2014), p. 104-121. doi : http://dx.doi.org/10.1016/j.spasta.2013.07.006.

[130] Christopher Paciorek et Mark Schervish. "Nonstationary covariance functions for Gaussian process regression". In : *Advances in neural information processing systems* 16 (2003).

[131] Antti Aitio et David A Howey. "Predicting battery end of life from solar off-grid system field data using machine learning". In : *Joule* 5.12 (2021), p. 3204-3220. doi : http://dx.doi.org/10.1016/j.joule.2021.11.006.

[132] Ryan Prescott Adams et Oliver Stegle. "Gaussian process product models for nonparametric nonstationarity". In : *Proceedings of the 25th international conference on Machine learning*. 2008, p. 1-8. doi : http://dx.doi.org/10.1145/1390156.1390157.

[133] John P Klein et Melvin L Moeschberger. *Survival analysis : techniques for censored and truncated data*. T. 2. Springer, 2003. doi : http://dx.doi.org/10.1007/b97377.

[134] Joaquin Quinonero-Candela et al. "Evaluating predictive uncertainty challenge". In : *Machine Learning Challenges Workshop*. Springer. 2005, p. 1-27. doi : http://dx.doi.org/10.1007/11736790_1.

[135] Philipp Dechent et al. "Estimation of Li-Ion Degradation Test Sample Sizes Required to Understand Cell-to-Cell Variability". In : *Batteries & Supercaps* 4.12 (2021), p. 1821-1829.

[136] Egoitz Martinez-Laserna et al. "Technical viability of battery second life : A study from the ageing perspective". In : *IEEE Transactions on Industry Applications* 54.3 (2018), p. 2703-2713. doi : https://doi.org/10.1109/TIA.2018.2801262.

[137] M. Raissi, P. Perdikaris et G.E. Karniadakis. "Physics-informed neural networks : A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations". In : *Journal of Computational Physics* 378 (2019), p. 686-707. issn : 0021-9991. doi : https://doi.org/10.1016/j.jcp.2018.10.045.

[138] George Em Karniadakis et al. "Physics-informed machine learning". In : *Nature Reviews Physics* 3.6 (2021), p. 422-440. doi : http://dx.doi.org/10.1038/s42254-021-00314-5.

[139] Wendi Guo et al. "Review of "grey box" lifetime modeling for lithium-ion battery : Combining physics and data-driven methods". In : *Journal of Energy Storage* 56 (2022), p. 105992.

[140] Guangxing Bai et al. "Prognostics of Lithium-Ion batteries using knowledge-constrained machine learning and Kalman filtering". In : *Reliability Engineering & System Safety* 231 (2023), p. 108944. ISSN : 0951-8320. DOI : https://doi.org/10.1016/j.ress.2022.108944.

[141] Peter M Attia et al. ""Knees" in lithium-ion battery aging trajectories". In : *Journal of The Electrochemical Society* 169.6 (2022), p. 060517. DOI : https://doi.org/10.1149/1945-7111/ac6d13.

[142] Jeremy Neubauer et Ahmad Pesaran. "The ability of battery second use strategies to impact plug-in electric vehicle prices and serve utility energy storage applications". In : *Journal of Power Sources* 196.23 (2011), p. 10351-10358. DOI : http://dx.doi.org/10.1016/j.jpowsour.2011.06.053.

[143] Wei He et al. "Prognostics of lithium-ion batteries based on Dempster–Shafer theory and the Bayesian Monte Carlo method". In : *Journal of Power Sources* 196.23 (2011), p. 10314-10321. DOI : http://dx.doi.org/10.1016/j.jpowsour.2011.08.040.

[144] Xuebing Han et al. "Cycle life of commercial lithium-ion batteries with lithium titanium oxide anodes in electric vehicles". In : *Energies* 7.8 (2014), p. 4895-4909. DOI : http://dx.doi.org/10.3390/en7084895.

[145] "Evaluation of commercial lithium-ion cells based on composite positive electrode for plug-in hybrid electric vehicle applications. Part II. Degradation mechanism under 2C cycle aging". In : *Journal of Power Sources* 196.23 (2011), p. 10336-10343. DOI : https://doi.org/10.1016/j.jpowsour.2011.08.078.

[146] Weiping Diao et al. "Algorithm to determine the knee point on capacity fade curves of lithium-ion cells". In : *Energies* 12.15 (2019), p. 2910. DOI : http://dx.doi.org/10.3390/en12152910.

[147] Fangfang Yang et al. "A coulombic efficiency-based model for prognostics and health estimation of lithium-ion batteries". In : *Energy* 171 (2019), p. 1173-1182. DOI : https://doi.org/10.1016/j.energy.2019.01.083.

[148] Sara Kohtz et Pingfeng Wang. "Capacity Degradation Modeling for Li-Ion Batteries using a Multiscale Gamma Process Approach". In : *2021 Annual Reliability and Maintainability Symposium (RAMS)*. IEEE. 2021, p. 1-6. DOI : https://doi.org/10.1109/RAMS48097.2021.9605712.

[149] Salman Jahani et al. "Remaining useful life prediction based on degradation signals using monotonic B-splines with infinite support". In : *IISE Transactions* 52.5 (2020), p. 537-554. DOI : https://doi.org/10.1080/24725854.2019.1630868.

[150] Datong Liu et al. "Satellite lithium-ion battery remaining cycle life prediction with novel indirect health indicator extraction". In : *Energies* 6.8 (2013), p. 3654-3668. DOI : https://doi.org/10.3390/en6083654.

[151] Xiu Yang et al. "Physics-informed CoKriging : A Gaussian-process-regression-based multifidelity method for data-model convergence". In : *Journal of Computational Physics* 395 (2019), p. 410-431. DOI : http://dx.doi.org/10.1016/j.jcp.2019.06.041.

[152] Laura P Swiler et al. "A survey of constrained Gaussian process regression : Approaches and implementation challenges". In : *Journal of Machine Learning for Modeling and Computing* 1.2 (2020). DOI : http://dx.doi.org/10.1615/JMachLearnModelComput.2020035155.

[153] Christian AGRELL. "Gaussian Processes with Linear Operator Inequality Constraints". In : *Journal of Machine Learning Research* 20.135 (2019), p. 1-36. URL : http://jmlr.org/papers/v20/19-065.html.

[154] Jouko LAMPINEN. "Gaussian Processes with Monotonicity Constraint for Big Data". In : (2014).

[155] Hugh SALIMBENI, Stefanos ELEFTHERIADIS et James HENSMAN. "Natural gradients in practice : Non-conjugate variational inference in Gaussian process models". In : *International Conference on Artificial Intelligence and Statistics*. PMLR. 2018, p. 689-697.

[156] James HENSMAN, Nicolas DURRANDE, Arno SOLIN et al. "Variational Fourier Features for Gaussian Processes." In : *J. Mach. Learn. Res.* 18.1 (2017), p. 5537-5588.

[157] Fangfang YANG et al. "Prognostics of Li(NiMnCo)O2-based lithium-ion batteries using a novel battery degradation model". In : *Microelectronics Reliability* 70 (2017), p. 70-78. DOI : https://doi.org/10.1016/j.microrel.2017.02.002.

[158] C BIRKL. "Diagnosis and prognosis of degradation in lithium-ion batteries". Thèse de doct. University of Oxford, 2017.

[159] Arpit MAHESHWARI et al. "Optimizing the operation of energy storage using a non-linear lithium-ion battery degradation model". In : *Applied Energy* 261 (2020), p. 114360. DOI : http://dx.doi.org/10.1016/j.apenergy.2019.114360.

[160] Bolun XU et al. "Modeling of lithium-ion battery degradation for cell life assessment". In : *IEEE Transactions on Smart Grid* 9.2 (2016), p. 1131-1140. DOI : http://dx.doi.org/10.1109/TSG.2016.2578950.

[161] Jorn M RENIERS, Grietus MULDER et David A HOWEY. "Review and performance comparison of mechanical-chemical degradation models for lithium-ion batteries". In : *Journal of The Electrochemical Society* 166.14 (2019), A3189-A3200. DOI : http://dx.doi.org/10.1149/2.0281914jes.

[162] M LUCU et al. "Data-driven nonparametric Li-ion battery ageing model aiming at learning from real operation data-Part B : Cycling operation". In : *Journal of Energy Storage* 30 (2020), p. 101410. DOI : http://dx.doi.org/10.1016/j.est.2020.101410.

[163] Kailong LIU et al. "Gaussian process regression with automatic relevance determination kernel for calendar aging prediction of lithium-ion batteries". In : *IEEE Transactions on Industrial Informatics* 16.6 (2019), p. 3767-3777. DOI : http://dx.doi.org/10.1109/TII.2019.2941747.

[164] Cédric VILLANI. *Optimal transport : old and new*. T. 338. Springer, 2009. URL : https://cedricvillani.org/sites/dev/files/old_images/2012/08/preprint-1.pdf.

[165] Gabriel PEYRÉ, Marco CUTURI et al. "Computational optimal transport : With applications to data science". In : *Foundations and Trends® in Machine Learning* 11.5-6 (2019), p. 355-607. DOI : http://dx.doi.org/10.1561/2200000073.

[166] Haitao LIU et al. "When Gaussian process meets big data : A review of scalable GPs". In : *IEEE transactions on neural networks and learning systems* 31.11 (2020), p. 4405-4423.

[167] Siwei LAI et al. "How to Generate a Good Word Embedding". In : *IEEE Intelligent Systems* 31.6 (2016), p. 5-14. DOI : 10.1109/MIS.2016.45.

[168] HongYun CAI, Vincent W. ZHENG et Kevin Chen-Chuan CHANG. "A Comprehensive Survey of Graph Embedding : Problems, Techniques, and Applications". In : *IEEE Transactions on Knowledge and Data Engineering* 30.9 (2018), p. 1616-1637. DOI : 10.1109/TKDE.2018.2807452.

[169] Maurice Fréchet. "Les éléments aléatoires de nature quelconque dans un espace distancié". fr. In : (1948), p. 97.

[170] P Thomas Fletcher. "Geodesic regression and the theory of least squares on Riemannian manifolds". In : *International journal of computer vision* 105 (2013), p. 171-185. doi : https://doi.org/10.1007/s11263-012-0591-y.

[171] Jacob Hinkle et al. "Polynomial regression on Riemannian manifolds". In : *Computer Vision–ECCV 2012 : 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part III 12*. Springer. 2012, p. 1-14. doi : https://doi.org/10.1007/978-3-642-33712-3_1.

[172] Monami Banerjee et al. "Nonlinear regression on Riemannian manifolds and its applications to Neuro-image analysis". In : *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015 : 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part I 18*. Springer. 2015, p. 719-727. doi : https://doi.org/10.1007/978-3-319-24553-9_88.

[173] Gaspard Monge. "Mémoire sur la théorie des déblais et des remblais". In : *Histoire de l'Académie Royale des Sciences* (1781), p. 666-704.

[174] Leonid Kantorovitch. "On the transfer of masses (in russian)." In : *Doklady Akademii Nauk, 37(2)* (1942), p. 227-229.

[175] George B Dantzig. "Programming of interdependent activities : II mathematical model". In : *Econometrica, Journal of the Econometric Society* (1949), p. 200-211.

[176] Yann Brenier. "Décomposition polaire et réarrangement monotone des champs de vecteurs". In : *CR Acad. Sci. Paris Sér. I Math.* 305 (1987), p. 805-808.

[177] Yann Brenier. "Polar factorization and monotone rearrangement of vector-valued functions". In : *Communications on pure and applied mathematics* 44.4 (1991), p. 375-417.

[178] Cédric Villani. *Topics in Optimal Transportation*. 58. American Mathematical Soc., 2003.

[179] Filippo Santambrogio. "Optimal transport for applied mathematicians". In : *Birkäuser, NY* 55.58-63 (2015), p. 94.

[180] Yossi Rubner, Carlo Tomasi et Leonidas J Guibas. "The earth mover's distance as a metric for image retrieval". In : *International journal of computer vision* 40 (2000), p. 99-121.

[181] Zihao Zhou et al. "Few-Shot Cross Domain Battery Capacity Estimation". In : *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*. 2021, p. 703-711. doi : http://dx.doi.org/10.1145/3460418.3480409.

[182] Martin Arjovsky, Soumith Chintala et Léon Bottou. "Wasserstein generative adversarial networks". In : *International conference on machine learning*. PMLR. 2017, p. 214-223. url : https://proceedings.mlr.press/v70/arjovsky17a.html.

[183] Nicholas Kolkin, Jason Salavon et Gregory Shakhnarovich. "Style transfer by relaxed optimal transport and self-similarity". In : *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, p. 10051-10060. url : 10.1109/CVPR.2019.01029.

[184] Brenier. "Décompositions polaire et réarrangements monotono". In : (1987).

[185] Yann Brenier. "Polar factorization and monotone rearrangement of vector-valued functions". en. In : *Communications on Pure and Applied Mathematics* 44.4 (juin 1991), p. 375-417. issn : 00103640, 10970312. doi : 10.1002/cpa.3160440402. url : https://onlinelibrary.wiley.com/doi/10.1002/cpa.3160440402 (visité le 21/06/2023).

[186] Remi Flamary et al. "POT : Python Optimal Transport". en. In : (2021), p. 8.

[187] Marco Cuturi. "Sinkhorn distances : Lightspeed computation of optimal transport". In : *Advances in neural information processing systems* 26 (2013). url : https://papers.nips.cc/paper_files/paper/2013/hash/af21d0c97db2e27e13572cbf59eb343d-Abstract.html.

[188] Anton Mallasto et Aasa Feragen. "Learning from uncertain curves : The 2-Wasserstein metric for Gaussian processes". In : *Advances in Neural Information Processing Systems* 30 (2017).

[189] Anton Mallasto, Augusto Gerolin et Hà Quang Minh. "Entropy-regularized 2-Wasserstein distance between Gaussian measures". In : *Information Geometry* 5.1 (2022), p. 289-323. doi : https://doi.org/10.1007/s41884-021-00052-8.

[190] Eustasio del Barrio et Jean-Michel Loubes. "The statistical effect of entropic regularization in optimal transportation". In : *arXiv preprint arXiv :2006.05199* (2020).

[191] Julien Rabin et al. "Wasserstein barycenter and its application to texture mixing". In : *Scale Space and Variational Methods in Computer Vision : Third International Conference, SSVM 2011, Ein-Gedi, Israel, May 29–June 2, 2011, Revised Selected Papers 3*. Springer. 2012, p. 435-446.

[192] Facundo Mémoli. "Gromov–Wasserstein distances and the metric approach to object matching". In : *Foundations of computational mathematics* 11 (2011), p. 417-487. doi : https://doi.org/10.1007/s10208-011-9093-5.

[193] Titouan Vayer et al. "Fused Gromov-Wasserstein distance for structured objects". In : *Algorithms* 13.9 (2020), p. 212.

[194] Khang Le et al. "Entropic gromov-wasserstein between gaussian distributions". In : *International Conference on Machine Learning*. PMLR. 2022, p. 12164-12203. url : https://proceedings.mlr.press/v162/le22a.html.

[195] Jean-David Benamou. "Numerical resolution of an "unbalanced" mass transport problem". In : *ESAIM : Mathematical Modelling and Numerical Analysis* 37.5 (2003), p. 851-868.

[196] Thibault Séjourné, Gabriel Peyré et François-Xavier Vialard. "Unbalanced optimal transport, from theory to numerics". In : *arXiv preprint arXiv :2211.08775* (2022).

[197] Jean-David Benamou et Yann Brenier. "A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem". In : *Numerische Mathematik* 84.3 (2000), p. 375-393.

[198] Pedro C. Álvarez-Esteban et al. "A fixed-point approach to barycenters in Wasserstein space". In : *Journal of Mathematical Analysis and Applications* 441.2 (2016), p. 744-762. doi : https://doi.org/10.1016/j.jmaa.2016.04.045.

[199] Marco Cuturi et Arnaud Doucet. "Fast computation of Wasserstein barycenters". In : *International conference on machine learning*. PMLR. 2014, p. 685-693. url : https://proceedings.mlr.press/v32/cuturi14.html.

[200] Alexander Petersen et Hans-Georg Müller. "Fréchet regression for random objects with Euclidean predictors". In : *The Annals of Statistics* 47.2 (2019), p. 691-719. doi : 10.1214/17-AOS1624.

[201] Luc Brogat-Motte et al. "Learning to predict graphs with fused Gromov-Wasserstein barycenters". In : *International Conference on Machine Learning*. PMLR. 2022, p. 2321-2335. URL : https://icml.cc/virtual/2022/spotlight/18000.

[202] N. Bonneel, G. Peyré et M. Cuturi. "Wasserstein barycentric coordinates : histogram regression using optimal transport". In : *ACM Trans. Graph.* 35.4 (2016), p. 153-179. DOI : http://dx.doi.org/10.1145/2897824.2925918.

[203] Ioannis Tsochantaridis et al. "Large margin methods for structured and interdependent output variables." In : *Journal of machine learning research* 6.9 (2005).

[204] Sinho Chewi et al. "Gradient descent algorithms for Bures-Wasserstein barycenters". In : *Conference on Learning Theory*. PMLR. 2020, p. 1276-1304. URL : https://proceedings.mlr.press/v125/chewi20a.html.

[205] Jianing Fan et Hans-Georg Müller. "Conditional Wasserstein Barycenters and Interpolation/Extrapolation of Distributions". In : *arXiv preprint arXiv :2107.09218* (2021).

[206] Roger Penrose. "A generalized inverse for matrices". In : *Mathematical proceedings of the Cambridge philosophical society*. T. 51. 3. Cambridge University Press. 1955, p. 406-413. DOI : http://dx.doi.org/10.1017/S0305004100030401.

[207] Alexander Petersen, Xi Liu et Afshin A. Divani. "Wasserstein F-tests and confidence bands for the Fréchet regression of density response curves". In : *The Annals of Statistics* 49.1 (2021), p. 590-611. DOI : 10.1214/20-AOS1971.

[208] Isabel Haasler et Pascal Frossard. *Bures-Wasserstein Means of Graphs*. arXiv :2305.19738 [cs, eess, stat]. Mai 2023. URL : http://arxiv.org/abs/2305.19738 (visité le 13/07/2023).

[209] Witold Marańda. "Capacity degradation of lead-acid batteries under variable-depth cycling operation in photovoltaic system". In : *2015 22nd International Conference Mixed Design of Integrated Circuits & Systems (MIXDES)*. IEEE. 2015, p. 552-555.

[210] Sang-Hoon Hwang et al. "Reconfigurable hybrid resonant topology for constant current/-voltage wireless power transfer of electric vehicles". In : *Electronics* 9.8 (2020), p. 1323.

[211] Alexander G. de G. Matthews et al. "GPflow : A Gaussian process library using TensorFlow". In : *Journal of Machine Learning Research* 18.40 (avr. 2017), p. 1-6. URL : http://jmlr.org/papers/v18/16-537.html.

[212] Martín Abadi et al. *TensorFlow : Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL : https://www.tensorflow.org/.

[213] Ciyou Zhu et al. "Algorithm 778 : L-BFGS-B : Fortran subroutines for large-scale bound-constrained optimization". In : *ACM Transactions on mathematical software (TOMS)* 23.4 (1997), p. 550-560. DOI : http://dx.doi.org/10.1145/279232.279236.

[214] Richard H Byrd et al. "A limited memory algorithm for bound constrained optimization". In : *SIAM Journal on scientific computing* 16.5 (1995), p. 1190-1208. DOI : http://dx.doi.org/10.1137/0916069.

[215] Dmitry Kobak, Jonathan Lomond et Benoit Sanchez. "The optimal ridge penalty for real-world high-dimensional data can be zero or negative due to the implicit ridge regularization". In : *The Journal of Machine Learning Research* 21.1 (2020), p. 6863-6878. URL : https://jmlr.org/papers/v21/19-844.html.