

Titre: Prédiction par apprentissage machine de la vitesse de propagation de la zone normale d'un ruban supraconducteur à haute température critique
Title:

Auteur: Maxime Moreau
Author:

Date: 2023

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Moreau, M. (2023). Prédiction par apprentissage machine de la vitesse de propagation de la zone normale d'un ruban supraconducteur à haute température critique [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie.
Citation: <https://publications.polymtl.ca/55071/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/55071/>
PolyPublie URL:

Directeurs de recherche: Frédéric Sirois
Advisors:

Programme: Génie électrique
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Prédiction par apprentissage machine de la vitesse de propagation de la zone
normale d'un ruban supraconducteur à haute température critique**

MAXIME MOREAU

Département de génie électrique

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

Août 2023

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

Prédiction par apprentissage machine de la vitesse de propagation de la zone normale d'un ruban supraconducteur à haute température critique

présenté par **Maxime MOREAU**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

a été dûment accepté par le jury d'examen constitué de :

Jean-Pierre DAVID, président

Sirois FRÉDÉRIC, membre et directeur de recherche

Daniel ALOISE, membre

REMERCIEMENTS

Je tiens à remercier mon directeur Frédéric Sirois et Christian Lacroix pour m'avoir proposé ce projet et par le fait même de m'avoir fait découvrir le domaine de l'intelligence artificielle. Je les remercie pour le soutien et la confiance que j'ai reçus d'eux tout au long de ma maîtrise. Je tiens également à remercier le Conseil national de recherches du Canada pour son soutien pendant ma maîtrise et surtout Julio J. Valdés et Alain Tchagang pour nos discussions qui m'ont tant appris sur le domaine de l'intelligence artificielle. Je remercie également Jean-Hugues Fournier-Lupien pour son soutien et ses conseils au début de ma maîtrise.

Je tiens également à remercier tous mes amis du laboratoire avec qui j'ai eu la chance de partager mes questionnements et mes problèmes. Vous m'avez aidé plus que vous ne le pensez. Je remercie également ma famille, mes amis et ma conjointe Sabrina pour leurs encouragements et leur soutien tout au long de mes études.

RÉSUMÉ

Les fils (rubans) supraconducteurs à haute température critique sont des dispositifs prometteurs pour la conception de dispositif comme des moteurs à haut rendement, des limiteurs de courant de court-circuit et des génératrices d'éoliennes compactes. Les procédés de fabrication actuels font que certaines propriétés comme le courant critique peuvent varier de $\pm 20\%$ le long du ruban. En injectant un courant supérieur au courant critique, le ruban devient non-supraconducteur, ou "normal", et risque d'être détruit si la chaleur générée localement est trop grande. Il est donc nécessaire que cette zone normale se propage rapidement pour uniformiser la génération de chaleur.

La dynamique de la propagation de la zone normale dans un ruban supraconducteur est un procédé hautement non-linéaire et difficile à simuler numériquement. Ce mémoire tente d'appliquer des outils d'apprentissage machine pour déterminer l'impact des paramètres géométriques et des propriétés de matériaux de rubans supraconducteurs sur la vitesse de propagation de la zone normale (NZPV) dans le but d'obtenir un outil de conception pour ceux-ci. L'objectif de ce projet est d'entraîner un modèle avec un minimum de données en obtenant une erreur aussi faible que possible sur la prédiction de la NZPV. Afin d'y parvenir, il est nécessaire de choisir quelles données ont le plus de valeur à être ajoutées au jeu de données d'entraînement afin d'améliorer la qualité et la robustesse du modèle. Pour minimiser le nombre de simulations nécessaire pour produire un jeu de données, une méthodologie d'apprentissage actif par requête de comité a été implémentée.

Les deux objectifs du mémoire sont de développer une méthodologie pour construire un jeu de données fournissant suffisamment d'information pour entraîner un modèle de régression avec une erreur de prédiction donnée et de trouver quel modèle de régression est le plus adapté pour prédire la vitesse de propagation de la zone normale d'un ruban supraconducteur. La recherche faite dans ce mémoire se concentre sur les méthodes d'intelligence artificielle telles que l'apprentissage actif et les ensembles de modèles. La méthodologie a été développée sur un jeu de données généré à partir du modèle analytique de la vitesse de propagation de la zone normale créé par Lawrence Dresner.

La méthodologie a été testée sur des données générées par simulation par éléments finis. La première architecture de ruban est une architecture ressemblant aux rubans commerciaux de la compagnie SuperPower Inc. La génération d'un modèle permettant de prédire la NZPV sur ce type de ruban a nécessité 40 données. Le jeu de données initial contenait 25 données et il a fallu 3 itérations d'apprentissage actif où 5 données ont été ajoutées à chaque itération.

pour atteindre une RMSE de 1.89 cm/s sur le jeu de données test. La deuxième architecture étudiée est l'architecture CFD hybride développée par Polytechnique Montréal. Le jeu de données initiales pour ce problème contenait 27 données. L'ensemble de modèles a atteint une RMSE de 5.63 cm/s après une itération et 32 données d'entraînement.

Le travail fait dans ce projet de maîtrise montre qu'il est possible de construire des jeux de données avec un petit nombre de données tout en prédisant la NZPV avec une RMSE satisfaisante.

ABSTRACT

High-temperature superconducting (HTS) tapes are promising devices for the design of high-efficiency motors, fault current limiters, and compact wind turbine generators. Current manufacturing processes result in certain properties, such as critical current, varying by around $\pm 20\%$ around their average values. When a current higher than the critical current is injected, the tape becomes non-superconducting or "normal" and may be destroyed if the locally generated heat is too high. Therefore, it is necessary for the normal zone to propagate rapidly to ensure uniform heat generation in the tape.

The dynamics of normal zone propagation in a tape is a highly nonlinear and challenging process to simulate numerically. This master's thesis aims to apply machine learning tools to determine the impact of geometric parameters and material properties of HTS tapes on the normal zone propagation velocity (NZPV) in order to develop a design tool for these tapes. The project's objective is to train a model with minimal data and achieve the lowest possible error in predicting the NZPV. To achieve this, it is necessary to choose which data are most valuable to add to the training dataset to improve the quality and robustness of the model. To minimize the number of simulations required to generate a dataset, an active learning methodology using query by committee has been implemented.

The two objectives of this master's thesis are to develop a methodology for constructing a dataset that provides enough information to train a regression model with a given prediction error and to find the most suitable regression model for predicting the normal zone propagation velocity in an HTS tape. The research in this thesis focuses on artificial intelligence methods such as active learning and model ensembles. The methodology was developed on a dataset generated from Lawrence Dresner's analytical model of normal zone propagation velocity.

The methodology was tested on data generated by finite element simulation. The first tape architecture studied is similar to the commercial tapes from the company SuperPower Inc. Generating a model to predict NZPV for this type of tape required 40 data points. The initial dataset contained 25 data points, and it took 3 active learning iterations, with 5 data points added in each iteration, to achieve an RMSE of 1.89 cm/s on the test dataset. The second architecture studied is the hybrid CFD architecture developed by Polytechnique Montréal. The initial dataset for this problem contained 27 data points. The ensemble of models achieved an RMSE of 5.63 cm/s after one iteration and 32 data points.

The work in the master's thesis has shown that the methodology makes it possible to generate

datasets with a small amount of data that allow to predict the NZPV of superconductive tapes with a satisfactory RMSE.

TABLE DES MATIÈRES

REMERCIEMENTS	iii
RÉSUMÉ	iv
ABSTRACT	vi
TABLE DES MATIÈRES	viii
LISTE DES TABLEAUX	x
LISTE DES FIGURES	xi
LISTE DES SIGLES ET ABRÉVIATIONS	xiii
CHAPITRE 1 INTRODUCTION	1
1.1 Mise en contexte et objectif	1
1.2 Problématique	2
1.3 Objectif	3
CHAPITRE 2 REVUE DE LITTÉRATURE : RUBANS SUPRACONDUCTEURS	4
2.1 Caractéristique des matériaux supraconducteurs	4
2.1.1 Courant critique	4
2.1.2 Stabilité des supraconducteurs	4
2.1.3 Vitesse de propagation de la zone normale (NZPV)	6
2.2 Rubans supraconducteurs	6
2.2.1 Ruban uniforme	6
2.2.2 Architecture CFD	7
2.2.3 Architecture hybride	10
2.3 Modélisation de la propagation de la zone normale	10
2.3.1 Modèle analytique 1-D de la NZPV	12
2.3.2 Limitation du modèle de Dresner	15
CHAPITRE 3 MÉTHODES DE RÉGRESSION ET MODÈLES D'APPRENTISSAGE	
MACHINE	17
3.1 Modèles de régression	17
3.2 Ensemble de modèles	20

3.3	Apprentissage actif	21
3.3.1	Apprentissage actif pour la prédiction de la NZPV	23
3.3.2	Génération des données	25
3.3.3	Algorithme	26
3.4	Métriques d'évaluations	27
3.4.1	Racine de l'écart quadratique moyen	27
3.4.2	Test Gamma	28
CHAPITRE 4 CHOIX DE LA MÉTHODOLOGIE : UNE APPROCHE PRATIQUE		32
4.1	Données d'entraînement - modèle de Dresner	32
4.1.1	Évaluation de la qualité des données	34
4.2	Analyse des modèles de prédiction	36
4.3	Apprentissage actif	39
4.3.1	Erreur généralisée et requête par comité	41
4.4	Performance des modèles	43
4.5	Discussion	47
CHAPITRE 5 PRÉDICTION DE LA NZPV D'UN RUBAN SUPRACONDUCTEUR		48
5.1	Générations des données	49
5.1.1	Simulation par éléments finis	49
5.1.2	Calcul de la NZPV	51
5.2	Performance d'un modèle entraîné sur une simulation d'un ruban régulier . .	55
5.2.1	Entraînement des modèles	56
5.3	Performance d'un modèle entraîné sur une simulation d'un ruban SuperPower CFD hybride	62
5.3.1	Entraînement des modèles	64
CHAPITRE 6 CONCLUSION		67
6.1	Synthèse des travaux	67
6.2	Limitations de la solution proposée	68
6.3	Améliorations futures	68
RÉFÉRENCES		70

LISTE DES TABLEAUX

4.1	Propriétés des matériaux	33
4.2	Paramètres du système	33
5.1	Paramètres du ruban régulier	55

LISTE DES FIGURES

1.1	Géométrie d'un ruban 2G-HTS avec architecture hybride-CFD	3
2.1	Coupe transversale d'un ruban supraconducteur commercial	8
2.2	Coupe transversale d'un ruban CFD	9
2.3	Chemin du courant pour l'architecture uniforme et pour l'architecture CFD	9
2.4	Effet de la résistance d'interface sur la NZPV	10
2.5	Coupe transversale d'un ruban hybride	11
2.6	Solution de l'équation 2.8	14
2.7	Schéma de la coupe transversale d'un ruban supraconducteur	15
3.1	Schéma d'un réseau de neurones	20
3.2	Schéma du principe d'un ensemble de modèles	22
3.3	Schéma d'une boucle d'apprentissage actif.	24
3.4	Distribution des données initiales et valeurs de NZPV associées	24
3.5	Exemple de la racine carrée de l'erreur quadratique moyenne partielle	29
3.6	Distribution de la RMSE partielle sur la prédiction selon les différents paramètres	30
3.7	Régression du test Gamma	31
4.1	Distribution de la vitesse de propagation de la zone normale	33
4.2	Effet du bruit sur Γ en fonction du nombre de donnée	35
4.3	Effet du bruit sur la MSE d'entraînement, de validation et de test . .	35
4.4	Apprentissage d'un réseau de neurones	37
4.5	Performance des modèles	38
4.6	Régression de quatre différents modèles de prédiction sur les données de tests.	40
4.7	Distribution des données initiales et des données ajoutées	43
4.8	Comparaison du désaccord du comité et de la RMSE partielle	44
4.9	Effet des données d'entraînement sur la RMSE	46
5.1	Géométrie d'un ruban 2G-HTS avec architecture hybride-CFD	48
5.2	Distribution du ratio du courant critique selon la longueur du courant	50
5.3	Propriétés des matériaux utilisés dans le logiciel COMSOL.	51
5.4	Propagation de la zone normale	52
5.5	Mesures de températures des sondes dans une simulation par éléments finis	53

5.6	Calcul de la NZPV	53
5.7	Géométrie simulée par éléments finis d'un ruban BARE	54
5.8	Distribution du jeu de données initial pour le ruban régulier	56
5.9	Outil d'entraînement de réseau de neurones (nntraintool) du “ <i>Deep Learning Toolbox</i> ” sur MATLAB.	57
5.10	Désaccord du comité à la première itération	58
5.11	Évolution de la RMSE de l'ensemble de modèles	60
5.12	Courbe de prédictions selon les NZPV de rubans réguliers simulés . .	60
5.13	Distribution des données finales du ruban régulier	61
5.14	Surface de prédiction de la NZPV d'un ruban BARE	61
5.15	Distribution du jeu de données initial d'un ruban hybride	62
5.16	Évolution de la RMSE des données du ruban hybride	63
5.17	Courbe de prédictions selon les NZPV de rubans Hybride simulés . .	64
5.18	Distribution des données d'entraînement, de validation et de test dans l'espace des entrées.	66
5.19	Surface de prédiction de la NZPV d'un ruban hybride	66

LISTE DES SIGLES ET ABRÉVIATIONS

1-D	Une dimension
2G HTS CC	Second generation high temperature superconductive coated conductor - Supraconducteur recouvert de seconde génération à haute température critique
BCS	Théorie de Bardeen, Cooper et Schrieffer
CFD	Current flow diverter - Déviateur du flux du courant
b-CFD	Buffer-layer-CFD - Couche tampons-CFD
CTL	Current transfer length - Longueur de transfert du courant
GE	Generalized error - Erreur généralisée
GPR	Gaussian process regression - Régression par processus gaussien
NN	Neural network - Réseau de neurones
NZPV	Normal zone propagation velocity - Vitesse de propagation de la zone normale
RMSE	Root mean square error - Racine carrée de l'erreur quadratique moyenne
s.d.	Sans dimension
SVR	Support vector regression - Régression par vecteur de support
λ	fraction du volume du supraconducteur (s.d.)
ρ	Résistivité ($\Omega \cdot m$)
ρ_m	Densité massique ($kg \cdot m^{-3}$)
A	Surface de la section d'un ruban (m^2)
B_c	Champ magnétique critique (T)
C_p	Capacité thermique ($J \cdot K^{-1}$)
f	Fraction du volume du stabilisateur
f	Ratio de couverture (s.d.)
L	Longueur du ruban (m)
i	Ratio du courant appliqué sur le critique appliqué (s.d.)
I	Courant (A)
I_c	Courant critique (A)
J	Densité de courant ($A \cdot m^{-2}$)
J_c	Densité de courant critique ($A \cdot m^{-2}$)
k	Conductivité thermique ($W \cdot m^{-1} \cdot K^{-1}$)
Q	Pertes par effet Joules ($W \cdot m^{-2}$)
S	Capacité calorifique ($J \cdot K^{-1} \cdot m^{-3}$)

T	Température (K)
T_b	Température du bain thermique (K)
T_c	Température critique (K)
T_{cs}	Température de partage du courant (K)
w	Largeur du ruban (m)
w_f	Largeur de la surface résistive (m)

CHAPITRE 1 INTRODUCTION

1.1 Mise en contexte et objectif

Utilité de la supraconductivité

La supraconductivité est un comportement électromagnétique des matériaux décrit par deux phénomènes : une conductivité parfaite et un diamagnétisme parfait. Les recherches et le développement de ces matériaux ont débuté au 20e siècle et se poursuivent encore aujourd'hui. L'avantage principal des matériaux supraconducteurs est qu'ils supportent de fortes densités de courant sans générer de chaleur. Les supraconducteurs de pointe peuvent atteindre des densités de courant allant jusqu'à 3.125 MA/cm^2 à 77 K, ce qui est deux à trois ordres de grandeur supérieurs à celles du cuivre. Sans les supraconducteurs, il n'y aurait pas d'imagerie par résonance magnétique, d'ordinateurs quantiques et il y aurait moins de train à lévitation magnétique à très grande vitesse. D'autres développements intéressants avec ces matériaux sont les câbles de transport d'énergie, le stockage direct d'électricité ou encore l'utilisation de puces électroniques à très faible génération de chaleur [1]. Plus de 110 ans de recherche et 5 prix Nobel ont permis d'approfondir la connaissance de la supraconductivité telle qu'elle est aujourd'hui.

Historique

En 1911, Heike Kamerlingh Onnes a mesuré une résistivité nulle pour le mercure dans un bassin d'hélium liquide, ce qui lui a valu un prix Nobel de physique en 1913, initiant le monde à la supraconductivité [2, 3]. En 1933, Meissner et Ochsenfeld ont découvert que le champ magnétique est exclu d'un matériau à l'état supraconducteur. Ce phénomène porte le nom d'effet Meissner en l'honneur de son découvreur [4]. La première théorie acceptée de la supraconductivité a été décrite en 1950 par les deux physiciens russes Vitaly Ginzburg et Lev Landau. Cette théorie a permis à Aleksei Abrikossov en 1957 d'interpréter le changement de phase d'un matériau supraconducteur de type II lors de l'apparition de vortex entre deux champs magnétiques critiques [5]. Ginzburg, Abrikossov et Anthony J. Leggett se sont partagé le prix Nobel de 2003 pour ces travaux. En 1957, une deuxième théorie de la supraconductivité a vu le jour. La théorie de Bardeen, Cooper et Schrieffer (BCS) se concentre sur le comportement des électrons et leurs interactions à très basse température. Ils expliquent comment deux électrons peuvent s'attirer à travers le réseau atomique d'un matériau pour former une paire d'électrons. Ces paires d'électrons portent le nom de paires de Cooper [6–9]. Il a fallu 15 ans, soit en 1972, pour que ces physiciens reçoivent leurs prix Nobel. C'est en

1986 qu’une autre étape majeure vers le développement d’applications à base de matériaux supraconducteurs a été franchie lorsque Johannes G. Bednorz et Karl A. Müller ont découvert un matériau de type céramique dont la température critique est plus grande que 77 K, le point d’ébullition de l’azote liquide. Avant cette découverte, les matériaux supraconducteurs ne fonctionnaient qu’à des températures extrêmement basses.

La découverte de ces céramiques supraconductrices à haute température critique a eu un impact sur plusieurs domaines technologiques. Elle a mené entre autres à la fabrication de rubans supraconducteurs. La caractérisation de ces dispositifs est importante pour le développement de cette technologie, afin d’en maximiser la performance et pour minimiser les coûts de fabrication. Les propriétés de ces dispositifs, notamment le courant critique I_C , varient fortement avec la présence de défauts. Ces défauts surviennent lors du procédé de fabrication et sont difficiles à prévoir et ils entraînent une variation des propriétés du ruban le long de celui-ci.

Lorsque le courant injecté est supérieur au courant critique local, le supraconducteur devient normal à cet endroit, créant un point chaud. Si la production de chaleur n’est pas interrompue, la température du dispositif augmente et dépasse la température critique localement. Le point chaud crée une zone dans la céramique qui n’est plus supraconductrice que l’on nomme la zone normale et qui se propage alors longitudinalement dans le ruban. Dans ce cas, il est nécessaire que la chaleur se propage rapidement pour accélérer la détection de ces points chauds. S’ils ne sont pas détectés, le dispositif sera détruit de façon irréversible.

1.2 Problématique

Lors de la conception d’un ruban supraconducteur, la sélection des paramètres géométriques et des matériaux est une tâche difficile. Il n’existe pas de modèle analytique permettant d’obtenir la vitesse de propagation de la zone normale (*Normal zone propagation velocity*, ou NZPV) prenant en compte tous les matériaux utilisés et l’architecture du ruban. La méthode courante consiste à simuler le problème électrothermique par éléments finis et de calculer la NZPV à partir des résultats des simulations. Un exemple du domaine de simulation d’un ruban de type hybride-CFD est présenté à la figure 1.1. Chaque simulation peut prendre plusieurs heures et certains paramètres numériques de simulation sont difficiles à ajuster afin d’obtenir une solution qui converge. Il n’est donc pas encore possible de simuler de longs rubans, entre autres choses.

Ce mémoire propose d’utiliser des méthodes d’apprentissage machine pour assister à la conception de rubans supraconducteurs et développer un outil de prédiction de la NZPV

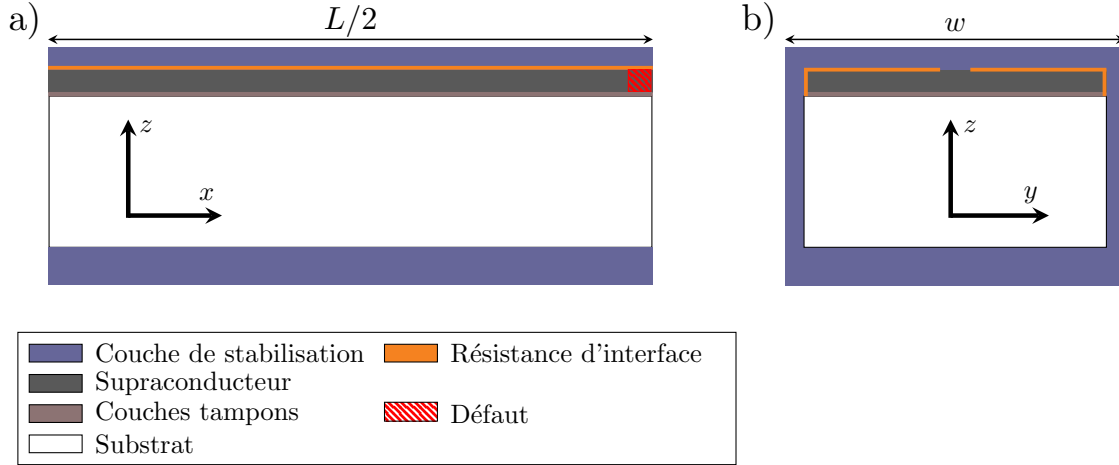


FIGURE 1.1 Géométrie d'un ruban 2G-HTS avec architecture hybride-CFD utilisée pour la simulation par éléments finis. a) Section du ruban selon la longueur (axe x), et b) section selon la largeur. Les dimensions ne sont pas à l'échelle et $L \gg w$.

en fonction des caractéristiques du ruban et du courant qui y est injecté.

1.3 Objectif

L'objectif de cette maîtrise est de développer une méthodologie pour entraîner un modèle de prédiction de la NZPV en utilisant un minimum de données provenant de simulation par éléments finis tout en garantissant une précision satisfaisante. Ce modèle doit capturer l'impact des paramètres géométriques et physiques sur la NZPV pour assister à la conception de ces rubans.

La méthodologie est validée avec un modèle analytique simple qui approxime le comportement de la NZPV. Ensuite, un modèle de régression est entraîné sur des simulations par éléments finis des rubans supraconducteurs. Ces simulations ont été au préalable calibrées expérimentalement. Le cœur de ce travail consiste donc à déterminer quel ensemble de paramètres doit être choisi pour obtenir le plus d'information utile à l'entraînement d'un modèle prédictif. La qualité d'un modèle prédictif a une forte dépendance avec la qualité et la quantité des données utilisées pour l'entraîner.

Ce mémoire est organisé en plusieurs chapitres, notamment une revue de littérature sur le comportement des rubans supraconducteurs (chapitre 2) et un aperçu des méthodes et des modèles de prédiction (chapitre 3), afin d'introduire les concepts nécessaires à la compréhension du travail. Le choix de la méthodologie est présenté dans le chapitre 4, les résultats dans le chapitre 5, et la conclusion dans le chapitre 6.

CHAPITRE 2 REVUE DE LITTÉRATURE : RUBANS SUPRACONDUCTEURS

2.1 Caractéristique des matériaux supraconducteurs

Dans un matériau supraconducteur, le courant circule sans résistance, ce qui signifie qu'il n'y a pas de perte d'énergie sous forme de chaleur. Cependant, pour qu'un matériau soit considéré comme supraconducteur, il doit également présenter un diamagnétisme parfait, c'est-à-dire qu'il repousse complètement le champ magnétique externe. La supraconductivité est une propriété qui se manifeste lorsque le matériau est refroidi sous une certaine température critique T_c et qu'il est soumis à un champ inférieur à son champ critique B_c . Il existe d'autres facteurs pouvant affecter la supraconductivité d'un matériau comme sa structure atomique, la pression ou le courant transporté. Il est donc important de comprendre comment ce type de matériau fonctionne pour pouvoir utiliser efficacement ses propriétés.

2.1.1 Courant critique

L'état supraconducteur dans un matériau ne peut pas exister au-dessus d'une température critique T_c propre à chaque matériau. Les supraconducteurs sont donc plongés dans un bain thermique comme l'azote liquide ou l'hélium liquide. Ceci dit, il y a d'autres conditions qui peuvent détruire la supraconductivité d'un matériau comme le champ magnétique critique B_c et la densité de courant critique J_c . Lorsque le courant injecté est supérieur au courant critique, le matériau devient résistif et il commence à chauffer. Le courant critique varie avec la température, plus la température est élevée, plus le courant critique est faible et vice versa.

2.1.2 Stabilité des supraconducteurs

Afin de tirer pleinement parti des caractéristiques d'un matériau supraconducteur, il est couramment nécessaire de l'incorporer dans un système particulier, tel qu'un fil ou un ruban. Ce procédé vise à obtenir les propriétés électriques et mécaniques souhaitées en fonction de l'application visée. Plusieurs procédés de microfabrication sont utilisés industriellement pour la production de rubans supraconducteurs, tels que le *Rolling Assisted Biaxially Textured Substrate* (RABiTS), l'*Ion Beam Assisted Deposition* (IBAD) ou encore la *solution deposition planarization* (SDP). Cependant, ces procédés de fabrication peuvent produire des défauts qui affectent les propriétés du système.

Les défauts dans un ruban supraconducteur font varier localement les propriétés de celui-ci.

Lorsque le courant critique (I_c) diminue, il devient fort probable qu'il perde son état supraconducteur et qu'il commence à dissiper de la chaleur. La transition de l'état supraconducteur à l'état normal est appelée un quench. Puisque les matériaux supraconducteurs utilisés pour les rubans sont des céramiques, ils sont très résistifs électriquement lorsqu'ils sont à l'état normal. Les rubans sont donc dotés d'une couche de stabilisateur comme du cuivre ou de l'argent afin que le courant passe à travers cette couche en cas de quench. Cette couche génère moins de chaleur que le supraconducteur devenu résistif, ce qui réduit les risques de bris du dispositif.

Lorsqu'un courant circule dans un supraconducteur, il y a trois régimes de stabilité. Si le système ne génère pas de température, il est dans le régime cryostable. Si le refroidissement est plus fort que la génération de chaleur, il est conditionnellement stable. Lorsque la génération de chaleur est plus grande que le refroidissement, il est dans un régime instable. Quand le courant dans tout le ruban est inférieur au courant critique, aucune chaleur n'est produite. Si une perturbation se produit dans le ruban et qu'elle génère une zone normale, elle s'estompe uniformément du système. En 1965, Stekly et Zar [10] ont mené une expérience où un supraconducteur a été perturbé en zone normale, mais est redevenu supraconducteur après refroidissement à l'hélium liquide. La chaleur émise par le matériau par effet Joule doit être compensée par ce refroidissement. Les conditions de stabilité sont exprimées par le critère de Stekly à l'équation (2.1), à savoir que la génération de chaleur doit être inférieure à la capacité de refroidissement du système :

$$Q = \rho J^2 A / (fP) < h(T_c - T_b). \quad (2.1)$$

Ici, Q est la chaleur générée par unité de surface refroidie, ρ est la résistivité du stabilisateur tel que le cuivre ou l'argent, J est la densité de courant du système, A est l'aire de la coupe transversale du système, f est la fraction de volume de stabilisateur, P est le périmètre du système, h est le coefficient de transfert thermique, T_c et T_b sont respectivement la température critique et la température de bain du liquide de refroidissement. En 1969, Maddock, James et Norris montrent qu'il est possible pour un système avec un supraconducteur d'être stable pour un critère moins restrictif que le critère de Stekly. Ils expliquent qu'au lieu de refroidir uniformément la zone normale, celle-ci se résorbe par ces deux extrémités si le courant est plus grand que celui imposé par le critère de Stekly, mais plus petit que le courant minimum de propagation. Lorsque la chaleur générée est plus grande que la chaleur dissipée, le supraconducteur devient normal localement et diminue donc le courant critique aux extrémités de cette zone normale et elle s'agrandit. Ce phénomène d'emballement thermique fait grandir la zone normale. La vitesse à laquelle la zone normale augmente est la NZPV. La

zone normale devient de plus en plus grande jusqu'à ce que le ruban se détruit ou que tout le ruban devienne normal.

2.1.3 Vitesse de propagation de la zone normale (NZPV)

La NZPV est une caractéristique importante d'un ruban supraconducteur et elle fait l'objet central de ce projet de maîtrise. Il s'agit de la vitesse à laquelle la zone normale s'agrandit après l'apparition d'un quench. Elle varie entre autres avec l'architecture du ruban, les matériaux utilisés, la température externe, le courant imposé et la quantité de stabilisateur utilisée. La valeur de NZPV est utile à connaître puisqu'elle permet de déterminer la vitesse à laquelle un défaut sera détecté. Lorsque la NZPV est rapide, la détection du quench est faite plus rapidement et donc les risques de bris sont réduits [11]. Si l'emballement thermique n'est pas détecté, le ruban peut être endommagé et perdre ses propriétés électriques ou même se rompre. Les rubans supraconducteurs peuvent avoir plusieurs mètres de long et même des kilomètres, il est donc important d'avoir une grande NZPV pour réduire le nombre de détecteurs de quench nécessaire et réduire les frais de fonctionnement [12, 13].

2.2 Rubans supraconducteurs

Les supraconducteurs possèdent des propriétés électriques très intéressantes, mais leur utilisation peut être difficile. Leurs propriétés mécaniques et thermiques sont des facteurs limitant à l'utilisation de ceux-ci. L'utilisation moderne des supraconducteurs est sous la forme d'un ruban flexible et est appelée *second generation high temperature superconductor coated conductor* ou 2G HTS CC. Ces matériaux sont dit à haute température critique puisque la température critique de ceux-ci est plus grande que la température d'ébullition de l'azote liquide ($T_c > 77K$).

2.2.1 Ruban uniforme

Les rubans utilisés actuellement dans l'industrie sont des rubans dits uniformes. Ces rubans sont fabriqués sur un substrat en alliage de nickel ou en acier inoxydable, ce qui offre de bonnes propriétés mécaniques comme une bonne résistance mécanique. Sur cet alliage, plusieurs couches tampons sont déposées pour prévoir la déposition du supraconducteur. Il faut 1) une barrière de diffusion pour éviter la contamination du supraconducteur, 2) une couche dite de texture pour déposer le supraconducteur avec la bonne structure atomique, et 3) une ou plusieurs couches d'adaptation du paramètre de maille afin de diminuer les tensions et de conserver la structure atomique du supraconducteur [14]. Les supraconducteurs de deuxième

génération sont souvent faits de plans d'oxyde de cuivre et de terres rares, appelés REBCO pour *Rare-Earth Barrium Copper Oxide*. Il existe plusieurs stœchiométries permettant d'obtenir des propriétés supraconductrices, comme le $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ ou YBCO. La concentration d'oxygène varie avec la variable x , qui prend des valeurs de 0 à 1. Ce paramètre doit être bien ajusté pour permettre un matériau supraconducteur à haute température critique [15]. Un revêtement de stabilisateur est déposé autour de l'empilement pour ajouter une masse thermique et électrique. Une coupe transversale de ce type de ruban est montrée à la figure 2.1.

L'architecture de ces rubans est faite de manière à ce que le courant circule dans le matériau supraconducteur. Lorsqu'il y a un événement causant un quench, le courant est dévié vers le stabilisateur puisque celui-ci possède une plus grande conductivité que la céramique à l'état normal. Une couche épaisse de stabilisateur implique une grande masse thermique et une grande conductivité, ce qui diminue le chauffage du ruban. Cependant, une plus grande conductivité peut être problématique pour certaines applications comme les limiteurs de courant. Avec une plus grande conductivité, il faut une plus grande longueur de ruban pour atteindre la limitation de courant désirée.

Une autre solution pour éviter une grande génération de chaleur aux points chauds est de détecter rapidement l'apparition de ceux-ci. Pour ce faire, une stratégie est d'augmenter la NZPV. En augmentant celle-ci, la résistivité du ruban augmente plus rapidement, ce qui permet d'accélérer la détection d'un quench. Plusieurs approches permettent d'augmenter la NZPV sans changer la quantité de stabilisateur. Une solution est d'ajouter une résistance d'interface entre le supraconducteur et le stabilisateur. Lorsque le courant passe du supraconducteur au stabilisateur, le courant doit traverser cette interface résistive, ce qui augmente la longueur de transfert du courant (*CTL*). Il a été montré par Levin et al. que la NZPV varie quasi linéairement avec la *CTL* [16].

Cette architecture a cependant un problème majeur : cette surface résistive génère trop de chaleur à l'interface, ce qui augmente le risque de générer un quench. Les travaux de Lacroix *et al.* [17–20] se concentrent à modifier l'architecture des rubans en ajoutant de résistances seulement à certains endroits des rubans afin d'augmenter la NZPV sans sacrifier la stabilité thermique de ceux-ci.

2.2.2 Architecture CFD

L'architecture de ruban CFD (*current flow diverter*) tire son nom de l'effet de la géométrie sur le parcours du courant. La déviation du flux du courant est causée par l'interface supraconducteur-stabilisateur qui est résistive localement. La paroi résistive est au centre

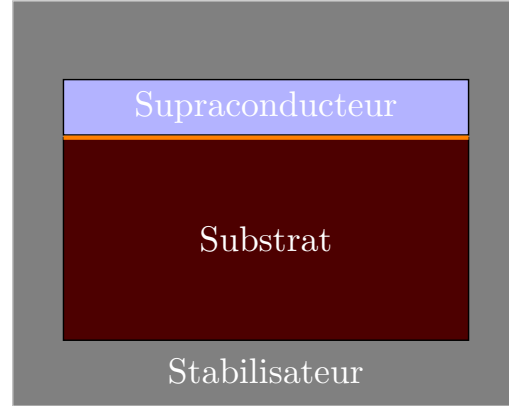


FIGURE 2.1 Coupe transversale d'un ruban supraconducteur commercial. Les grandeurs ne sont pas à l'échelle et la ligne orange entre le supraconducteur et le substrat représente les couches tampons.

du dispositif, ce qui laisse le courant facilement passer sur les côtés du ruban. La figure 2.2 montre la coupe transversale d'un ruban CFD et la figure 2.3 montre le parcours du courant dans le ruban. Le courant contourne la paroi résistive avant d'atteindre le stabilisateur.

Cette architecture augmente la CTL tout en limitant la génération de chaleur à l'interface résistive. La chaleur est générée sur une plus longue région du ruban, ce qui minimise l'apparition de point chaud. La figure 2.4 montre un autre avantage majeur de cette architecture. La NZPV peut être plus élevée d'un ordre de grandeur pour l'architecture CFD que pour un ruban avec une résistance d'interface uniforme.

Lorsque le courant traverse la barrière résistive de l'architecture CFD, de la chaleur est générée à l'interface. En 2018, Fournier *et al.* [21] proposent l'architecture b-CFD (*buffer layer Current Flow Diverter*) qui augmente la CTL sans ajouter de barrière résistive. Pour ce faire, il ajoute une couche de stabilisateur plus grande sous le ruban tout en gardant une mince couche de stabilisateur sur le dessus. Le courant traverse l'épaisseur du ruban par le côté avant d'atteindre la couche de stabilisateur sous le ruban. Cette architecture a l'avantage de garder l'effet CFD tout en ayant une plus grande masse thermique, ce qui rend cette technologie applicable pour des électroaimants nécessitant beaucoup de stabilisateur. Pour un ruban commercial, augmenter la quantité de stabilisateur réduit la NZPV. Ce n'est plus toujours le cas avec l'architecture b-CFD, car la réduction de la NZPV est compensée par l'augmentation de la CTL.

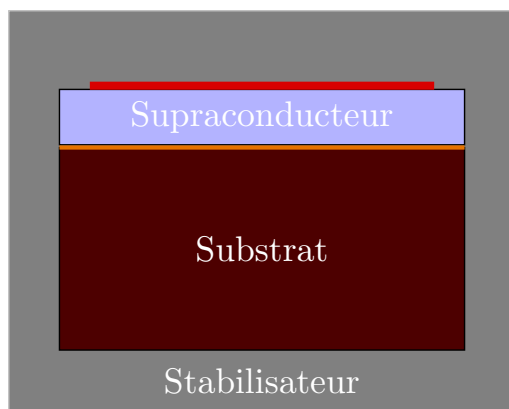


FIGURE 2.2 Coupe transversale d'un ruban CFD. La résistance d'interface est indiquée en rouge et les couches tampons sont en oranges.

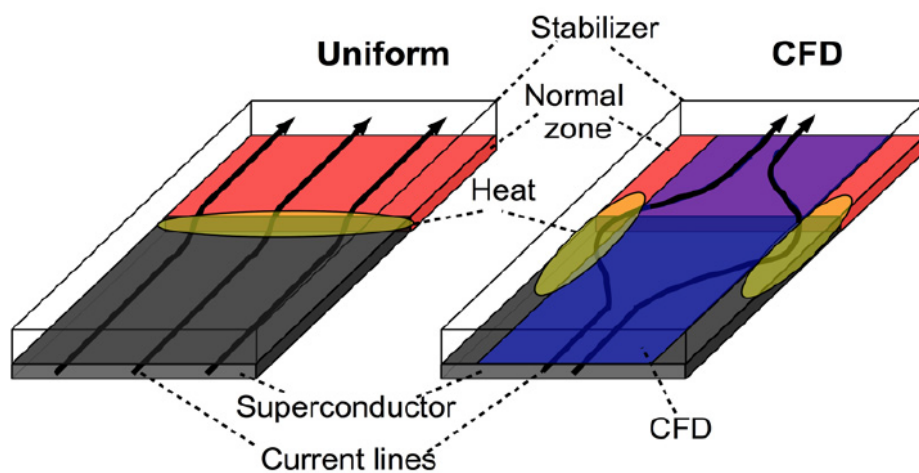


FIGURE 2.3 Parcours du courant pour l'architecture uniforme et pour l'architecture CFD. La zone mauve représente la résistance d'interface élevée et la zone rouge indique la zone normale [18].

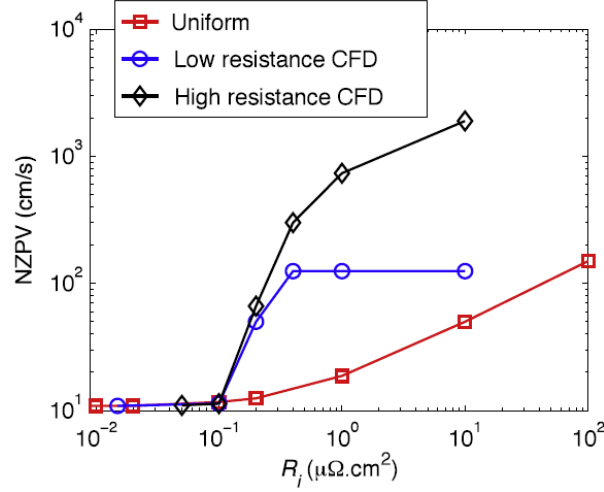


FIGURE 2.4 Effet sur la NZPV de la résistance d'interface (R_i) de ruban supraconducteur ayant une architecture uniforme ou CFD [18].

2.2.3 Architecture hybride

Giguère *et al.* [19] ont proposé une nouvelle architecture pour les rubans CFD et b-CFD, qui utilise une paroi résistive sur une partie de l'interface stabilisateur-supraconducteur, ainsi qu'une couche épaisse de stabilisateur sous le ruban. Ces mécanismes augmentent la CTL, rendant possible d'ajouter du stabilisateur et d'augmenter la NZPV en même temps, ce qui n'était pas possible avec les architectures précédentes.

L'ajout de stabilisateur au bas du ruban permet à plus de courant de passer sous le dispositif en passant par les côtés, ce qui augmente la CTL. La figure 2.5 montre une coupe transversale du ruban ainsi que le parcours du courant.

2.3 Modélisation de la propagation de la zone normale

La vitesse de propagation de la zone normale dans les supraconducteurs de type II est difficile à modéliser, car il y a plusieurs causes physiques impliqués dans le problème. Lorsque de la chaleur est générée par l'effet Joule dans le ruban, cela entraîne un changement des propriétés électriques du ruban. De plus, cette élévation de température peut également provoquer un changement de phase dans le ruban et le rendre non-supraconducteur. L'ensemble de ces interactions contribue à rendre le comportement du système extrêmement complexe et non linéaire. Plusieurs méthodes numériques peuvent être utilisées pour modéliser la transition de l'état supraconducteur à l'état normal, comme les différences finies [22], les circuits équivalents avec les éléments groupés [23] et les éléments finis.

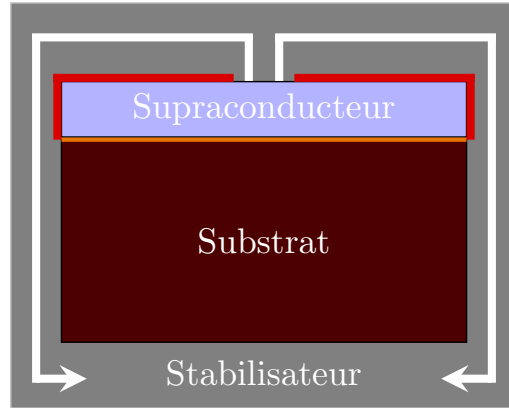


FIGURE 2.5 Coupe transversale d'un ruban hybride. Les flèches indiquent la trajectoire du courant.

Par exemple, certains chercheurs étudient le supraconducteur MgB_2 dans le contexte d'une machine imagerie par résonance magnétique (IRM). Ils étudient différents mécanismes de protection pour le supraconducteur en cas de quench [24,25]. Cependant, la plupart des chercheurs étudient les supraconducteurs à base de terre rare comme le $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$. Wang *et al* étudient l'effet du matériau choisi comme stabilisateur en haut et en bas du ruban supraconducteur [26]. Wan-kan Chan *et al* étudient l'influence de la résistance d'interface entre le supraconducteur et le stabilisateur sur la NZPV. Ces travaux sont orientés vers la protection des dispositifs contre les points chauds apparaissant lors des quenches. Ces chercheurs ont également fait l'étude de la NZPV et de l'énergie nécessaire à l'amorçage d'un quench dans une bobine de ruban supraconducteur [27,28]. D'un point de vue de notre propre laboratoire, les travaux de Lacroix *et al* se concentrent sur l'architecture du ruban, avec objectif de réduire les risques de bris d'un ruban supraconducteur. Ils ont présenté des architectures utilisant des interfaces résistives pour augmenter la NZPV, tout en limitant la génération de chaleur que ces interfaces peuvent provoquer, tel qu'expliqué dans les publications de Lacroix *et al* [17,18] et Giguère *et al* [19]. Les différentes architectures ont été développées avec les éléments finis et testées en laboratoire.

Il existe également des modèles analytiques, mais ceux-ci ne prennent pas en compte certains éléments importants des rubans supraconducteurs, comme l'architecture du fil ou la résistance d'interface. Néanmoins, ces modèles analytiques peuvent servir comme guide pour développer des outils pour assister la conception des rubans supraconducteurs.

2.3.1 Modèle analytique 1-D de la NZPV

Dans le manuel *Advances in Cryogenic Engineering Materials* [29], Lawrence Dresner présente une expression de la NZPV en fonction des paramètres géométriques du supraconducteur et du stabilisateur en utilisant les travaux précédents de Maddock et de Altov [30,31]. La solution en 1 dimension de Dresner est la première expression analytique qui considère le partage de courant entre le supraconducteur et le stabilisateur. La NZPV est dérivée à partir de l'équation de la chaleur sur un ruban supraconducteur avec d'une couche de stabilisation faite de matériaux comme le cuivre ou l'argent. Le dispositif est plongé dans un bain thermique, tel que l'azote liquide à 77 K ou l'hélium liquide à 4.2 K. Le problème est modélisé comme :

$$S \left(\frac{\partial T}{\partial t} \right) = \frac{\partial}{\partial z} \left(k \frac{\partial T}{\partial z} \right) + (Q - q) \frac{P}{A}, \quad (2.2)$$

où S est la capacité calorifique [$\text{J}/(\text{K} \cdot \text{m}^3)$], T est la température à une position z et à un temps t , k est la conductivité thermique du matériau, Q est la chaleur générée par effet Joule, q est le flux de chaleur transféré vers le liquide refroidissant, P est la surface en contact avec le liquide et A est la section d'aire de la coupe transversale du conducteur. En observant la température du front d'onde et non celle de tout le ruban, on cherche maintenant une solution de l'équation de chaleur sous la forme d'une onde progressive, soit $T(z, t) = T(z + vt)$. Le front d'onde de la zone normale se déplace vers la gauche à une vitesse v . Cette vitesse est la vitesse de propagation de la zone normale. La dérivée temporelle devient $\left(\frac{\partial T}{\partial t} \right) = \left(\frac{dT}{dz} \right) \frac{dz}{dt} = \left(\frac{dT}{dx} \right) v$, et en posant $x = vt + z$, l'équation (2.2) devient :

$$-vS \left(\frac{dT}{dx} \right) + \frac{d}{dx} \left(k \frac{dT}{dx} \right) + (Q - q) \frac{P}{A} = 0, \quad (2.3)$$

Le flux par convection (q) est exprimé par la loi de refroidissement de Newton, soit $q = h(T - T_b)$, où h est le coefficient de refroidissement du stabilisateur. La chaleur générée par effet Joule dans le stabilisateur (Q) dépend du courant injecté dans celui-ci. Lorsque la température est proche T_c , le courant est partagé entre le supraconducteur et le stabilisateur. On note T_{cs} la température où le courant commence à être partagé entre le supraconducteur et le stabilisateur. Lorsque la température se rapproche de la température critique, le courant critique du supraconducteur se met à diminuer et plus de courant passe dans le stabilisateur. La chaleur générée selon la température est exprimée à l'équation suivante :

$$\frac{QP}{A} = \begin{cases} \rho_{Ag} J^2 / f & T_c < T, \\ (\rho_{Ag} J^2 / f) \frac{T - T_{cs}}{T_c - T_{cs}} & T_{cs} < T < T_c, \\ 0 & T_b < T < T_{cs}. \end{cases} \quad (2.4)$$

Ici, ρ_{Ag} est la résistivité de l'argent et f la fraction de stabilisateur. Avec cette expression de la chaleur en fonction de la température, on retrouve une expression du partage de courant $g(\tau)$ en fonction de $\tau = T - T_{env}$, T_{env} est la température de l'environnement et $i = I/I_c$ qui est la fraction du courant critique, c'est à dire :

$$g(\tau) = \begin{cases} 1 & \tau_c < \tau, \\ [\tau + \tau_c(i - 1)] / i\tau_c & (1 - i)\tau_c < \tau < \tau_c, \\ 0 & 0 < \tau < (1 - i)\tau_c. \end{cases} \quad (2.5)$$

Avec ces définitions, le système d'unité adimensionnelle et en posant le changement de variable proposé par Maddock *et al.* [30] $s = k(\frac{dT}{dz})$, (2.3) devient :

$$s \left(\frac{ds}{d\tau} \right) - vs + \alpha i^2 g(\tau) - \tau = 0, \quad (2.6)$$

avec $\alpha i^2 = (\rho_{Ag} J^2 / f) / (hP\tau_c / A)$. Cette équation a été résolue pour v en 1979 par Lawrence Dresner [32]. L'expression analytique de la vitesse de propagation de la zone normale pour le cas spécifique sans refroidissement ($h \rightarrow 0$) est :

$$v = \left(\frac{2\lambda J_c}{S} \right) \left(\frac{\rho_{Ag} k}{f\tau_c} \right)^{1/2} \left[\frac{c(i)i^{1/2}}{2} \right]. \quad (2.7)$$

Ici, λ est la fraction de supraconducteur, J_c est la densité de courant critique et $c(i)$ est un paramètre variant avec le courant selon l'équation non-linéaire (2.8).

$$\left[c / (4 - c^2)^{1/2} \right] \arctan \left[c (4 - c^2)^{1/2} / (2 - c^2) \right] + 2 \ln c = \ln[i / (1 - i)]. \quad (2.8)$$

Avec l'équation 2.7, l'équation 2.8, les paramètres géométriques et les propriétés des matériaux, il est possible d'obtenir la vitesse de propagation de la zone normale dans un ruban

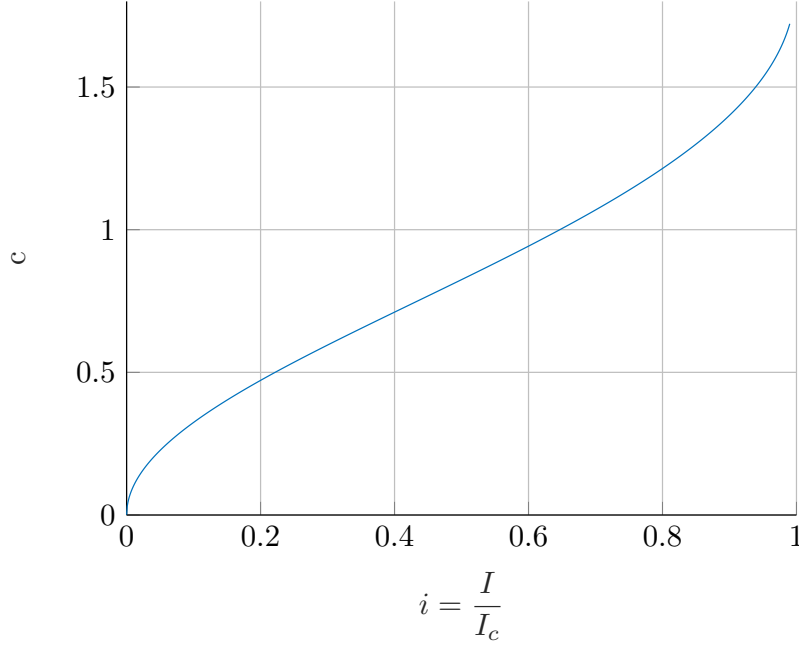


FIGURE 2.6 Valeur de la variable c permettant de résoudre numériquement l'équation 2.8 pour les différentes valeurs de courant réduit.

supraconducteur qui prend en compte le partage de courant.

Ces équations permettent de résoudre le problème pour différentes quantités de stabilisateur et de supraconducteurs, ainsi que pour différentes valeurs de propriétés des matériaux et de courant. L'architecture du ruban n'est pas prise en compte dans le modèle de Dresner, mais il tient compte de la fraction de stabilisateur f et de la fraction de supraconducteur λ . De plus, les propriétés des matériaux utilisés dans l'équation (2.7) sont les termes effectifs qui dépendent de la quantité de stabilisateur utilisée. L'équation (2.7) peut être réécrite en fonction des paramètres variables h_{Ag} et i :

$$v(h_{Ag}, i) = \left(\frac{2\lambda(h_{Ag})J_c}{S_{eff}(h_{Ag})} \right) \left(\frac{\rho_{Ag}k_{eff}(h_{Ag})}{f(h_{Ag})\tau_c} \right)^{1/2} \left[\frac{c(i)i^{1/2}}{2} \right], \quad (2.9)$$

$$\lambda(h_{Ag}) = \frac{h_{YBCO}}{h_{Ag} + h_{YBCO} + h_{Hast}}, \quad (2.10)$$

$$f(h_{Ag}) = \frac{h_{Ag}}{h_{Ag} + h_{YBCO} + h_{Hast}}, \quad (2.11)$$

$$k_{eff}(h_{Ag}) = \frac{k_{Ag}h_{Ag} + k_{YBCO}h_{YBCO} + k_{Hast}h_{Hast}}{h_{Ag} + h_{YBCO} + h_{Hast}}, \quad (2.12)$$

$$S_{eff}(h_{Ag}) = \frac{S_{Ag}h_{Ag} + S_{YBCO}h_{YBCO} + S_{Hast}h_{Hast}}{h_{Ag} + h_{YBCO} + h_{Hast}}. \quad (2.13)$$

La quantité de stabilisateur impact plusieurs éléments de l'équation (2.9). Les figures 2.6 et 2.7 et les équations (2.10), (2.11), (2.12) et (2.13) montrent comment le courant et l'épaisseur de stabilisateur impactent la NZPV. Il est à noter que la capacité calorifique S de l'équation (2.13) s'exprime en fonction de la capacité thermique ainsi que de la masse volumique ($S = C_p\rho_m$). Ainsi, $S_{Ag} = C_{p,Ag}\rho_{m,Ag}$ et pareillement pour S_{YBCO} et S_{Hast} . Dans ces équations, la dépendance en température des matériaux n'est pas considérée, la valeur de chacune des propriétés est prise à 90 K, soit la température critique du supraconducteur. Le choix de cette température est fait puisque le calcul de la NZPV pour des simulations est fait à cette température.

2.3.2 Limitation du modèle de Dresner

Le modèle de la NZPV est très utile, mais il ne permet pas de prendre en considération plusieurs éléments importants à la conception de fils supraconducteur. Par exemple, ce modèle

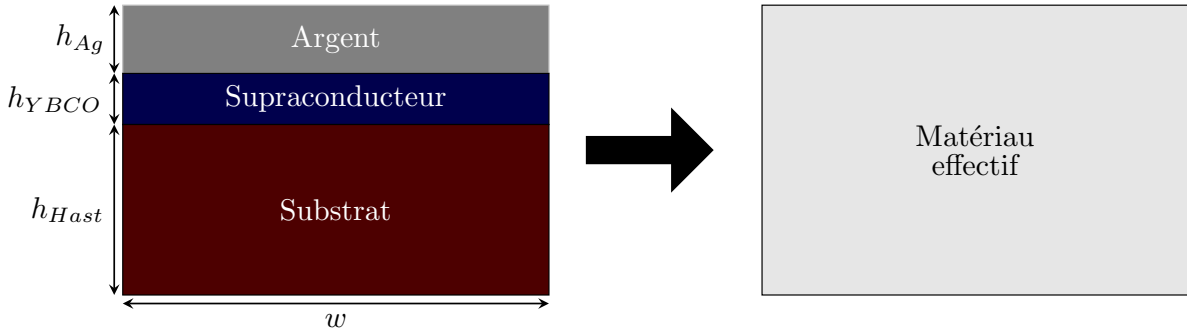


FIGURE 2.7 Schéma de la coupe transversale d'un ruban supraconducteur à gauche et d'un matériau avec des propriétés électriques et thermiques effectives à droite.

ne prend pas en compte la dépendance avec la température des propriétés des matériaux. Il ne prend pas non plus en considération l'architecture du ruban et il ne peut donc pas prédire l'effet d'une résistance d'interface entre le supraconducteur et le stabilisateur ni les effets de l'architecture CFD montrés à la figure 2.5. Néanmoins, le modèle de Dresner démontre une reproduction satisfaisante de la physique inhérente au problème, ce qui lui permet de représenter les tendances observées. Ainsi, il devient une ressource pertinente pour générer efficacement des ensembles de données destinés à l'expérimentation de différentes techniques et de divers modèles d'apprentissage automatique.

CHAPITRE 3 MÉTHODES DE RÉGRESSION ET MODÈLES D'APPRENTISSAGE MACHINE

La conception d'un modèle prédictif de la vitesse de propagation de la zone normale dans un 2G HTS CC se fait en plusieurs étapes. Il faut générer des données, entraîner un modèle, valider la qualité des modèles et réentraîner au besoin. Pour obtenir un modèle avec une précision donnée, il faut trouver un bon modèle et il faut suffisamment de données pour l'entraîner.

3.1 Modèles de régression

Les modèles de régression sont utiles en physique expérimentale et en science de l'information puisqu'il est très rare qu'il existe d'avoir un modèle analytique pour des problèmes réalistes en pratique. Il faut souvent se fier à un nombre limité de données. L'objectif des modèles de régression est d'établir une relation entre des données d'entrée ($\mathbf{X}$) et des données de sortie (y) en utilisant une base de données constituée d'une liste de paire $(\mathbf{X}, y)$. Pour faire un modèle de régression, il faut déterminer les variables indépendantes et étiqueter les variables dépendantes. Il faut ensuite trouver un modèle permettant d'établir une relation entre ceux-ci en identifiant les paramètres (qui sont nos degrés de liberté) et on calcule l'erreur afin d'estimer la qualité du modèle. Cette section présente quelques modèles permettant de mettre en lien les variables indépendantes et dépendantes.

Régression polynomiale

La régression polynomiale est une méthode qui suppose une relation polynomiale entre les n variables indépendantes $\mathbf{X}$ et la variable dépendante y . Ces variables sont reliées avec les coefficients réels $\beta_{i,j}$ selon l'équation suivante :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{1,1} x_1^2 + \beta_{2,2} x_2^2 + \beta_{1,2} x_1 x_2 + \varepsilon. \quad (3.1)$$

Cette équation prend en compte 2 variables indépendantes x_1 et x_2 , un polynôme de degré 2 et un résidu ε . L'équation (3.1) peut être écrite sous forme vectorielle comme :

$$\begin{aligned}
y &= \boldsymbol{\beta} \mathbf{X} + \varepsilon, \\
\boldsymbol{\beta} &= [\beta_0 \quad \beta_1 \quad \beta_2 \quad \beta_{1,1} \quad \beta_{2,2} \quad \beta_{1,2}], \\
\mathbf{X} &= \begin{bmatrix} 1 & x_1 & x_2 & x_1^2 & x_2^2 & x_1 x_2 \end{bmatrix}'.
\end{aligned} \tag{3.2}$$

Cette méthode se généralise facilement à plusieurs variables indépendantes et à n'importe quel degré. Il faut trouver quelles valeurs de $\boldsymbol{\beta}$ minimise l'erreur sur les données. Cette méthode utilise souvent la minimisation de l'erreur quadratique moyenne pour déterminer les valeurs de $\boldsymbol{\beta}$. Il est important de commencer avec des degrés plus petits pour minimiser la possibilité de sur-entraînement et de maximiser la généralisation. Les performances de cette méthode peuvent être médiocres si les variables sont corrélées. En effet, lorsqu'il y a beaucoup de variables, il a de fortes chances que certaines soient interdépendantes. Cette méthode peut donc être plus difficile à généraliser à plus de paramètres d'entrées.

Régression par vecteur de support

La méthode ε -SVR [33] cherche à déterminer un hyperplan qui minimise la déviation des données d'entraînement par rapport aux valeurs à prédire en fixant une limite sur cette déviation avec le paramètre ε . Cette méthode est avantageuse dans la mesure où elle est résistante aux données aberrantes et peut être facilement généralisée en utilisant différents paramètres ε . Pour modéliser des relations non linéaires entre les variables, la méthode ε -SVR utilise un noyau ϕ pour transformer l'espace original des variables en un espace de plus grande dimension où la séparation linéaire des données est possible. L'équation (3.3) montre la relation entre la prédiction de la valeur à prédire $f(\mathbf{X})$ et les données d'entrées $\mathbf{X}$:

$$f(\mathbf{X}) = (w \cdot \phi(\mathbf{X})) + b. \tag{3.3}$$

Ainsi, la méthode ε -SVR est capable de capturer des relations complexes et non linéaires entre les variables en ajustant les paramètres w et b pour minimiser l'erreur sur les données. Les noyaux les plus utilisés sont le noyau linéaire, polynomial et les noyaux gaussiens.

Processus gaussien

La régression par processus gaussien est un modèle statistique utilisé pour prédire des valeurs à partir de données d'entraînement. Le modèle est construit en spécifiant une distribution de fonctions gaussiennes, dont les paramètres sont mis à jour en utilisant les données d'en-

traînement. La prédiction sur de nouvelles données se fait en utilisant leur covariance avec les données d'entraînement, calculée à l'aide d'un noyau définissant la similarité entre les points de l'espace des prédicteurs. Les noyaux les plus couramment utilisés en régression par processus gaussien sont les fonctions à base radiale, les exponentielles quadratiques, le noyau matérn, le noyau linéaire et le noyau périodique [34].

Réseau de neurones

Un réseau de neurones est un modèle mathématique inspiré du cerveau humain et des connexions existantes entre les neurones de celui-ci. Ils sont composés de trois types de couches, soit la couche d'entrée, les couches cachées et la couche de sortie. Chaque couche est constituée de neurones qui sont des fonctions linéaires avec une fonction d'activation non-linéaire. Les neurones d'une couche sont reliés aux neurones de la couche précédente et de la couche suivante. La figure 3.1 montre un schéma d'un réseau de neurones avec chaque connexion entre les couches. Le modèle mathématique d'un neurone est exprimé comme suit :

$$z_i^{(j+1)} = \sum_{m=1}^M w_m a_m^{(j)} + b_i^{(j)}, \quad (3.4)$$

$$a_i^{(j+1)} = \text{act} \left(z_i^{(j+1)} \right). \quad (3.5)$$

La fonction $\text{act}(z)$ est une fonction d'activation qui permet au réseau d'exprimer des fonctions non-linéaires. Les fonctions d'activation les plus populaires sont $\text{ReLU}(z)$, $\tanh(z)$ ou $\text{sigmoid}(z)$. Il n'existe pas de processus permettant de déterminer quelle fonction d'activation choisir pour un certain problème, il faut les essayer et voir laquelle est la plus adaptée.

Il y a plusieurs raisons expliquant la popularité des réseaux de neurones. La plus importante est qu'un réseau de neurones possède la qualité d'approximation universelle, c'est-à-dire qu'il est possible d'approximer n'importe quelle fonction continue avec n'importe quelle précision si le réseau possède suffisamment de neurones [35]. Il est donc possible d'exprimer la relation entre des variables, peu importe la complexité du problème. Les réseaux de neurones peuvent donc être appliqués à un grand nombre de problèmes complexes et non-linéaires. Un autre avantage des réseaux de neurones est que la prédiction et l'entraînement peuvent être faits avec des multiplications matricielles, ce qui rend le procédé plus rapide. La propagation du réseau de la figure (3.1) peut s'écrire sous forme matricielle comme suit :

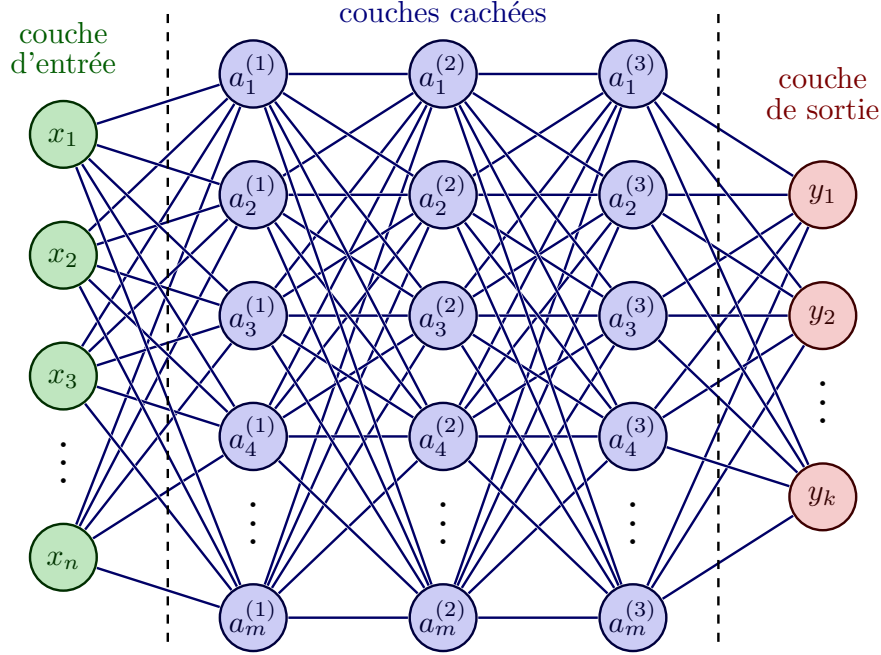


FIGURE 3.1 Réseau de neurones avec n paramètre d'entrée, 3 couches cachées avec m neurones par couche et k sorties. Chaque ligne est une multiplication entre le paramètre de la couche précédente et un poids w à déterminer.

$$\begin{aligned}
 \mathbf{y} &= \sigma \left(\mathbf{W}^{(4)} \mathbf{a}^{(3)} + \mathbf{b}^{(4)} \right), \\
 \mathbf{a}^{(3)} &= \text{act} \left(\mathbf{W}^{(3)} \mathbf{a}^{(2)} + \mathbf{b}^{(3)} \right), \\
 &\vdots, \\
 \mathbf{a}^{(1)} &= \text{act} \left(\mathbf{W}^{(1)} \mathbf{x} + \mathbf{b}^{(1)} \right),
 \end{aligned}$$

avec $\mathbf{W}^{(i)}$ les matrices de poids, $\mathbf{b}^{(i)}$ des vecteurs de biais, $\mathbf{y}$ les valeurs à prédire et $\mathbf{X}$ les entrées du réseau. La propagation et l'entraînement reviennent donc à faire des opérations sur des matrices, ce qui se fait rapidement sur une carte graphique.

3.2 Ensemble de modèles

Les ensembles de modèles ou les comités de modèles en apprentissage machine consistent à combiner les prédictions de plusieurs modèles afin d'améliorer leurs performances. Ils peuvent être utilisés pour faire de la classification ou pour faire de la régression et ils sont connus

pour augmenter la précision et la robustesse des prédictions [36].

Il existe plusieurs types d'ensemble de modèles comme le *Bagging* (*Bootstrap Aggregating*), qui consiste à entraîner plusieurs modèles sur des sous-ensembles des données d'entraînement qui peuvent se superposer, puis à utiliser la moyenne des prédictions comme valeur finale [37]. Cette méthode aide à réduire la variance des prédictions surtout lorsque les sous-modèles peuvent être sur-entraînés (*overfitting*). Cette méthode possède l'avantage que les sous-modèles peuvent être entraînés en parallèle, ce qui réduit le temps de calcul. Une autre technique des ensembles de modèles est le *Boosting*. Cette technique itérative consiste à faire en sorte que les poids des données d'entraînement ayant les pires performances sur le modèle précédent soit ajustés de manière à ce qu'ils aient plus d'importance pour les prochains modèles à entraîner. Ces prochains modèles sont axés sur les cas problématiques de la base de données, ce qui réduit les biais des prédictions. Un autre modèle d'ensemble est le *Random forest*. Cette technique combine plusieurs arbres de décisions et les joint de manière à obtenir de meilleures prédictions. Le terme *forest* provient de l'utilisation de plusieurs arbres de décisions pour obtenir une prédiction. Cette méthode est très utile pour des problèmes avec beaucoup de paramètres variables. [38]

Les ensembles de modèles ont plusieurs avantages sur les modèles uniques, le plus important étant la réduction de la variance sur les prédictions. En combinant la prédiction de plusieurs modèles, l'ensemble devient plus robuste et précis que chaque modèle individuel. De plus, il réduit le biais présent dans les modèles sur-entraînés et il réduit l'impact des données bruitées. Un des problèmes avec les ensembles de modèles est qu'ils sont plus coûteux en calcul puisqu'il faut entraîner plusieurs modèles. La figure 3.2 montre un schéma montrant comment les données traversent l'ensemble de modèle afin de produire des prédictions.

3.3 Apprentissage actif

L'apprentissage actif est basé sur le principe de permettre au modèle de sélectionner les données à ajouter à la base de données utilisée pour l'entraînement. Bien que l'ajout de données soit une méthode efficace pour améliorer les performances d'un modèle, toutes les données ne se valent pas. Certaines données sont plus utiles que d'autres et peuvent apporter une information plus précise au système. Certaines zones de l'espace échantillons peuvent être monotones et linéaires et d'autres peuvent être non-linéaires et irrégulières. Les zones irrégulières doivent être plus échantillonnées pour mieux capter la dynamique réelle de la problématique.

L'apprentissage actif est aussi appelé la conception expérimentale optimale (*Optimal Expe-*

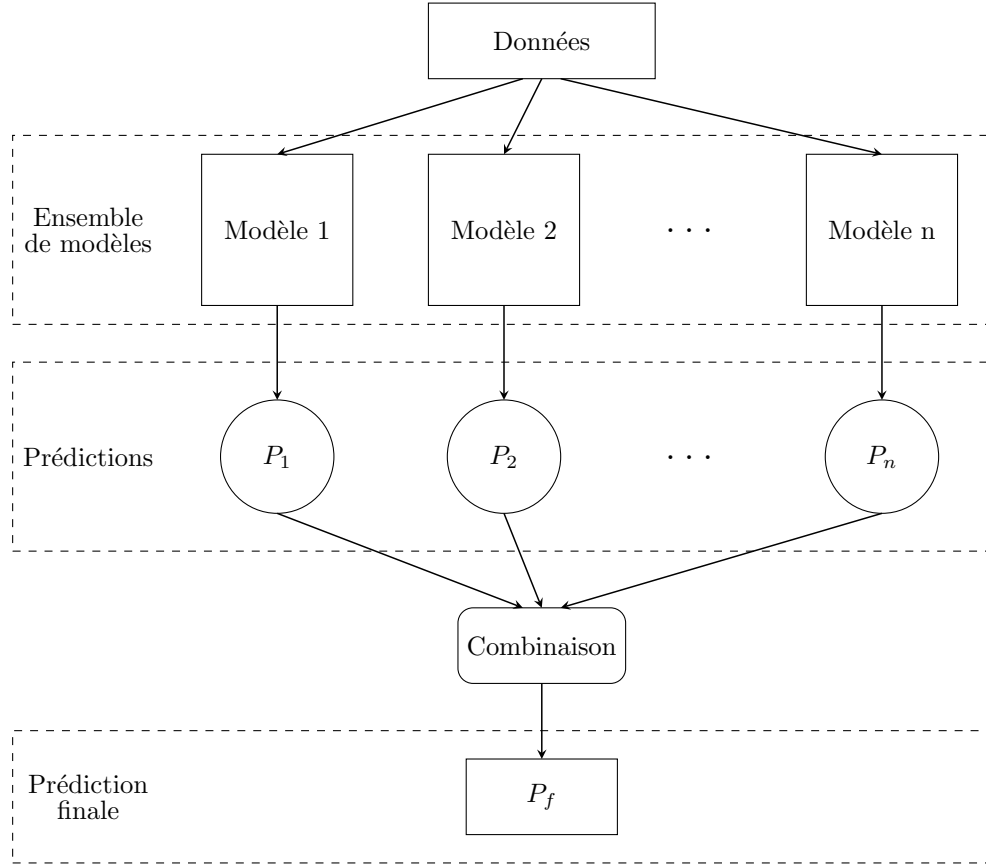


FIGURE 3.2 Schéma du principe d'un ensemble de modèles.

perimental Design) [39]. On y fait référence lorsque le choix des données est fait à partir de l'incertitude du modèle de prédiction. Pour augmenter la précision du modèle, les données à ajouter sont choisies à partir de l'incertitude et des erreurs du modèle. L'apprentissage actif est important puisque si la base de données d'entraînement est mal générée ou s'il n'y a pas suffisamment de données, le modèle devient susceptible de ne pas être généralisable et d'avoir une mauvaise performance sur certaines régions de l'espace échantillon. À la figure 3.3, la recherche des données à générer trouve les zones de l'espace échantillons des entrées qui ont des performances médiocres. Ce bloc trouve de nouvelles entrées $\mathbf{X}'$ à étiqueter pour augmenter les performances du modèle. La figure 3.3 présente un schéma des étapes du processus d'apprentissage actif.

Ce type d'entraînement aide également à trouver des valeurs aberrantes. Lorsqu'une donnée ne présente pas un comportement similaire aux autres, le modèle a de la difficulté à prédire cette variable. Après l'entraînement sur le nouveau jeu de données, le modèle offre de meilleures performances dans cette région malgré la donnée aberrante. Une méthode importante d'apprentissage actif est la requête par comité (*Query-by-Committee*). Cette méthode

se base sur les ensembles de modèles et sur la différence entre les prédictions de ceux-ci sur une même donnée à prédire. Si les modèles prédisent tous la même valeur, la variance de celle-ci est nul et le modèle performe bien. Si les prédictions sont très différentes, alors la variance sera grande car le modèle est inefficace près de cette donnée d'entrée $\mathbf{X}$ [40], il est donc mieux d'ajouter une donnée d'entraînement près de ce prédicteur. La variance des modèles est aussi appelé l'erreur généralisée puisque la distribution d'erreur est relié à la variance du modèle d'ensemble.

3.3.1 Apprentissage actif pour la prédiction de la NZPV

La valeur de la vitesse de propagation de la zone normale v_n d'un ruban supraconducteur est influencée par plusieurs paramètres comme la géométrie, la température, la composition atomique des différentes couches et le courant. La conception d'un modèle prédictif de la NZPV doit pouvoir apprendre l'effet de chaque paramètre sur celle-ci. Pour le cas simplifié de Dresner, on s'intéresse à l'effet de l'épaisseur du stabilisateur h_{Ag} et de la fraction du courant critique i sur la NZPV obtenu avec l'équation (2.7). La section 4.1 explique précisément comment les données sont générées avec le modèle analytique de la NZPV de Dresner.

La génération de la base de données initiale est importante puisque chaque simulation est associée à un coût en temps de calcul. Il faut donc choisir judicieusement quelle donnée acquérir pour générer les données d'entraînement. L'épaisseur du stabilisateur h_{Ag} prend des valeurs dans l'intervalle $[1\mu\text{m}, 5\mu\text{m}]$ et le courant réduit i est dans l'intervalle $[0, 1]$. Les prédicteurs prennent des valeurs sur des intervalles différents. Pour la base de données initiale, il y a deux aspects pris en considération guider l'échantillonnage des données. Premièrement, la frontière de l'espace échantillons des entrées $\mathbf{X}$ doit faire partie de la base de données d'entraînement, spécifiquement les valeurs extrêmes. Cette règle est importante, car les modèles prédictifs sont toujours meilleurs pour interpoler que pour extrapoler et cela permet de limiter les prédictions à des valeurs raisonnables. Deuxièmement, il faut échantillonner davantage les régions ayant une plus grande variation de la valeur à prédire. La seconde règle est utile, car les grandes variations de la valeur apportent une plus grande incertitude. En échantillonnant davantage, on obtient une meilleure représentation de la distribution de la NZPV selon les paramètres et la précision du modèle devient meilleure. Dans un cas où il n'y a aucune information *a priori* sur la distribution de la valeur à prédire, il est mieux de prendre une distribution uniforme des indicateurs.

Pour éviter le sur-ajustement d'un modèle de régression, il est nécessaire d'avoir des données pour la validation. Il est commun d'utiliser aléatoirement des données parmi le lot et de les utiliser pour éviter le sur-entraînement. Il y a un problème avec cette méthode lorsqu'il y a

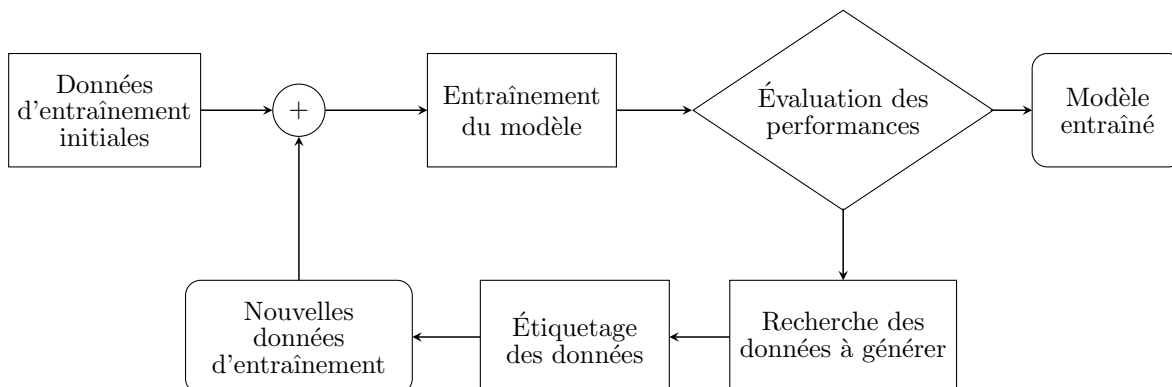


FIGURE 3.3 Schéma d'une boucle d'apprentissage actif.

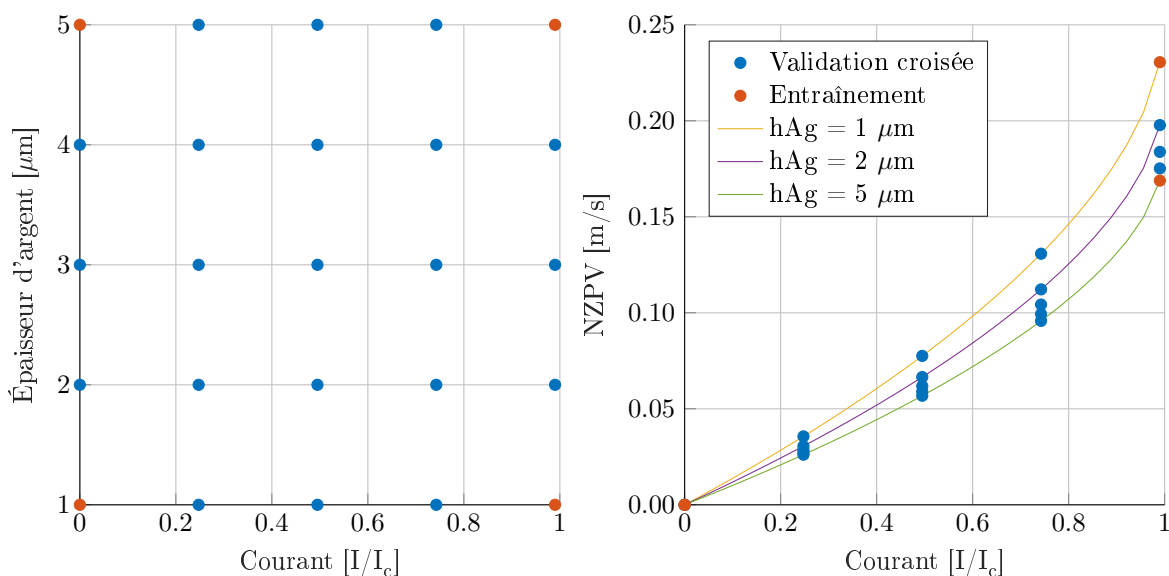


FIGURE 3.4 Données d'entrée pour la prédiction de la NZPV. Le graphique de gauche montre la distribution uniforme de 25 points dans l'espace des entrées pour l'entraînement, dont les 4 données aux extrêmes sont strictement pour l'entraînement. Le graphique de droite compare la prédiction de la NZPV en fonction du courant pour différentes épaisseurs de stabilisateur. Les lignes pleines représentent les distributions de NZPV du modèle de Dresner et les cercles représentent les points échantillonnés.

peu de données : si un point aux extrémités est pris, alors le modèle fera une extrapolation pour cette zone, ce qui réduit grandement la qualité du modèle [41]. Pour les petites bases de données, il est recommandé d'utiliser la validation croisée (*cross-validation*) pour utiliser les données au maximum. La base de données est divisée en k sous-groupes et le modèle est entraîné avec $k - 1$ des sous-groupes. Les performances sont validées avec le sous-groupe restant. Cet exercice est répété k fois pour avoir entraîné et valider avec chacun des sous-groupes. Cette méthode nécessite plus de calcul, car le modèle doit être entraîné k fois, mais offre une meilleure estimation de la précision du modèle de prédiction. La figure 3.4 montre un choix des données d'entraînement et de validation permettant d'avoir toujours les valeurs extrêmes. Ce choix permet de restreindre les prédictions des modèles entraînés sur ces données à de l'interpolation.

Le choix des données à générer est fait par requête par comité, c'est-à-dire que la variation de prédiction des différents modèles d'un ensemble permet de définir des régions qui nécessitent plus de données. Les modèles utilisés sont des réseaux de neurones pour les raisons suivantes :

1. Les réseaux de neurones sont facilement généralisables à un plus grand nombre de paramètres d'entrée. Il suffit d'augmenter le nombre de neurones de la première couche.
2. Il y a plusieurs ensembles de poids et de biais qui conduisent à des performances de prédictions similaires lors de l'entraînement d'un réseau de neurones. Il est donc possible d'utiliser toutes les données pour l'entraînement de tous les réseaux de neurones.

3.3.2 Génération des données

Il y a deux types de données qui sont générées dans ce travail. Les données utilisées pour déterminer le modèle le plus approprié et les données pour le modèle réel de prédiction de la NZPV. Les premières données doivent être faciles à générer et elles doivent avoir un comportement similaire au comportement réel de la NZPV d'un ruban supraconducteur. Le modèle Dresner de la NZPV répond à ces exigences. Le deuxième type de donnée à générer est la simulation du comportement électrothermique d'un ruban supraconducteur avec un courant imposé sur celui-ci. Ces données sont longues à générer, mais il s'agit des données nécessaires à la résolution de la problématique. Ces données sont générées par éléments finis avec le logiciel COMSOL. Lors de la génération des données, il faut choisir quels paramètres doivent être étudiés. Les paramètres sur lesquels il est possible d'effectuer une conception plus efficace de la NZPV des rubans supraconducteurs sont souvent liés au courant et à la quantité de stabilisateur. Il faut donc faire plusieurs simulations en fonction de ces paramètres pour obtenir une base de données.

3.3.3 Algorithme

Algorithme 1

Avec les données suivantes :

T – Données d’entraînement

U – Données non-étiquetées

x_j – Donnée j de l’ensemble U avec $j \in [1, N]$

$sModel$ – Sous-modèles

k – Nombre maximal d’itérations

m – Nombre de données étiquetées par itération

1. **Repète** k fois :
 2. Entraîne un ensemble de modèle avec T ,

$$E^* = EnsembleMethode(sModel, T)$$
 3. Calcul $STD(E^*, \mathbf{x}_j) \forall$ les données non-étiquetées, $\mathbf{x}_j \in U$
 4. Sélectionne un sous-ensemble S de m données non-étiquetées qui maximise l’utilité
 5. Étiquette les donnés dans S
 6. Retire les données de S de U et ajoute-les à T
 7. **Retourne** $EnsembleMethod(sModel, T)$
-

L’algorithme d’apprentissage actif par requête de comité nécessite un jeu de données d’entraînement initial T , des données x_j non-étiquetées dans l’ensemble U et un ensemble de modèles. L’algorithme est une boucle d’entraînement et d’ajout de données. À la ligne 2 du pseudocode, les sous-modèles de l’ensemble sont entraînés avec les données d’entraînement T . À l’étape 3, l’écart-type des prédictions est calculé sur toutes les données dans l’ensemble U . À la ligne 4, on sélectionne un ensemble S de m données à étiqueter. L’échantillonnage est fait aléatoirement en utilisant la distribution de l’écart-type comme fonction de poids. Les données de S sont ensuite étiquetées à la ligne 5. Les données de S sont ajoutées à T et retirées de U . Cette boucle est répétée k fois. L’objet que l’algorithme retourne est un ensemble de modèles entraîné sur les données T .

Cet algorithme a l’avantage de choisir automatiquement les données à étiqueter. Les données choisies sont celles qui ont le potentiel d’améliorer le plus les performances du modèle. Si le modèle ajoute une donnée à chaque itération ($m = 1$), alors la donnée à ajouter x devrait

être celle dont la prédiction donne le plus grand écart-type à l'étape 3. Si $m > 1$, alors il faut échantillonner aléatoirement U avec la fonction de poids basé sur l'écart-type. Si les données choisies à étiqueter sont toutes celles ayant le plus grand écart-type, celles-ci ont de fortes chances d'être très près les unes des autres dans l'espace échantillon des entrées, ce qui apportera moins d'information aux modèles.

3.4 Métriques d'évaluations

Le choix des métriques d'évaluations est un aspect essentiel à la conception d'un bon modèle de prédiction. L'importance d'une métrique d'évaluation provient de l'habileté à quantifier la précision et la justesse des modèles de prédictions. Cela permet également de vérifier si le modèle est généralisable pour de nouvelles données.

3.4.1 Racine de l'écart quadratique moyen

Une métrique importante est la racine carrée de l'erreur quadratique moyenne (RMSE). Cette métrique est l'erreur moyenne attendue sur n'importe quelle prédiction. Le lien entre cette valeur et la qualité du modèle est direct : un modèle avec une faible RMSE est très souvent meilleur qu'un modèle avec une plus grande RMSE. De plus, pour les réseaux de neurones, l'erreur quadratique moyenne (MSE) est très souvent utilisée pour déterminer les paramètres du modèle. Cette valeur est minimisée à chaque itération et le modèle choisi est celui avec la plus faible MSE. La RMSE est la racine carrée de la MSE et elle est plus utile, car elle a les mêmes unités que la valeur à prédire. Le calcul de la RMSE sur une fonction $f(x) = y + \epsilon$, est-ce fait comme suit :

$$RMSE(f) = \sqrt{\frac{1}{N} \sum_{i=1}^N (f(x_i) - y_i)^2} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i + \epsilon_i - y_i)^2} = \sqrt{\frac{1}{N} \sum_{i=1}^N \epsilon_i^2}, \quad (3.6)$$

avec f le modèle, x le prédicteur, y la valeur à prédire et ϵ l'erreur sur la prédiction. La racine carrée de l'erreur quadratique moyenne est utile pour comparer différents modèles, mais elle peut être adaptée pour déterminer les régions de l'espace échantillon ayant des performances moindres. Dans un problème avec M prédicteurs $\mathbf{X} = [x^{(1)}, x^{(2)}, \dots, x^{(M)}]$, dont chacun des paramètres d'entrées prend un nombre discret de valeurs $x_i^{(m)} \in [x_i^{(m)}]$, avec $i = (1, 2, \dots, K)$. Il est intéressant de comprendre comment l'erreur varie selon un paramètre d'entrée. Au lieu de faire la somme sur toutes les données, il est possible de faire des sommes partielles pour paramétrer la RMSE sur le prédicteur à étudier. Les sommes partielles sont une méthode de projections de l'erreur sur l'axe d'un des paramètres d'entrées. La figure 3.5 montre un

exemple en deux dimensions de la RMSE partielle selon les deux paramètres d'entrée $x^{(1)}$ et $x^{(2)}$.

Pour le cas où $x^{(1)}$ prend K valeurs différentes, la RMSE partielle est calculé sur chaque valeurs $x_k^{(1)}$ en sommant sur tous les autres dimensions pour chaque k .

$$RMSE(f, x_k^{(1)}) = \sqrt{\frac{1}{N_k} \sum_{x^{(2)}} \sum_{x^{(3)}} \dots \sum_{x^{(M)}} \left(f(x_k^{(1)}, x^{(2)}, \dots, x^{(M)}) - y_i \right)^2} \quad \forall k \in [1, K], \quad (3.7)$$

avec N_k le nombre de données où $x^{(1)} = x_k^{(1)}$. Faire ce travail permet d'afficher la RMSE en fonction du paramètre $x^{(1)}$. Le même développement peut être fait sur chacun des autres paramètres d'entrées. La figure 3.6 montre comment la RMSE partielle est utile pour visualiser à quel endroit dans l'espace échantillon les performances du modèle sont moindres et là où elles sont meilleures.

3.4.2 Test Gamma

Le test Gamma est un outil statistique permettant de déterminer à quel point une base de données d'entrées et de sortie $[\mathbf{x}, y]$ peut permettre de retrouver une relation lisse entre les données. Cette métrique donne une approximation de l'erreur quadratique moyenne sur le bruit de y et par le fait même, par un modèle lisse entraîné sur ces données. Le test gamma permet de calculer une estimation de la variance sur y que n'importe quel modèle lisse pourrait avoir s'il est basé uniquement du jeu de données [42].

En supposant des données de la forme $\{(\mathbf{x}_i, y_i) : 1 \leq i \leq M\}$ avec les entrées $\mathbf{x} \in \mathbb{R}^m$ et sorties $y \in \mathbb{R}$ et un modèle $f(\mathbf{x})$ tel que :

$$y = f(\mathbf{x}) + r, \quad (3.8)$$

où f est une fonction inconnue et lisse représentant la relation entre $\mathbf{x}$ et y , tandis que r représente le bruit du système. Le test gamma approxime la variance de la variable du bruit $\text{Var}(r)$. La moyenne du bruit r est nulle, car n'importe quel biais peut être inclus dans la fonction inconnue $f(\mathbf{x})$. Pour un modèle lisse, deux points $\mathbf{x}$ et $\mathbf{x}'$ prêts l'un de l'autre dans l'espace échantillon doivent aussi être près dans l'espace des prédictions $y(\mathbf{x})$ et $y' = y(\mathbf{x}')$. En regardant le cas limite où $\mathbf{x}' \rightarrow \mathbf{x}$ et en prenant l'espérance, on retrouve le résultat suivant :

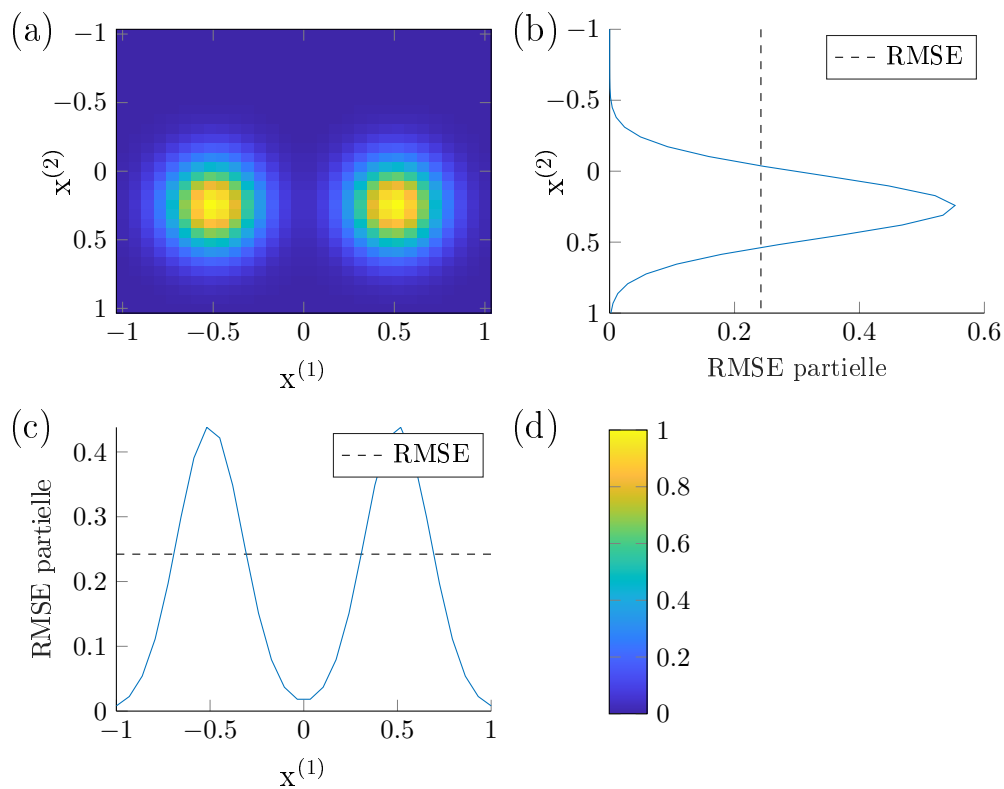


FIGURE 3.5 Exemple de la racine carrée de l'erreur quadratique moyenne partielle de deux pics gaussiens en deux dimensions. La figure (a) montre la distribution de l'erreur selon les deux paramètres $x^{(1)}$ et $x^{(2)}$, la distribution de la RMSE partielle selon (b) $x^{(2)}$ et (c) $x^{(1)}$ ainsi que (d) l'échelle d'intensité de l'erreur.

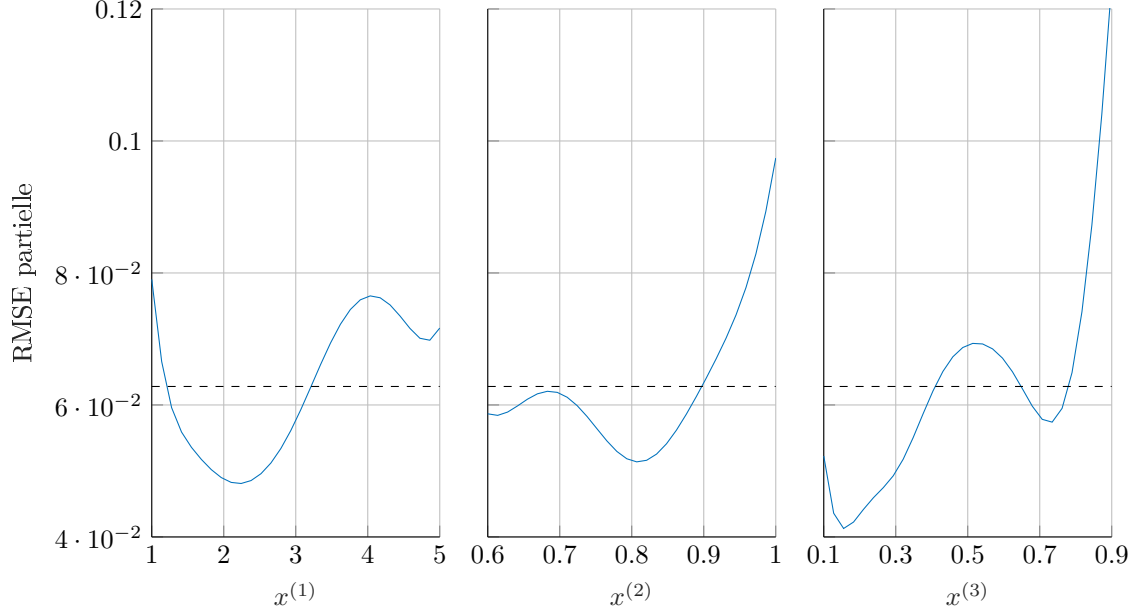


FIGURE 3.6 Distribution de la RMSE partielle sur la prédiction selon les différents paramètres x^i . La ligne pointillée est la vraie RMSE.

$$\mathbb{E} \left(\frac{1}{2} (y - y')^2 \right) \rightarrow \text{Var}(r). \quad (3.9)$$

En d'autres mots, le carré de la différence entre une donnée y et son voisin le plus près y' tend vers la variance du bruit lorsque les paramètres x et x' sont près. Ce résultat n'est cependant pas atteignable à cause de la quantité limitée de données disponibles. Pour estimer cette limite, on observe la distance des k premier voisin de chaque entrée $\mathbf{x}_i$ et on les annote $\mathbf{x}_{N[i,k]}$. Il faut ensuite calculer la distance euclidienne moyenne entre chaque valeur de sortie y_i et les k $y_{N[i,k]}$ associés au voisin de $\mathbf{x}_i$. La valeur moyenne pour k est notée $\gamma_M(k)$ pour y et $\delta_M(k)$ pour $\mathbf{x}$. Les équations 3.10 et 3.11 montrent comment ce calcul est fait :

$$\delta_M(k) = \frac{1}{2M} \sum_{i=1}^M \left| \mathbf{x}_{N[i,k]} - \mathbf{x}_i \right|^2, \quad (3.10)$$

$$\gamma_M(k) = \frac{1}{2M} \sum_{i=1}^M \left| y_{N[i,k]} - y_i \right|^2. \quad (3.11)$$

Le test Gamma est basé sur l'affirmation que la relation entre $\gamma_M(k)$ et $\delta_M(k)$ est environ linéaire près de $\delta_M(k) = 0$. La relation entre ces variables est $\gamma_M(k) \approx A\delta_M(k) + \Gamma$. La figure

3.7 montre un exemple typique de cette relation. Comme pour l'équation (3.9), lorsque les premiers voisins tendent vers le x associés, alors $\delta_M(k) \rightarrow 0$ et $\gamma_M(k) \rightarrow \Gamma = \text{Var}(r)$.

Le test Gamma offre plusieurs perspectives sur les jeux de données et sur les modèles entraînés sur ceux-ci. Il permet de déterminer s'il y a suffisamment de données, si elles sont de qualité suffisante pour obtenir un modèle avec une précision donnée. C'est également un outil permettant d'investiguer si toutes les variables pouvant affecter la sortie y sont considérées ou si les variables utilisées sont toutes importantes.

Une autre utilité du test gamma est qu'il peut servir de critère de précision lors de l'entraînement des modèles. Le modèle arrête d'être adapté lorsque la MSE sur les données d'entraînement est égale au Γ . Un modèle ayant une MSE plus faible que la variance du bruit des données est à risque d'être sur-entraîné et d'avoir des performances inférieures sur les données de validation et de test.

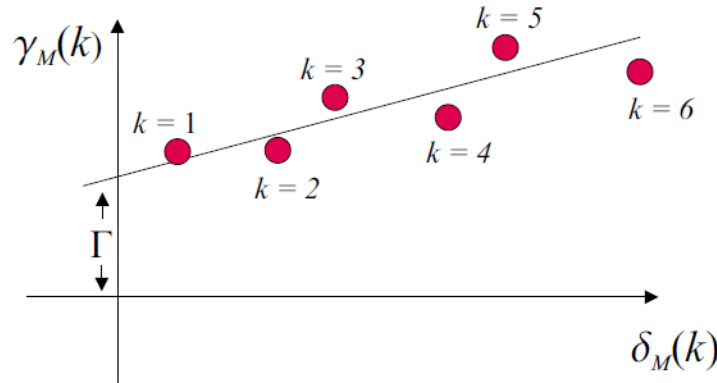


FIGURE 3.7 Régression du test Gamma [43].

CHAPITRE 4 CHOIX DE LA MÉTHODOLOGIE : UNE APPROCHE PRATIQUE

Dans cette section, on tente de répondre à deux questions : 1) Comment construire un jeu de données avec une quantité minimale de données fournissant suffisamment d'information pour entraîner un modèle de régression ? 2) Quel modèle est le plus efficace pour prédire la dynamique et la vitesse de propagation de la zone normale dans des rubans supraconducteurs ?

Les données utilisées pour répondre à ces questions sont générées à partir du modèle de Dresner, puisqu'il décrit une partie de la dynamique électrothermique qui nous intéresse et qu'il est beaucoup plus rapide de générer des données avec ce modèle qu'avec des simulations électrothermiques. En effet, avec ce modèle, il est possible de générer un jeu de données d'un million de données en environ 5 secondes. C'est beaucoup plus rapide qu'une simulation électrothermique de ruban supraconducteur par éléments finis, qui peut prendre entre 20 minutes et 4 heures pour générer une seule donnée.

4.1 Données d'entraînement - modèle de Dresner

Le modèle analytique de vitesse de propagation de la zone normale de Dresner est dérivé d'un problème de l'équation de la chaleur en une dimension avec une source de chaleur par effet Joule. Celui-ci permet de générer des données ayant un comportement similaire à celle des simulations éléments finis 3D. Cela permet de tester quel type de modèle peut apprendre la dynamique de la NZPV en utilisant un minimum de données. Les variables utilisées pour prédire la NZPV sont la fraction du courant critique injecté $i = I/I_c$ et l'épaisseur d'argent h_{Ag} . Un schéma de la géométrie utilisé est présenté à la figure 2.7. Les équations (2.8), (2.9), (2.10), (2.11), (2.12) et (2.13) sont résolus pour différentes valeurs de h_{Ag} et i , ainsi que les propriétés des matériaux présentées au tableau 4.1.

Les propriétés des matériaux et les paramètres des tableaux 4.1 et 4.2 permettent de résoudre l'équation 2.9 pour différentes épaisseurs de stabilisateur et différents courants. Le courant critique est calculé avec la densité de courant critique J_c et la surface de la coupe transversale du supraconducteur $I = J_c \times w \times h_{YBCO} = 100$ A. Le courant prend une valeur dans l'intervalle $[0, 100]$ A et le courant réduit (i) prend donc des valeurs entre $[0,1]$. La figure (4.1) montre le comportement de la NZPV pour différents paramètres.

TABLEAU 4.1 Propriétés des matériaux d'un ruban supraconducteur refroidi à 90 K

Matériau	Capacité thermique C_p [J/kgK]	Masse volumique ρ_m [kg/m ³]	Conductivité thermique k [W/m*K]	Résistivité électrique ρ [$\Omega \cdot m$]
YBaCuO	185.965	6390	8.5	-
Argent	177.315	10490	460	3.874e-9
Hastelloy	205.594	8890	7.97	-

TABLEAU 4.2 Paramètres du système

Propriété	Symbole	Valeur	Unité
Température critique	T_c	90	K
Température ambiante	T_b	77	K
Épaisseur du substrat	h_{Hast}	50	μm
Épaisseur du supraconducteur	h_{YBCO}	1	μm
Largeur du ruban	w	4	mm
Densité de courant critique	J_c	2.5e10	A/m ²

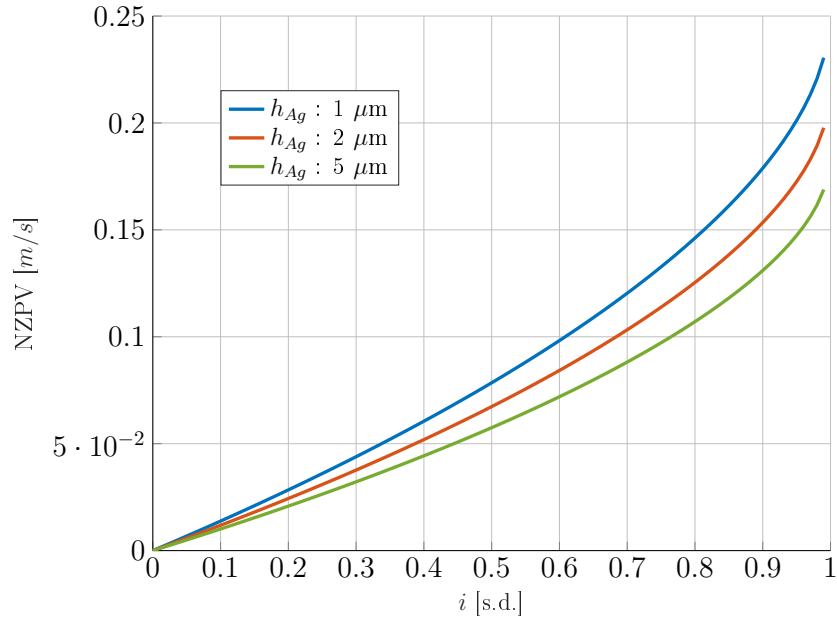


FIGURE 4.1 Distribution de la vitesse de propagation de la zone normale selon le courant du modèle de Dresner pour différente épaisseur d'argent.

4.1.1 Évaluation de la qualité des données

Le jeu de données est un aspect essentiel à la conception de modèle de prédiction. Plusieurs éléments ont une influence sur la qualité d'un jeu de données. Un premier point important est que les valeurs extrêmes du système soient dans le jeu de données. Cela permet au modèle de prédiction entraîné sur ces données de faire de l'interpolation et non de l'extrapolation. Il faut également que les zones où les variations de la NZPV sont plus grandes soient échantillonnées à plus forte densité. Ce deuxième point n'est pas facilement comblé s'il n'y a pas de connaissance au préalable sur le système étudié.

Deux méthodes sont utilisées ici pour évaluer numériquement la qualité d'un jeu de données : 1) le test Gamma et 2) la performance d'un modèle entraîné sur ces données. Le test Gamma est intéressant puisqu'il attribue une valeur Γ à un jeu de donnée indiquant une estimation de la variance du bruit dans les données. En pratique, un modèle dont la moyenne de l'erreur quadratique est inférieure à Γ a de forte chance d'être sur-entraîné. La figure 4.2 montre comment Γ varie en fonction du nombre de données pour un jeu de données avec bruit et sans bruit. La distribution des données est uniforme sur le courant et sur l'épaisseur de stabilisateur, le bruit suit une distribution gaussienne de moyenne nulle et de variance 10^{-4} (m/s)^2 .

Le premier élément intéressant de la figure 4.2 est que, pour le cas bruité, la valeur absolue de gamma atteint un plateau à 10^{-4} (m/s)^2 lorsqu'il y a environ 500 données. La courbe des données non-bruitées n'atteint pas de plateau et continue de diminuer avec l'ajout du nombre de données. La relation entre Γ et le nombre de donnée N n'est pas lisse, mais la relation entre ceux-ci est :

$$|\Gamma(N)| \approx 4.497N^{-0.924}. \quad (4.1)$$

La valeur de Γ est approximativement inversement proportionnelle à N , ce qui indique que la qualité de le jeu de données s'améliore à mesure que le nombre de données augmente, tel que prévu. Bien que le test gamma soit un outil utile pour évaluer la qualité d'un jeu de données, il n'est pas suffisant. Un test plus pertinent consiste à observer comment les performances d'un modèle entraîné sur ces données changent avec le nombre de données. La figure 4.3 montre la moyenne de l'erreur quadratique de la prédiction de la NZPV d'un réseau de neurones sur des données bruitées et non bruitées.

La figure 4.3 montre que les performances de prédiction d'un modèle varient selon les données utilisées pour l'entraînement. Les modèles entraînés sur les données non bruitées performant mieux que sur les données bruitées. Ce résultat est attendu puisque le bruit peut ajouter de

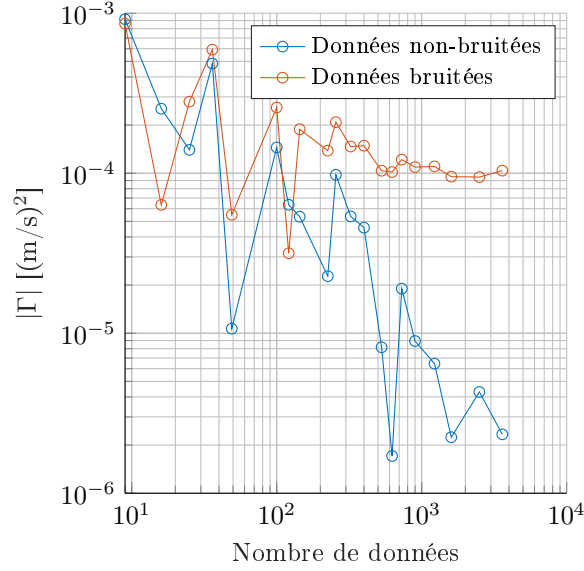


FIGURE 4.2 Évolution de l'indicateur Γ en fonction du nombre de données générées à l'aide du modèle de Dresner. Les données sans bruit sont représentées en bleu, tandis que les données avec un bruit gaussien de variance 10^{-4} (m/s)^2 sont représentées en orange.

l'information erronée, ce qui réduit la généralisabilité du modèle. Un résultat plus intéressant est que les performances sur les données de test des modèles bruités sont meilleures que celles sur les données d'entraînement et de validation. La MSE sur les données d'entraînement et de validation atteint un plateau à 10^{-4} (m/s)^2 et la MSE sur le jeu de test atteint un plateau à environ $3 \times 10^{-6} \text{ (m/s)}^2$.

Lors de l'entraînement, d'un réseau de neurones, les paramètres de celui-ci changent jusqu'à

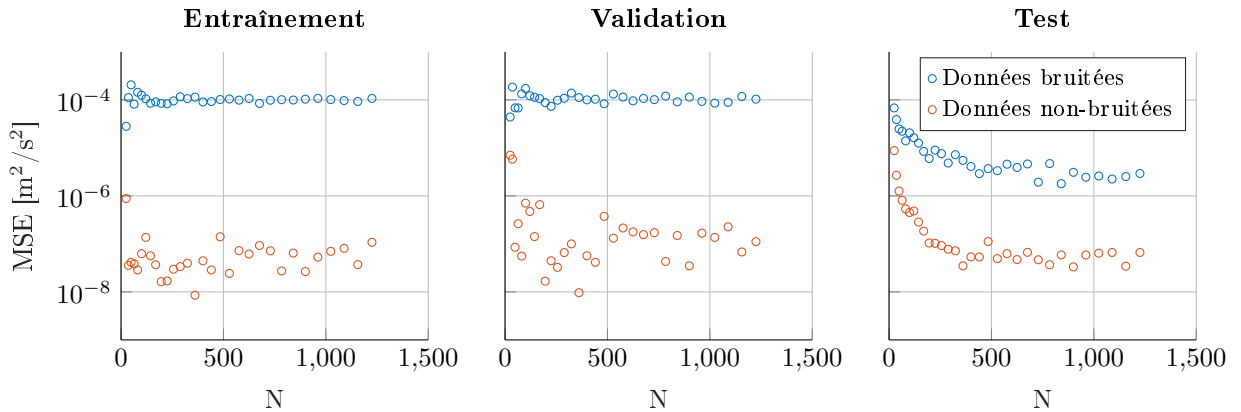


FIGURE 4.3 Effet du bruit sur la MSE de la prédiction de la NZPV d'un réseau de neurones sur les données d'entraînement, de validation et de test en fonction du nombre de données.

ce que les performances sur le jeu de validation se dégradent. Ceci est fait pour éviter de sur-entraîner le réseau de neurones. Puisqu'il y a un bruit gaussien de variance 10^{-4} (m/s)^2 , l'entraînement arrête alors à cette valeur, sans quoi le modèle apprendrait du bruit. Puisque la MSE sur le jeu de données tests est inférieure à celle d'entraînement et de validation, le modèle a appris la dynamique de la NZPV au-delà du bruit présent dans les données.

L'erreur quadratique moyenne du modèle bruité atteint un plateau à environ $2.5 \times 10^{-6} \text{ (m/s)}^2$ à 500 données, alors que le modèle entraîné avec les données non-bruité atteint un plateau de $5 \times 10^{-8} \text{ (m/s)}^2$ à 400 données. Pour le modèle sans bruit, la MSE sur les jeux d'entraînement, de validation et de test tend vers la même valeur. La raison de ce plateau est qu'on impose une limite sur l'algorithme d'ajustement des poids du réseau de neurones. L'entraînement arrête lorsque l'un des critères d'arrêt est atteint [44]. L'algorithme d'ajustement des poids utilise une descente de gradient et cet algorithme cesse l'entraînement entre autres lorsque le gradient atteint une valeur de 10^{-7} . La figure 4.4 montre comment la MSE sur la prédiction des données d'entraînement et de validation change avec les époques. On note que une époque est définit comme une itération complète de l'algorithme d'apprentissage, au cours de laquelle toutes les données d'entraînement ont été présentées au réseau de neurones.

4.2 Analyse des modèles de prédiction

Afin de déterminer quels modèles de régression permettent de bien représenter la relation entre la NZPV, l'épaisseur d'argent et le courant à partir d'un jeu de données, plusieurs modèles ont été entraînés sur un jeu de 100 données. Chaque donnée contient l'épaisseur d'argent, le courant et la NZPV associée du modèle de Dresner. Les modèles utilisés sont ceux de l'application *Regression Learner* de MATLAB, qui est un outil utile pour une première sélection des modèles et de leurs hyperparamètres ainsi que des modèles de régressions polynomiales. Les modèles requièrent souvent de déterminer plusieurs paramètres, mais trouver la combinaison de paramètres procurant les meilleures performances n'est pas une tâche triviale. Il faut entraîner les différents modèles et analyser leurs performances. Quatre familles de modèles sont présentée soit les régression par vecteur de support (SVR), la régression polynômiale, la régression par processus gaussien (GPR) et les réseaux de neurones. La figure 4.5 montre les performances de ces modèles avec différents hyper-paramètres en validation croisée entraînée sur 100 données divisées en 5 blocs de données. La validation croisée est une technique permettant de déterminer l'efficacité et la robustesse d'un modèle d'apprentissage automatique. Chaque modèle est entraîné sur 4 des 5 blocs de données et validé sur le 5e. Ce processus est répété 5 fois de façon à ce que chaque bloc ait servi de validation 1 fois. La validation croisée est la moyenne des 5 validations calculées avec les 5 entraînements. Cette

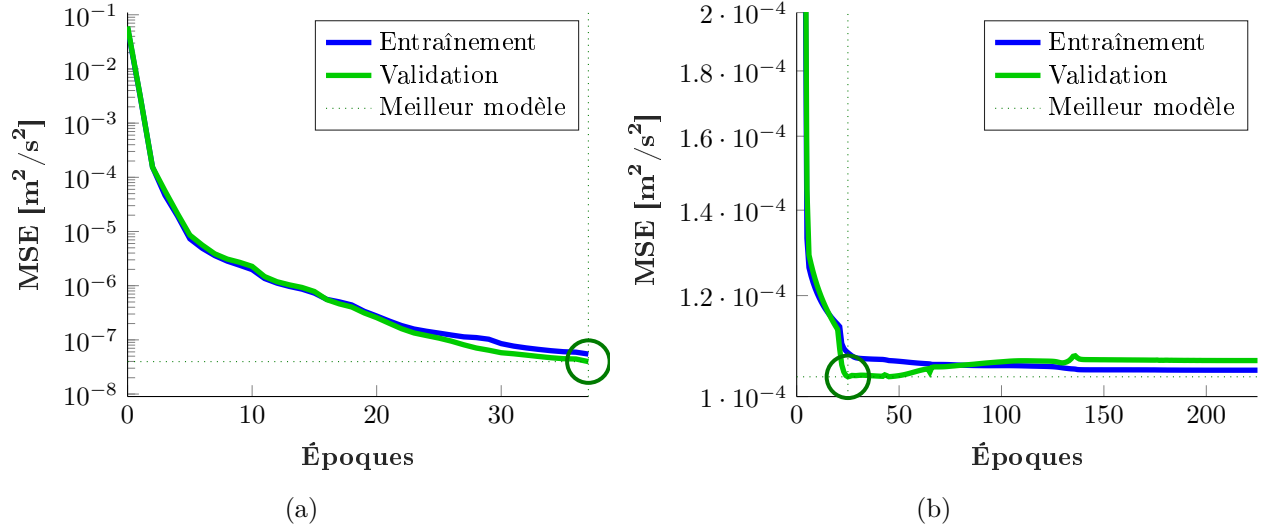


FIGURE 4.4 Courbe d'apprentissage d'un réseau de neurones sur un jeu de 1250 données, (a) sans bruit et (b) avec un bruit gaussien de variance 10^{-4} (m/s)^2 .

approche permet d'évaluer les performances du modèle de manière plus fiable, car elle réduit les effets de sur-ajustement et prend en compte la variabilité des données d'apprentissage.

La figure 4.5 permet de conclure plusieurs choses sur les différents modèles de régression et sur les données d'entraînement. Le premier résultat est qu'avec un jeu de 100 données, il est possible d'avoir une RMSE de la vitesse de propagation de la zone normale inférieure à 1 cm/s (0.01 m/s) pour tous les groupes de modèles. Les performances de chacun des sous-groupes varient selon les hyperparamètres des modèles. Par exemple, pour les modèles de régression par vecteurs de support, il y a un facteur 5 de différence entre la meilleure validation croisée et la pire. Les modèles avec les meilleures performances sont les réseaux de neurones.

Cette étude est faite sur un seul jeu de données, ce qui n'est pas suffisant pour affirmer que les modèles sont toujours plus performants pour prédire la NZPV. De plus, il y a d'autres hyperparamètres pouvant être optimisée sur les modèles SVR, GPR et les Réseaux de neurones. La figure 4.5 aide à réduire le nombre de modèles à tester. Les modèles de chaque catégorie avec les meilleures performances sont retenus pour faire l'analyse de la qualité de différentes bases de données. Le modèle de régression par vecteur de support (SVR) avec les meilleures performances est celui avec un noyau cubique, avec une RMSE de $3.98 \times 10^{-3} \text{ m/s}$. Le polynôme ayant les meilleures performances sur le jeu de test est le polynôme de degré 6 avec une RMSE de 1.17×10^{-3} . Les régressions par processus gaussien (GPR) performent toutes très bien, le modèle gardé est celui avec un noyau Matern 5/2 qui a la plus basse RMSE, soit $1.03 \times 10^{-3} \text{ m/s}$. Pour les réseaux de neurones, le modèle choisi est celui avec le moins de

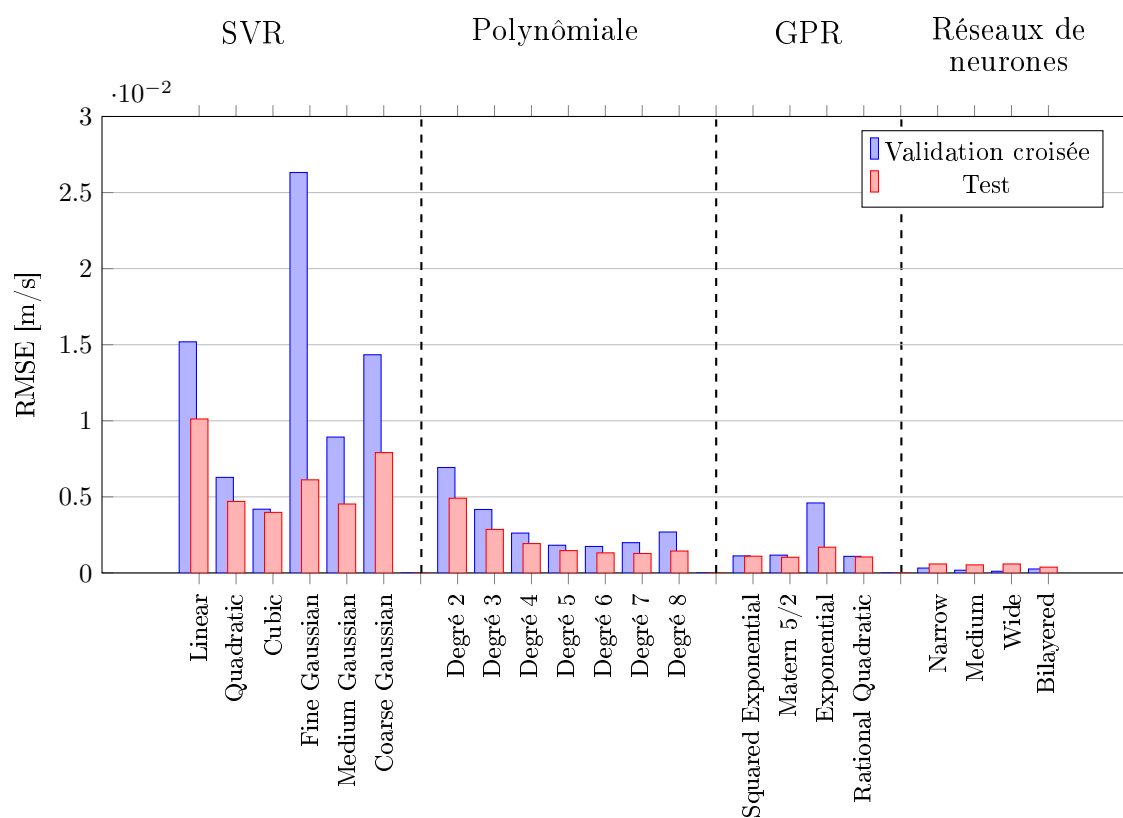


FIGURE 4.5 Performance des modèles entraînés sur un jeu de 100 données distribuées uniformément dans l'espace des paramètres d'entrée et testées sur 10 000 données.

neurones. Le modèle *narrow* a deux paramètres d'entrées, une couche de 10 neurones et une sortie, ou encore (2-10-1). Ce modèle est choisi pour éviter le sur-ajustement du modèle aux données d'entraînement. Les performances de ce modèle sont de 6.038×10^{-4} m/s.

La figure 4.6 montre les courbes de prédictions en fonctions de la vraie valeur de vitesse de propagation de la zone normale de ces quatre modèles, la RMSE ainsi que le coefficient de détermination R^2 . Ce coefficient est un indicateur de la capacité du modèle à décrire la réelle distribution des données. Chaque coefficient est très près de 1, car les modèles sont capables de bien décrire la dynamique. Ces courbes permettent de conclure plusieurs choses sur les modèles. Le premier point est qu'avec 100 données, il est possible d'entraîner plusieurs modèles avec une erreur moyenne inférieure à 1 cm/s. De plus, les pires prédictions sont situées à haute NZPV, ce qui est moins problématique puisque l'erreur relative reste basse. Le modèle de régression par vecteur de support (SVR) a un R^2 de 0.994 et il capte la majorité de la physique. Cependant, l'allure de la courbe de la prédiction en fonction de la vraie valeur montre que ce modèle n'est pas le plus adapté pour exprimer la dynamique de propagation de la zone normale.

Les modèles ont été entraînés sur un jeu de 100 données distribuées uniformément sur les entrées. Pour vérifier quel modèle est le plus adapté pour la problématique, il faut vérifier comment la qualité de prédiction de chaque modèle change selon le nombre de données. Certains modèles performant mieux avec peu de données d'entraînement et d'autres lorsqu'il y en a plus. De plus, puisque les erreurs des modèles présentés à la figure 4.6 sont majoritairement à haute NZPV, il faudrait ajouter plus de données dans cette région.

4.3 Apprentissage actif

Lors de la génération de données, la partie la plus coûteuse en ressource informatique est l'étiquetage des données. Étiqueter une donnée est l'action d'acquérir la valeur à prédire à partir des données d'entrée. Pour le cas du modèle de Dresner, les valeurs d'entrée sont l'épaisseur d'argent et le courant et la valeur à déterminer est la NZPV. Il est important de bien choisir les données à étiqueter, pour limiter le nombre de données à acquérir. Lorsqu'un modèle est entraîné, si les performances de prédictions ne sont pas suffisantes, il faut ajouter des données au jeu de données d'entraînement. L'apprentissage actif est utilisé pour améliorer les performances du modèle en trouvant quelles données sont à ajouter au jeu de données d'entraînement. Le choix des données à étiqueter est fait à partir des performances du modèle.

Une stratégie pour sélectionner les données à étiqueter consiste à utiliser la distribution des erreurs de prédiction des modèles. L'ajout d'une donnée dans une région où l'erreur de

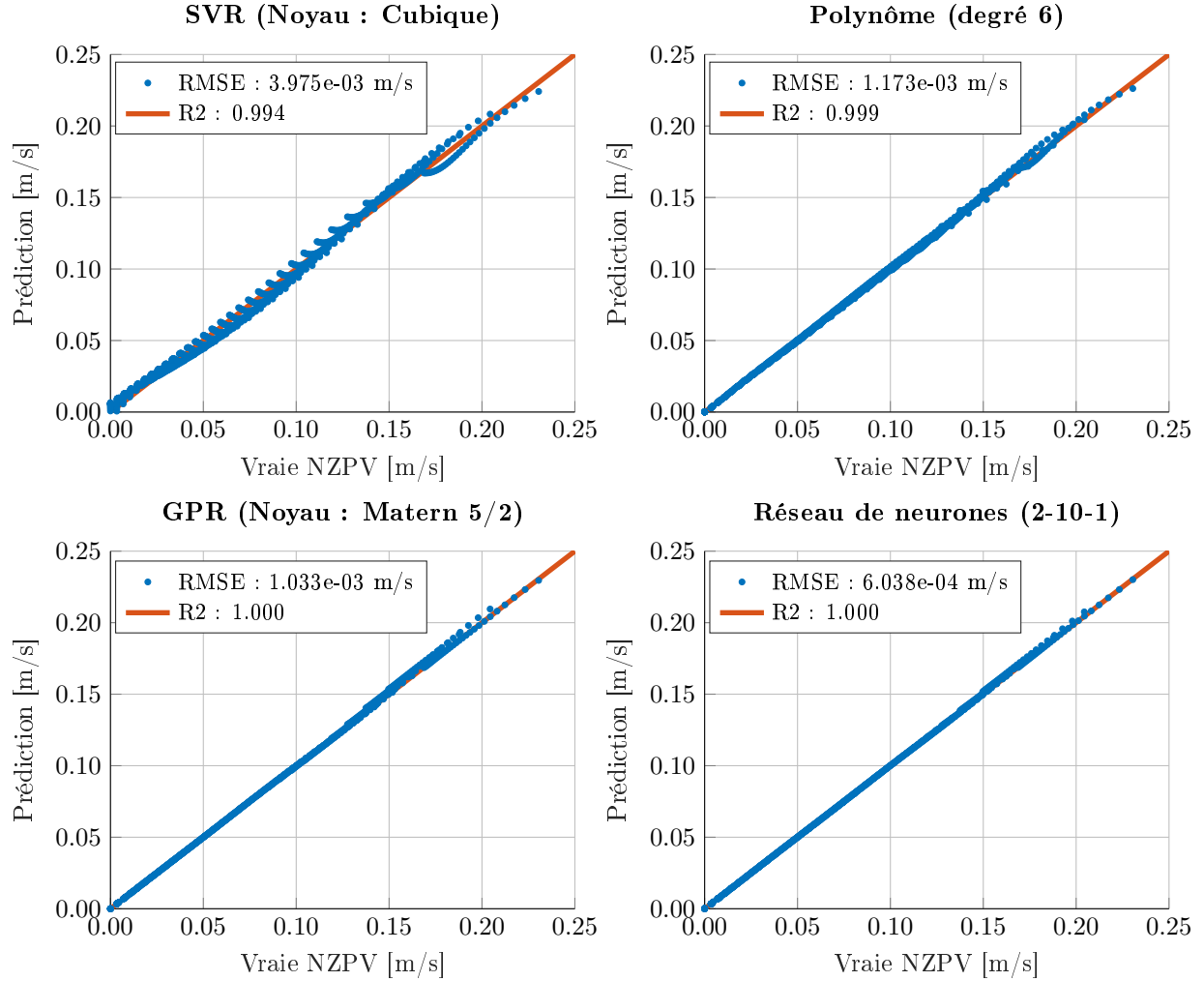


FIGURE 4.6 Régression de quatre différents modèles de prédiction sur les données de tests.

prédiction est élevée améliore davantage la qualité du modèle à cet endroit. Ajouter la donnée avec la plus grande erreur dans la base d'entraînement est un bon choix, mais cela devient problématique lorsqu'il y a plusieurs données à ajouter. En effet, il devient moins efficace de prendre les deux pires valeurs puisque les prédictions du modèle sont continues, la deuxième pire prédiction risque d'être un voisin immédiat de la pire valeur dans l'espace des entrées. Ainsi, l'information apportée par cette deuxième donnée est donc très similaire à celle de la première. Pour remédier à ce problème, l'échantillonnage est fait aléatoirement en utilisant la distribution de l'erreur comme probabilité d'ajouter les données au jeu. Les zones où les performances du modèle sont moindres sont plus souvent sélectionnées que les zones où les prédictions sont meilleures. Cette approche permet d'ajouter autant de données que voulu à chaque itération. Cependant, la distribution de l'erreur n'est pas accessible sur tout le

domaine, mais seulement à certains points. Il faut donc utiliser une métrique qui se comporte comme l'erreur, soit l'erreur généralisée.

4.3.1 Erreur généralisée et requête par comité

L'erreur généralisée est un outil qui permet d'estimer comment un ensemble de modèles performe sur des données qui ne sont pas dans le jeu d'entraînement. Pour les ensembles de modèles, chacun des modèles fait une prédiction et la valeur moyenne de ces prédictions forme la prédiction finale. L'erreur généralisée est l'écart-type des prédictions de chaque modèle de l'ensemble. Pour cette raison, l'erreur généralisée est également appelée le désaccord du comité. Les équations (4.2) et (4.3) montrent la NZPV prédite $\hat{v}$ par un ensemble de M modèles pour des valeurs d'entrées h_{Ag} et i ainsi que le calcul de l'erreur généralisé EG pour ces mêmes entrées :

$$\hat{v}(h_{Ag}, i) = \frac{1}{M} \sum_{i=1}^M f_i(h_{Ag}, i), \quad (4.2)$$

$$EG(h_{Ag}, i) = \sqrt{\frac{\sum_{i=1}^M (f_i(h_{Ag}, i) - \hat{v}(h_{Ag}, i))^2}{M}}. \quad (4.3)$$

Ici, f_i est la prédiction d'un des modèles. Ces équations montrent que le calcul de l'erreur généralisé n'a donc pas besoin de données étiquetées comme pour le calcul de l'erreur de prédiction, mais uniquement des variables d'entrées. Si les modèles prédisent tous la même valeur, alors l'écart-type est nul et l'ensemble performe bien pour cette valeur. Si les modèles prédisent des valeurs très différentes, alors l'ensemble est à risque de mal performer dans cette zone. Le gain majeur avec le calcul du désaccord du comité est qu'il est facile d'avoir la distribution de l'erreur selon les paramètres d'entrées. L'erreur et l'erreur généralisées ont un comportement similaire. Si l'ensemble de modèles a une mauvaise performance pour une certaine zone de l'espace échantillon, les sous-modèles feront des prédictions différentes, ce qui mène à une plus grande erreur généralisée.

Avec le modèle de Dresner, il est possible d'analyser la RMSE des prédictions de la NZPV en utilisant un jeu de données de test, car il est facile de générer ces données. Cela n'est pas pratique pour les données générées par simulation par éléments finis, dont chaque donnée peut prendre plusieurs heures à générer. Chaque point dans le jeu de données de test n'est pas utilisé pour l'entraînement. Afin de déterminer quelles données sont à générer pour améliorer le modèle, il est utile de connaître la distribution des erreurs. Étant donné que l'estimation de la distribution de l'erreur nécessite davantage de données étiquetées, il est

plus pratique d'utiliser l'erreur généralisée. La requête par comité utilise la distribution de l'erreur généralisée de l'ensemble de modèles pour sélectionner les données à étiqueter et les ajouter au jeu de données.

Afin de valider que l'erreur (RMSE) et l'erreur généralisée (désaccord du comité) se comportent de la même manière, un ensemble de modèles constitué de 8 réseaux de neurones est entraîné. Chaque réseau contient une couche cachée formée de 8 neurones avec une tangente hyperbolique comme fonction d'activation. Chaque modèle prend deux valeurs d'entrées, soit l'épaisseur d'argent (h_{Ag}) et le courant réduit ($i = I/I_c$) pour prédire la NZPV avec le modèle de Dresner. Le jeu de données initial contient 25 valeurs formant un quadrillage de $[5 \times 5]$ sur les valeurs d'entrées de manière à avoir une distribution uniformément distribuées, tel que montré à la figure 4.7. Les modèles sont entraînés sur ces données, puis 5 données sont ajoutées au jeu de données d'entraînement. Le processus est répété une deuxième fois pour obtenir un total de 35 données. Les données initiales et les données ajoutées sont montrées à la figure 4.7, et la figure 4.8 montre la distribution de l'erreur généralisée ainsi que la RMSE partielle de la prédiction de la NZPV sur ces jeux de données.

La figure 4.8 montre que la distribution de l'erreur et l'erreur généralisée sont très similaires. La partie (a) de la figure montre la distribution du désaccord du comité et de la RMSE partielle. La RMSE partielle, présentée à la section 3.4.1 montre la distribution de l'erreur sur de vraies données, qui ne seront pas accessibles pour le cas avec les simulations par éléments finis. Les deux métriques font de meilleures prédictions lorsque l'épaisseur d'argent est plus grande et la valeur maximale des erreurs est près de $2.2 \mu\text{m}$. La même analyse peut être faite pour la distribution du courant à droite de la figure 4.8 (a), et l'erreur maximale se situe à $i = 0.9$. Puisque la distribution de l'erreur est plus grande pour ces valeurs, il y a plus de chance que les données sélectionnées pour l'étiquetage soient proches de ces valeurs.

La figure 4.7 montre que c'est effectivement ce qui s'est produit. Plusieurs données ajoutées ont une valeur de $i \approx 0.9$ et de $h_{Ag} \approx 2.2 \mu\text{m}$. Un autre aspect intéressant est comment ces distributions changent lorsqu'il y a plus de données. La figure 4.8 (b) montre la RMSE, et le désaccord du comité diminue après avoir ajouté des données. La valeur de la RMSE totale et la valeur de l'erreur généralisée totale sont différentes, mais les distributions de celles-ci sont similaires. La figure 4.7 montre les 25 données initiales ainsi que les 10 données ajoutées par apprentissage actif.

Pour l'entraînement avec 25 données, le désaccord total est de 1.53×10^{-3} m/s, tandis que la RMSE totale est de 4.44×10^{-3} m/s. Pour l'entraînement avec 35 données, le désaccord total est de 4.27×10^{-4} m/s et que la RMSE totale est de 1.19×10^{-3} m/s. Les deux métriques diminuent après l'ajout de nouvelles données d'entraînement, ce qui indique que les prédic-

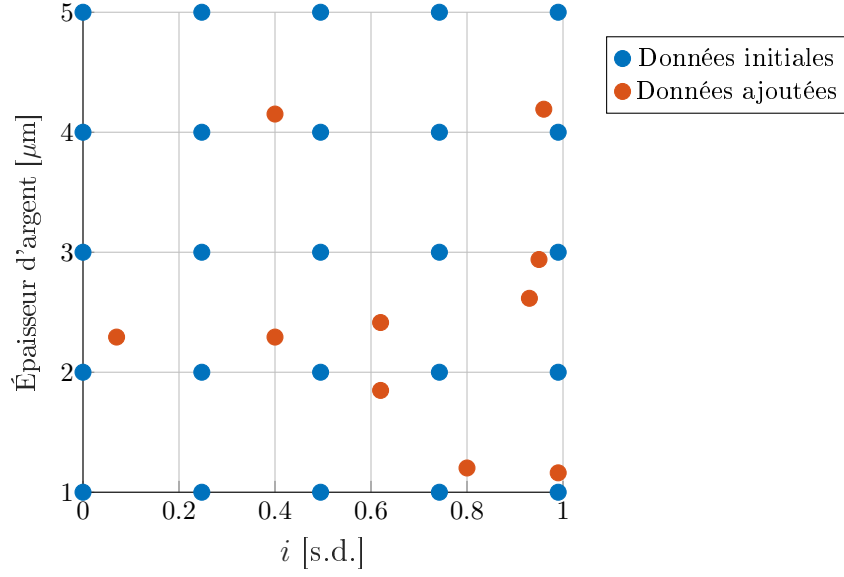


FIGURE 4.7 Distribution des données initiales et des données ajoutées par apprentissage actif.

tions sont meilleures. Puisqu'il y a une différence entre les valeurs des deux métriques, il n'est pas idéal d'utiliser le désaccord du comité comme critère d'arrêt. Utiliser la distribution de l'erreur généralisée comme fonction d'échantillonnage est donc un bon remplacement à l'erreur calculée avec des données tests pour l'apprentissage actif.

4.4 Performance des modèles

Dans cette section, plusieurs modèles sont entraînés sur la distribution uniforme et sur une distribution de données trouvée par apprentissage actif. L'objectif de cette section est de montrer que l'apprentissage actif donne de meilleures données pour l'entraînement et de trouver quel est le meilleur modèle pour prédire la NZPV. Les modèles choisis sont basés sur les résultats de la figure 4.5. Il y a le modèle de vecteur de support avec le noyau cubique, un polynôme de degré 6, un modèle de régression par processus gaussien avec un noyau Matern 5/2 et un réseau de neurones avec une couche cachée composée de 8 neurones. Il a donc l'architecture [2-8-1] pour les deux données d'entrées, les 8 neurones de la couche cachée et la valeur de sortie, soit la NZPV. Il y a également l'ensemble de 8 réseaux de neurones qui est utilisé pour trouver les données à ajouter par apprentissage actif. La figure 4.9 montre comment la RMSE de la prédiction de la NZPV se comporte pour les différents modèles. Chaque point des graphiques correspond à la moyenne de 10 entraînements avec différents

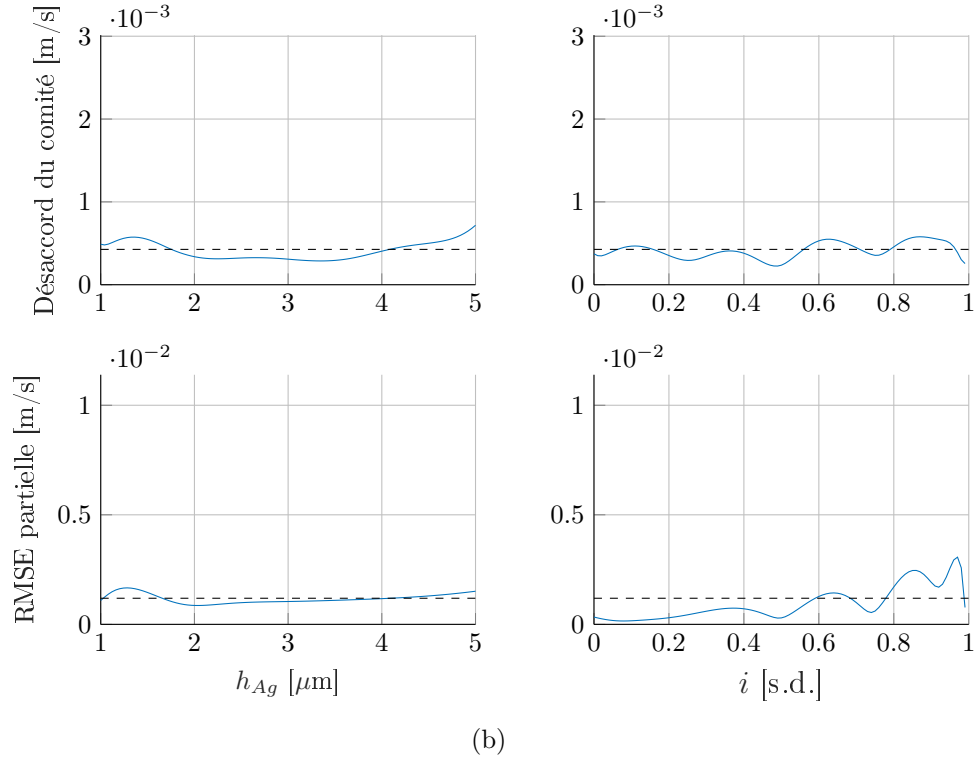
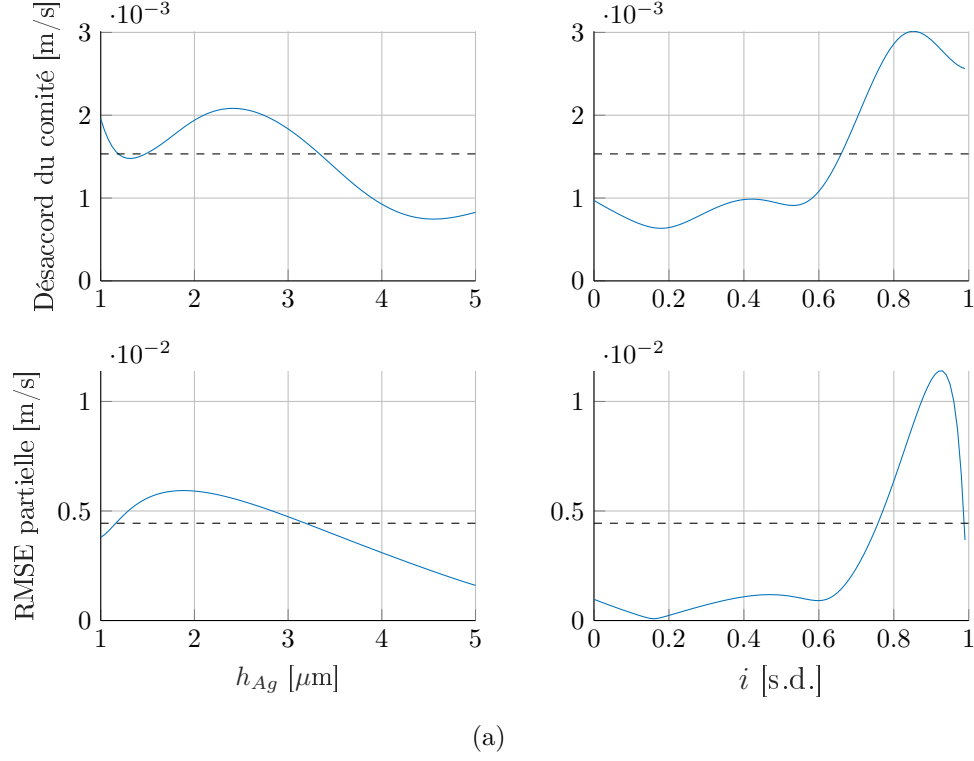


FIGURE 4.8 Comparaison du désaccord du comité et de la RMSE partielle selon l'épaisseur de stabilisateur d'argent h_{Ag} et selon le courant réduit i . (a) Les modèles du comité sont entraînés avec 25 données distribuées uniformément. (b) Les modèles sont entraînés avec les 25 données d'entraînement initiales plus 10 nouvelles données obtenues par requête de comité. La ligne pointillée correspond à la RMSE ou au désaccord moyen.

sous-ensembles de validation croisée. En réalisant chaque mesure 10 fois avec différents sous-groupes de données, les effets de la variabilité aux différents sous-ensembles de données de validation croisée sont réduits, ce qui permet d'obtenir une estimation plus robuste et précise de la qualité du modèle.

Le premier résultat marquant est que le réseau de neurones (NN) performe mieux que les autres modèles sur (a) les données uniformes et sur (b) les données trouvées par apprentissage actif. L'ensemble de modèles performe un peu mieux que le NN seul, mais le gain n'est pas majeur. Le deuxième résultat marquant est que les modèles entraînés par apprentissage actif (b) ont une erreur plus faible ou égale sur le jeu de données de test que les modèles entraînés avec les données uniformes. Ce résultat se voit particulièrement sur le réseau de neurones et légèrement sur le modèle GPR. Le réseau de neurones à la figure (b) atteint un plateau d'environ 1×10^{-4} m/s à 100 données. Le réseau de neurones en (a) atteint sa meilleure valeur de 2.7×10^{-4} m/s à 195 données. Le modèle SVM et les polynômes ne semblent pas être de bons candidats pour exprimer la relation entre le courant, l'épaisseur de stabilisateur et la vitesse de propagation de la zone normale.

Un autre point intéressant est que sur les données uniformes, les performances des réseaux de neurones et de l'ensemble de réseaux de neurones sur les jeux de données test sont moindres que celle sur les jeux d'entraînement et de validation. Il y a une mauvaise généralisation des modèles et donc un sur-ajustement sur les données d'entraînement. Lorsqu'il y a 100 données dans le jeu d'entraînement, le réseau de neurones (NN) atteint une RMSE inférieure à 2×10^{-3} m/s sur les données d'entraînement, mais il atteint une RMSE de 5×10^{-3} m/s sur les données test.

Une hypothèse est que la mauvaise généralisation provient de la structure des données d'entraînement et de validation. En effet, les données d'entraînement se retrouvent toutes sur un quadrillage, tel que montré à la figure 4.7, en bleu, et le jeu test ont des données pour des valeurs entre les lignes principales de ce quadrillage. Par exemple, pour le jeu contenant 25 données, chacune des données se retrouve sur un quadrillage de taille $[5 \times 5]$. La valeur de l'épaisseur d'argent peut prendre une valeur de 1, 2, 3, 4 ou de 5 μm . Le jeu de test contient des données dont l'épaisseur d'argent est de 4.5 μm , qui n'est pas présent dans le jeu d'entraînement. Les données d'entraînement et de validation sont donc toutes sur ce quadrillage, mais pas les données de test, ce qui expliquerait la mauvaise généralisation. La même logique est applicable pour le courant.

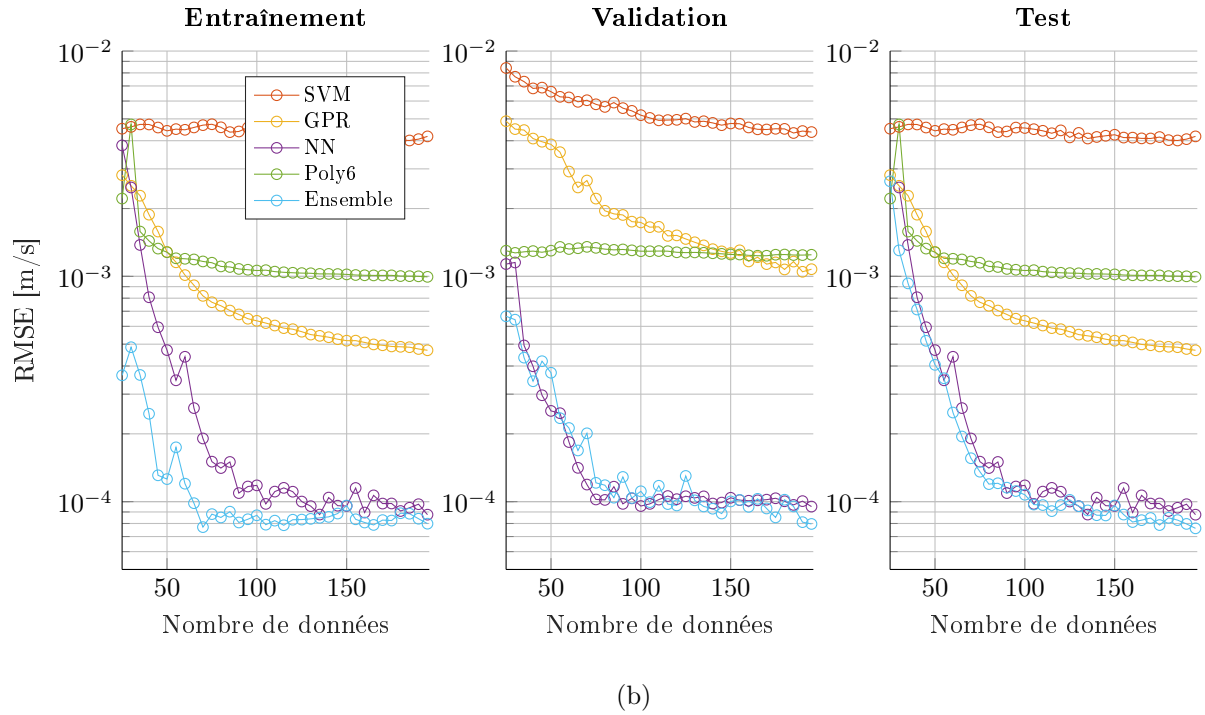
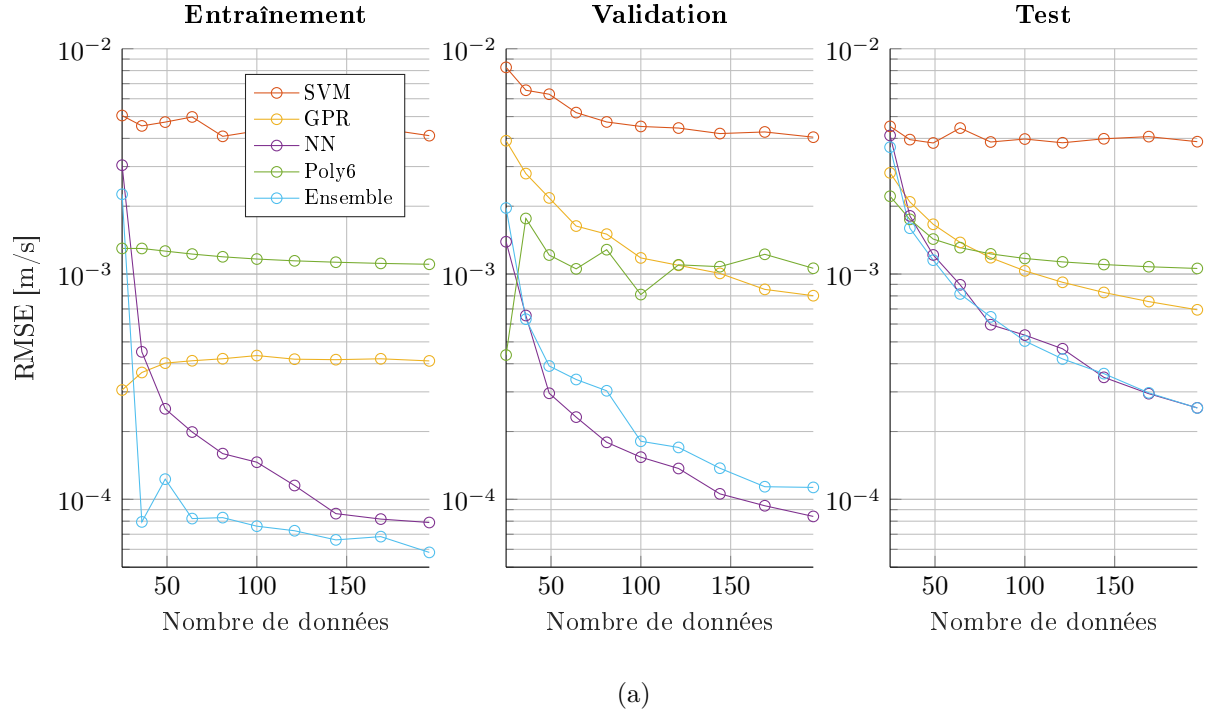


FIGURE 4.9 Racine de l'erreur quadratique moyenne de la prédiction de la NZPV de différents modèles de régression classiques selon le nombre de données dans le jeu de données d'entraînement. Les données sont distribuées uniformément sur les entrées (a) et déterminées par apprentissage actif (b). Les modèles sont entraînés en utilisant la validation croisée avec cinq sous-ensembles de donnée.

4.5 Discussion

La meilleure méthodologie trouvée pour construire un jeu de données de la NZPV en fonction de ses paramètres est d'utiliser l'apprentissage actif avec le désaccord du comité comme fonction d'échantillonnage de données à étiqueter. Le modèle entraîné est un ensemble de modèles constitué de 8 réseaux de neurones ayant tous la même architecture, soit deux neurones d'entrées, une couche cachée de huit neurones et un neurone de sortie. Le jeu de données initiales sera distribué uniformément.

La figure 4.9 montre l'erreur sur un jeu de données de test de 10 000 données, ce qui est possible à générer puisque l'équation de la NZPV de Dresner est analytique. Le jeu de données de test est construit ainsi pour pouvoir tester un grand nombre de valeurs. Cette métrique n'est pas disponible lorsque le jeu de données est construit à partir de simulations par éléments finis, qui prennent plus de temps à étiqueter.

Il faut toutefois une façon de vérifier que le modèle de régression fonctionne sur des données qui n'ont pas été utilisées lors du processus d'entraînement. Il est donc essentiel donc de construire un jeu de données de test, à part des données d'entraînement. Il est commun d'utiliser 20% de données pour le jeu de données de test. Pour un jeu de 50 données, il en faut donc 10 pour tester les modèles.

CHAPITRE 5 PRÉDICTION DE LA NZPV D'UN RUBAN SUPRACONDUCTEUR

Lors de la propagation de la zone normale dans un ruban supraconducteur, plusieurs phénomènes se produisent simultanément et s'influencent mutuellement. Lors d'un quench, la température dans le ruban change rapidement, ce qui affecte les propriétés des matériaux comme la capacité calorifique, la résistivité électrique et la conductivité thermique. Il y a également l'aspect tridimensionnel dont il faut tenir compte. La géométrie du ruban a un effet sur le parcours du courant et donc la propagation de la zone normale. Afin de reproduire la dynamique le plus fidèlement possible, des simulations par éléments finis sont faites en prenant en considération tous ces éléments.

Dans cette section, la méthode d'étiquetage des données par simulation par éléments finis est décrite. La figure 5.1 présente la géométrie d'un des rubans simulés. Un jeu de données est généré pour un ruban commercial dont l'architecture est similaire à celle des rubans fabriqués par la compagnie SuperPower Inc., appelés “*BARE tape*” ou ruban régulier. Un second jeu de données pour une architecture de type CFD hybride développé à Polytechnique Montréal.

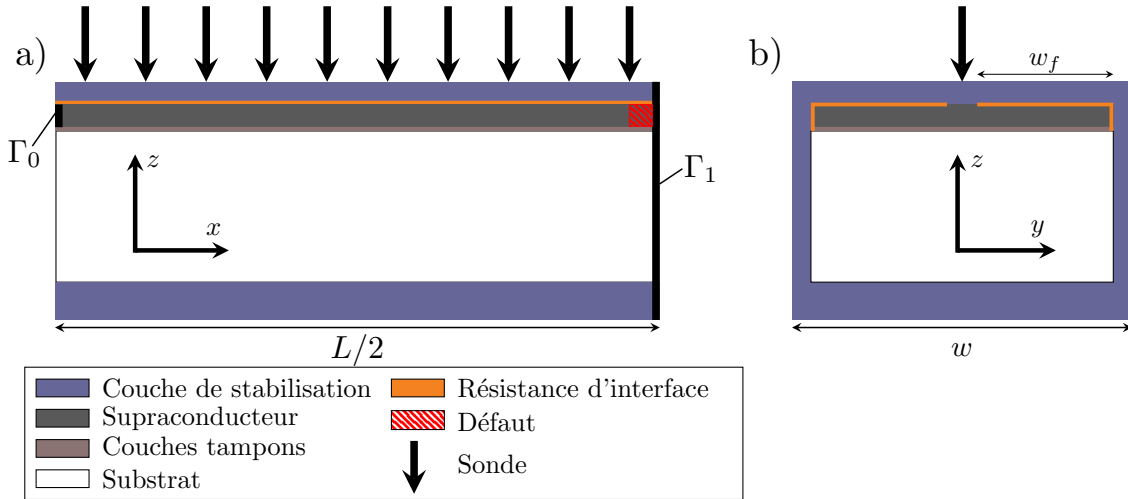


FIGURE 5.1 Géométrie d'un ruban 2G-HTS avec architecture hybride-CFD utilisée pour la simulation par éléments finis. a) Section du ruban selon la direction du courant et b) la coupe selon l'axe perpendiculaire au courant.

5.1 Générations des données

La génération de données se fait en trois étapes. Il faut (1) déterminer les paramètres de la simulation, (2) faire la simulation par éléments finis et (3) calculer la NZPV. La sélection des paramètres de la simulation est faite initialement en utilisant une distribution uniforme des paramètres variables choisis. Par la suite, les distributions des paramètres des simulations subséquentes sont choisies par apprentissage actif. Les simulations par éléments finis sont faites avec le logiciel COMSOL 5.6 avec les supercalculateurs de l'Alliance de recherche numérique du Canada [45]. À partir des simulations résolues, un script Matlab permet d'extraire la NZPV et de générer un fichier CSV contenant le jeu de données.

5.1.1 Simulation par éléments finis

La propagation de la zone normale peut être modélisée par une simulation électrothermique avec le logiciel COMSOL. Les modules de COMSOL utilisés sont *Current Conservation* pour la partie électrique et *Heat Transfer in Solids* pour l'aspect thermique. Les valeurs intéressantes de la simulation sont la distribution de la température dans le temps $T(\mathbf{x}, t)$ ainsi que la distribution de la densité de courant dans le temps $J(\mathbf{x}, t)$. Pour la partie électrique du problème, le système d'équations à résoudre est le suivant :

$$\left\{ \begin{array}{ll} \nabla \cdot (-\sigma(T)\nabla V) &= 0 \quad \text{dans } \Omega, \\ \int_{\partial\Omega} -\mathbf{n} \cdot (\sigma(\mathbf{T})\nabla \mathbf{V}) &= I_0(t) \quad \text{sur } \Gamma_0, \\ V &= 0 \quad \text{sur } \Gamma_1, \\ \mathbf{n} \cdot \nabla V &= 0 \quad \text{ailleurs,} \end{array} \right. \quad \begin{array}{l} (5.1) \\ (5.2) \\ (5.3) \\ (5.4) \end{array}$$

où $\sigma(T)$ est la conductivité des matériaux, $\mathbf{n}$ est un vecteur unitaire normal à la surface, $I_0(t)$ est le courant injecté dans le supraconducteur. La figure 5.1 montre le domaine de simulation Ω ainsi que les surfaces Γ_0 et Γ_1 . La conductivité électrique des matériaux varie fortement avec la température, surtout celle du supraconducteur. La conductivité de la couche supraconductrice est modélisée dans COMSOL comme deux conductivités en parallèle :

$$\sigma_{YBCO}(T) = \sigma_{sc}(T) + \sigma_n, \quad (5.5)$$

$$\text{avec } \sigma_{sc}(T) = \frac{J_c(T)}{E_c} \left(\frac{\|\mathbf{E}\|}{E_c} \right)^{\frac{1-n(T)}{n(T)}}, \quad (5.6)$$

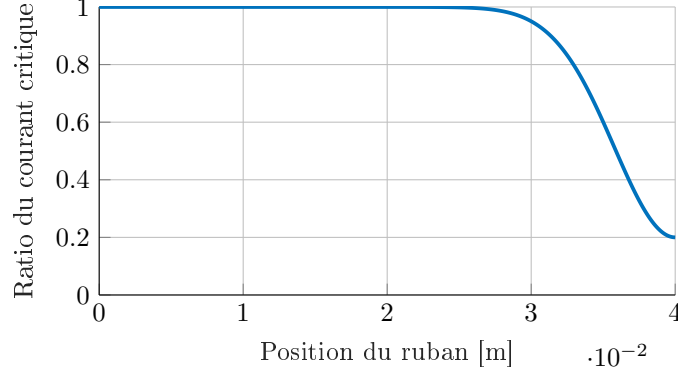


FIGURE 5.2 Distribution du ratio du courant critique selon la longueur d'un ruban avec un défaut à 0.04 m.

où $\sigma_{sc}(T)$ est la conductivité dans l'état supraconducteur et σ_n est la conductivité dans l'état normal. $J_c(T)$ est la densité de courant critique, $||\mathbf{E}||$ est la norme du champ électrique et E_c est le champ électrique critique. Pour le supraconducteur, la conductivité électrique dans l'état normal est de 10^6 Sm^{-1} et σ_{sc} est modélisée avec une loi de puissance. Ces valeurs sont basées sur le matériau supraconducteur YBaCuO. Pour introduire une zone normale, le paramètre σ_{sc} est réduit localement sur le ruban pour modéliser un défaut. La densité de courant critique est plus faible dans la région du défaut que dans celle du ruban, tel que montré à la figure 5.2. La réduction de la densité de courant critique suit une distribution gaussienne centrée avec le ruban et la valeur utilisée pour les simulations est de $J_{c,défaut} = 0.2J_c$. Le facteur $n(T)$ et la densité de courant $J_c(T)$ ont été déterminés à l'aide de données expérimentales et ont été ajustés avec des fonctions polynomiales lors d'un précédent projet (FASTGRID [20]). Les polynômes utilisés dans la simulation sont les suivants :

$$J_c(T) = 9 \times 10^9 \left(4.016 \times 10^8 T^4 - 8.9 \times 10^{-3} T^3 + 7.256 \times 10^{-1} T^2 - 35.7T + 1189 \right), \quad (5.7)$$

$$n(T) = 0.75 \left(-2.99 \times 10^{-7} T^5 + 7.546 \times 10^{-5} T^4 - 7.844 \times 10^{-3} T^3 + 4.259 \times 10^{-1} T^2 - 12.12T + 183.4 \right). \quad (5.8)$$

Le paramètre $I_0(t)$ représente le courant injecté dans le supraconducteur par la surface Γ_0 . La température est calculée à l'aide de l'équation de la chaleur :

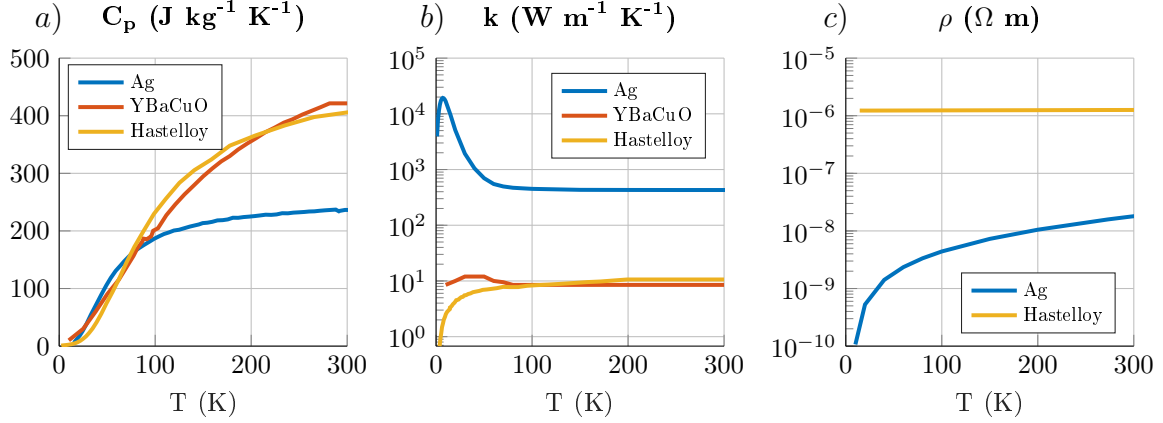


FIGURE 5.3 Propriétés des matériaux utilisés dans le logiciel COMSOL.

$$\begin{cases} \rho_m C_p(T) \frac{\partial T}{\partial t} + \nabla \cdot (-k(T) \nabla T) = Q(J) & \text{dans } \Omega, \\ \mathbf{n} \cdot \nabla T = 0 & \text{sur } \partial\Omega, \end{cases} \quad (5.9)$$

$$(5.10)$$

où $\rho_m(T)$ est la densité massique, $C_p(T)$ est la chaleur spécifique et $k(T)$ est la conductivité thermique. Le couplage entre le problème thermique et le problème électrique se fait par l'effet Joule, soit $Q(J) = E \cdot J = \sigma(T)(-\nabla V)^2$. Les dépendances en température de C_p , k et la résistivité électrique $\rho(T) = \sigma^{-1}(T)$ sont présentées à la figure 5.3 pour l'Hastelloy, l'argent et l'YBCO.

La figure 5.4 montre la propagation de la zone normale dans une simulation d'un quart de ruban avec une architecture CFD hybride. La zone blanche correspond à une température supérieure à 90 K, la température critique et la zone en rouge est la zone qui est encore supraconductrice.

5.1.2 Calcul de la NZPV

La simulation de la propagation de la zone normale d'un ruban supraconducteur permet d'obtenir la distribution de la température dans le ruban dans le temps. Les valeurs de température utilisées pour calculer la NZPV sont celles sur le dessus du ruban, puisque c'est à cet endroit que les données sont prises expérimentalement. Il est à noter que les données expérimentales sont des voltages mesurés aux différentes sondes et non la température. En effet, des électrodes de mesure sont uniformément réparties sur la longueur du ruban (axe x), à une position fixe sur la largeur (axe y). Pour calculer la NZPV à partir d'une simulation, il faut

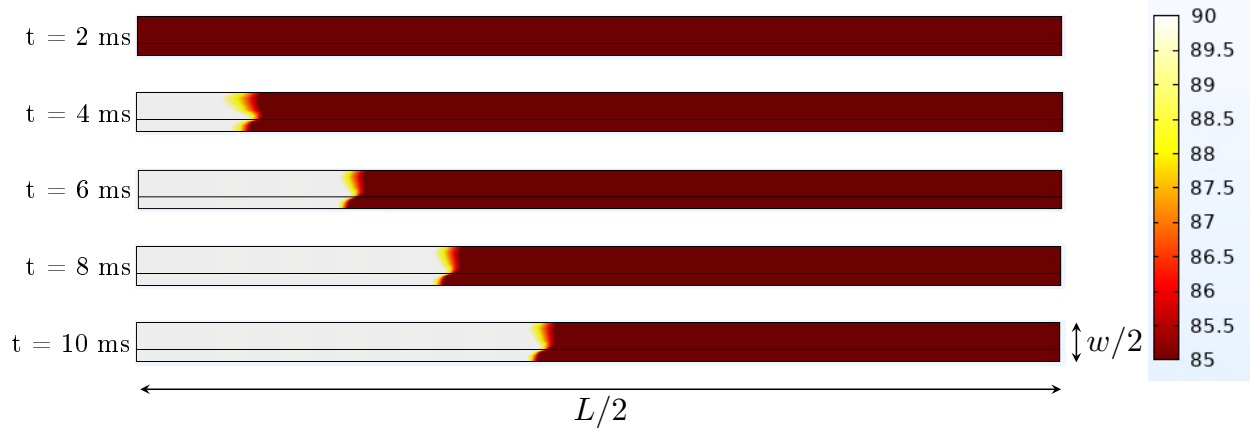


FIGURE 5.4 Propagation de la zone normale pour une simulation par éléments finis avec le logiciel COMSOL réalisée sur le quart d'un ruban CFD hybride. Le défaut est à l'extrémité droite et la ligne délimite la zone avec une résistance d'interface.

trouver le moment où la température mesurée pour chaque sonde est égale à la température critique $T(x_i, t'_i) = 90$ K, soit la température où le supraconducteur passe à l'état normal. Ici, i est l'index de la sonde i , $x_i = i \times \Delta x$ est la position de la sonde i , et t'_i est le temps auquel la sonde i atteint la température de 90 K. Le temps de la simulation est discret, il faut donc utiliser une interpolation pour obtenir une estimation du temps où la température atteint T_c . Pour chaque sonde, il est possible de calculer une valeur de la NZPV $v(x_i)$ avec le temps nécessaire pour que la zone normale progresse du point x_{i-1} au point x_i . L'équation 5.11 présente ce calcul :

$$v(x_i) = \frac{\Delta x}{\Delta t} = \frac{x_i - x_{i-1}}{t'_i - t'_{i-1}} \quad (5.11)$$

La partie gauche de la figure 5.5 montre comment la température mesurée par chaque sonde sur le ruban varie lors de la propagation de la zone normale. La partie droite de cette figure montre la valeur de NZPV calculée par chaque sonde, ainsi que la valeur de NZPV mesurée en prenant la moyenne ainsi que l'écart-type (SD). Chaque sonde permet d'obtenir une mesure de la NZPV. La valeur de NZPV calculée est la moyenne des valeurs les plus éloignées du défaut, car il la dynamique d'amorçage de la propagation de la zone normale ne nous intéresse pas. Seulement l'état stationnaire de la propagation de la zone normale nous intéresse.

La valeur de NZPV correspond à la moyenne des NZPV lorsque le plateau est atteint. Il est important de connaître la qualité de chaque donnée dans le jeu de données. Certaines simulations sont plus stables numériquement et l'erreur sur la NZPV est alors plus faible.

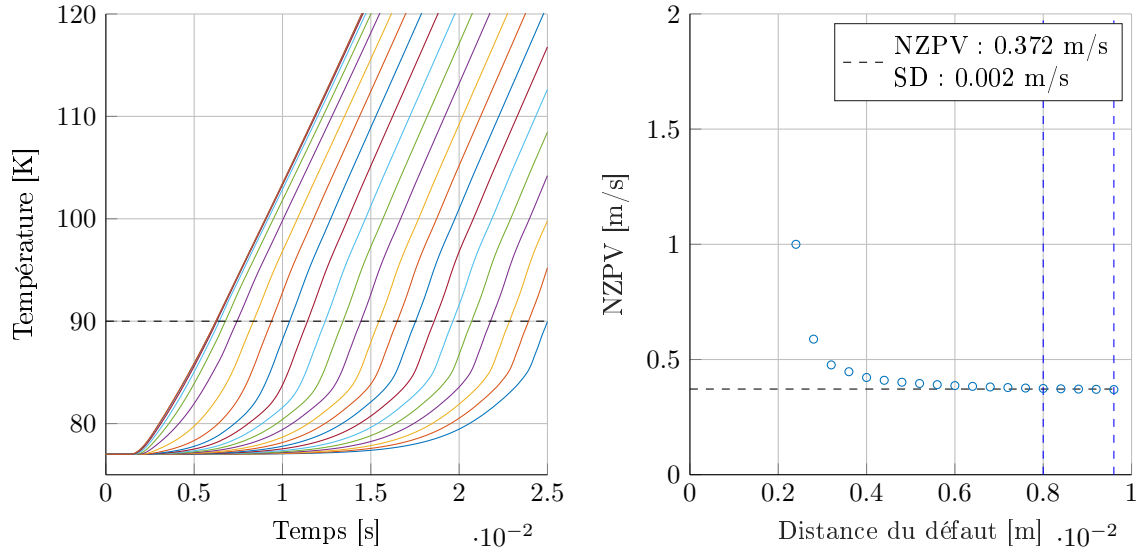


FIGURE 5.5 Gauche : Mesures de la température des sondes dans une simulation par éléments finis en fonction du temps. Droite : NZPV calculée à partir de chaque sonde. Les lignes pointillées verticales indiquent les points utilisés pour le calcul de la NZPV en régime stationnaire.

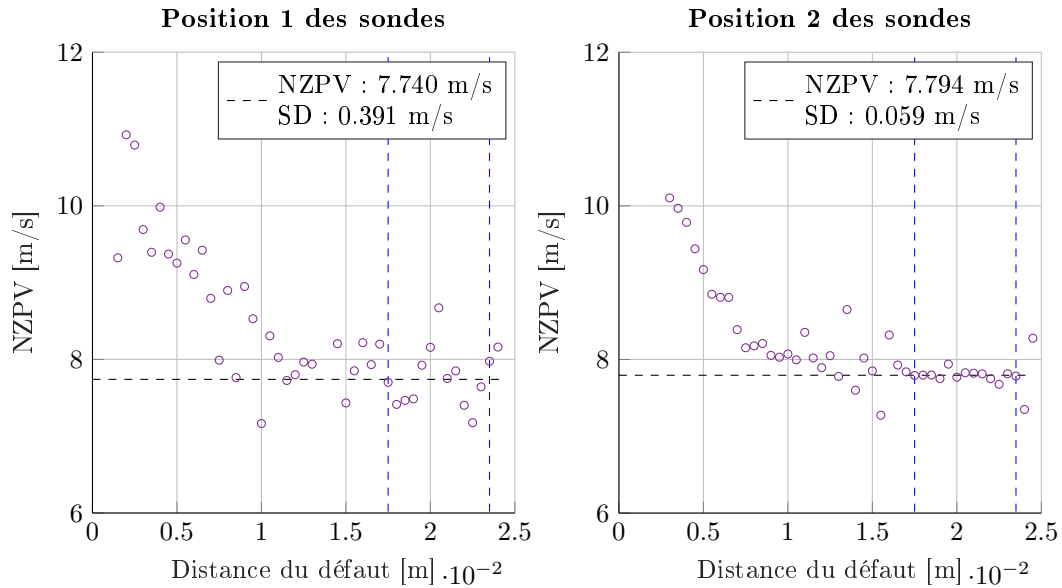


FIGURE 5.6 Vitesse de propagation de la zone normale d'un ruban CFD hybride simulé sur COMSOL pour un cas instable et mesurée à deux positions des sondes sur l'axe y.

Lorsque la simulation présente des fluctuations numériques importantes, la variance de la NZPV calculée augmente, et on dit que la simulation est “instable”. Cette variabilité de la NZPV est due à des erreurs numériques et ne reflète pas la dynamique de la propagation de la zone normale. La simulation de la figure 5.5 montre un cas dit “stable” et la figure 5.6 montre un cas dit “instable”. Pour calculer la NZPV d’un cas instable, il est possible de changer la position des sondes sur le ruban afin d’obtenir de meilleures valeurs. La figure 5.1 b) montre un exemple où les sondes sont au milieu du ruban. La figure 5.6 montre les valeurs de vitesse calculées à deux positions sur l’axe y du ruban. Les sondes de la figure de gauche sont à $1.2 \mu\text{m}$ du côté du ruban (axe y) et les sondes sont sur l’extrémité de la largeur pour la figure de droite. La figure 5.1 b) montre un exemple où les sondes sont placées au centre du ruban, soit à 2 mm du bord du ruban.

La position des sondes peut changer la valeur de la NZPV calculée ainsi que l’incertitude sur celle-ci. La figure 5.6 gauche donne une NZPV de $7.740 \pm 0.391 \text{ m/s}$. Cela correspond à une erreur relative de 5%. La NZPV calculée avec les sondes sur l’extrémité du ruban donne une valeur de $7.794 \pm 0.059 \text{ m/s}$ ou (0.76%). Les deux valeurs sont similaires, mais l’incertitude de la valeur prise au bord du ruban est nettement inférieure à celle prise à $1.2 \mu\text{m}$ du bord du ruban.

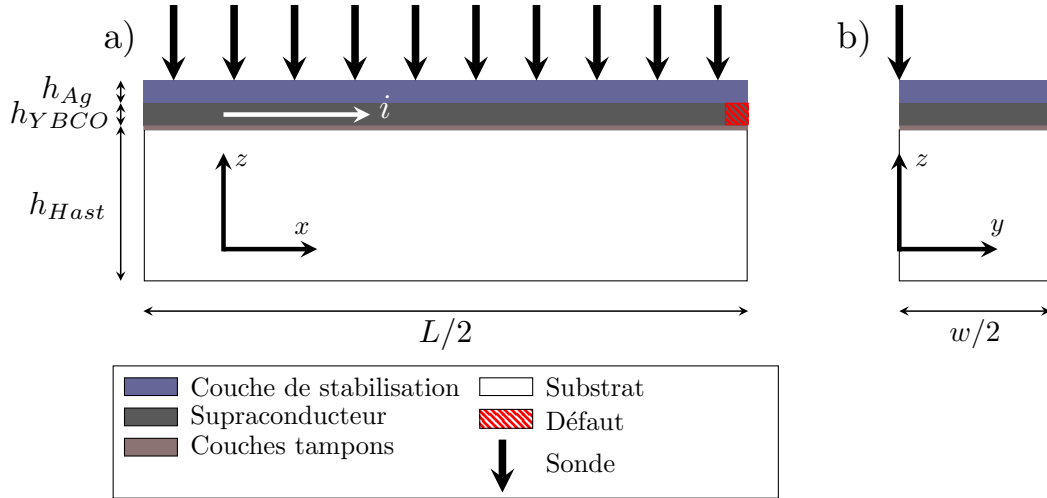


FIGURE 5.7 Géométrie simulée par éléments finis d’un ruban régulier. a) Section du ruban dans le sens de la longueur et b) section du ruban dans le sens de la largeur.

TABLEAU 5.1 Paramètres du ruban régulier

Propriété	Symbole	Valeur	Unité
Température critique	T_c	90	K
Température ambiante	T_b	77	K
Épaisseur du substrat	h_{Hast}	100	μm
Épaisseur du supraconducteur	h_{YBCO}	1	μm
Largeur du ruban	w	4	mm
Densité de courant critique	J_c	2.5×10^{10}	A/m ²
Ratio du courant critique au défaut	N/A	0.2	S.D.

5.2 Performance d'un modèle entraîné sur une simulation d'un ruban régulier

Les rubans supraconducteurs réguliers sont des *second generation high-temperature superconductor coated conductors* (2G HTS CC) fabriqués par la compagnie SuperPower Inc. L'architecture générale est montrée à la figure 2.1 de la section 2. Pour accélérer la simulation de ceux-ci, seulement la couche de stabilisation du dessus est simulée. De plus, il y a deux plans de symétrie, soit les plans $x - z$ et $y - z$, ce qui permet de simuler uniquement le quart du ruban et donc de réduire le temps de calcul. La géométrie simulée du ruban est montrée à la figure 5.7 et les paramètres utilisés en simulation pour ce type de ruban sont montrés dans le tableau 5.1.

Les deux paramètres étudiés sont l'épaisseur de la couche stabilisatrice d'argent h_{Ag} et le ratio i du courant appliqué I sur le critique appliqué I_c , i.e. $i = I/I_c$. Avec une épaisseur de supraconducteur de $1 \mu\text{m}$, une largeur de ruban de 4 mm et une densité de courant critique de $2.5 \times 10^{10} \text{ Am}^{-2}$, le courant critique du ruban est de :

$$I_c = J_c \cdot w \cdot h_{YBCO} = 100 \text{ A.} \quad (5.12)$$

L'intervalle de courant simulé est de 80 A à 100 A, ce qui implique que i prend des valeurs dans l'intervalle $[0.8, 1]$. L'intervalle d'intérêt de l'épaisseur de stabilisateur est de $[1, 5] \mu\text{m}$. Avec ces informations, un jeu de données initial est généré avec 25 données distribuées uniformément selon h_{Ag} et i . La figure 5.8 montre la distribution des données ainsi que les valeurs de NZPV. La figure 5.8 montre la répartition des données, mettant en évidence une variation plus prononcée de la NZPV entre $1 \mu\text{m}$ et $2 \mu\text{m}$ par rapport aux autres épaisseurs. La valeur minimale de NZPV est de 0.128 m/s lorsque $h_{Ag} = 5 \mu\text{m}$ et $i = 0.8$. La valeur maximale de la NZPV est de 2.073 m/s lorsque $h_{Ag} = 1 \mu\text{m}$ et $i = 1$.

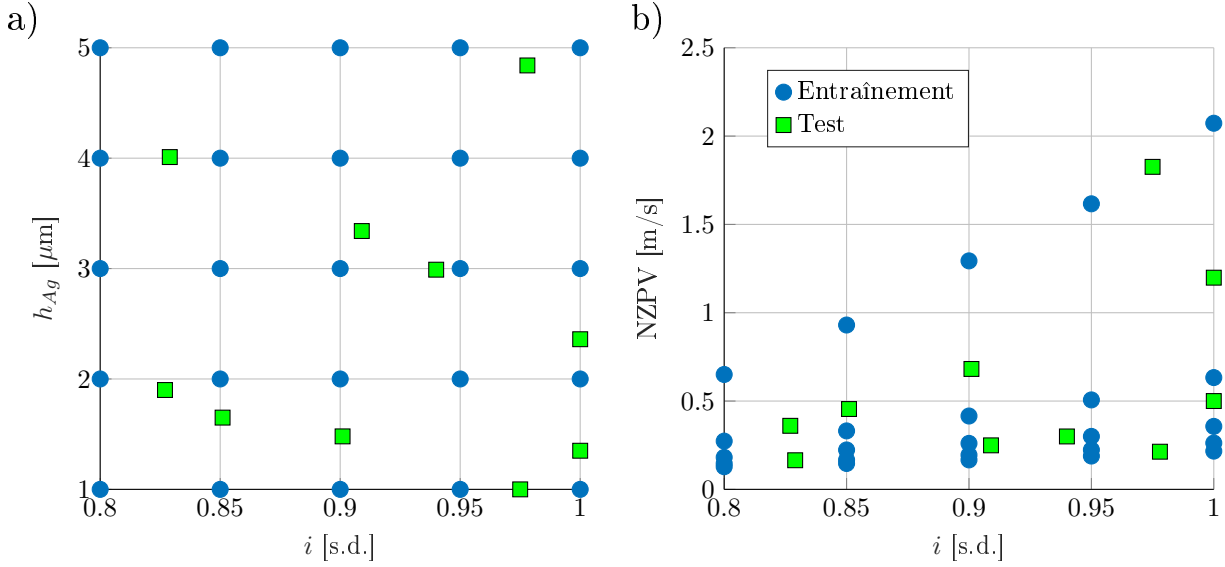


FIGURE 5.8 a) Distribution de données du jeu d'entraînement initial et du jeu de test pour les simulations de ruban régulier. b) NZPV selon le courant.

5.2.1 Entraînement des modèles

Lors de l'entraînement par apprentissage actif, il faut déterminer un critère pour déterminer si le modèle présente des performances satisfaisantes. Pour les rubans réguliers, le critère d'arrêt est que la racine carrée de l'erreur quadratique moyenne soit inférieure à 2 cm/s sur un jeu de test ($RMSE_{test} < 0.02$ m/s). Dix données tests sont générées de manière à ce que l'intervalle de NZPV possible soit testé et que le domaine des entrées soit bien échantillonné. La figure 5.8 montre la distribution de ces données. Si la RMSE sur les données du jeu de test est plus grande que le critère, cinq données sont ajoutées aux jeux d'entraînement et les modèles sont réentraînés.

Avec les données d'entraînement initiales, un ensemble de 8 réseaux de neurones est entraîné sur le logiciel Matlab et la boîte à outils *Deep Learning Toolbox*. L'algorithme de Levenberg-Marquardt avec régularisation bayésienne est utilisé pour actualiser les poids des réseaux de neurones à chaque utilisation. La fonction d'activation est une tangente hyperbolique. Pour terminer chaque entraînement, 20% des données sont utilisées comme un jeu de validation. Ce jeu permet de déterminer si les réseaux de neurones se sur-ajustent sur les données d'entraînement. La figure 5.9 montre l'architecture d'un réseau de neurones entraîné sur ces données. Cette figure montre également les différents critères d'arrêt d'entraînement de celui-ci de l'outil **nntraintool** de MATLAB. Cet outil est utilisé 8 fois à chaque entraînement, une fois pour chacun des réseaux de neurones de l'ensemble.

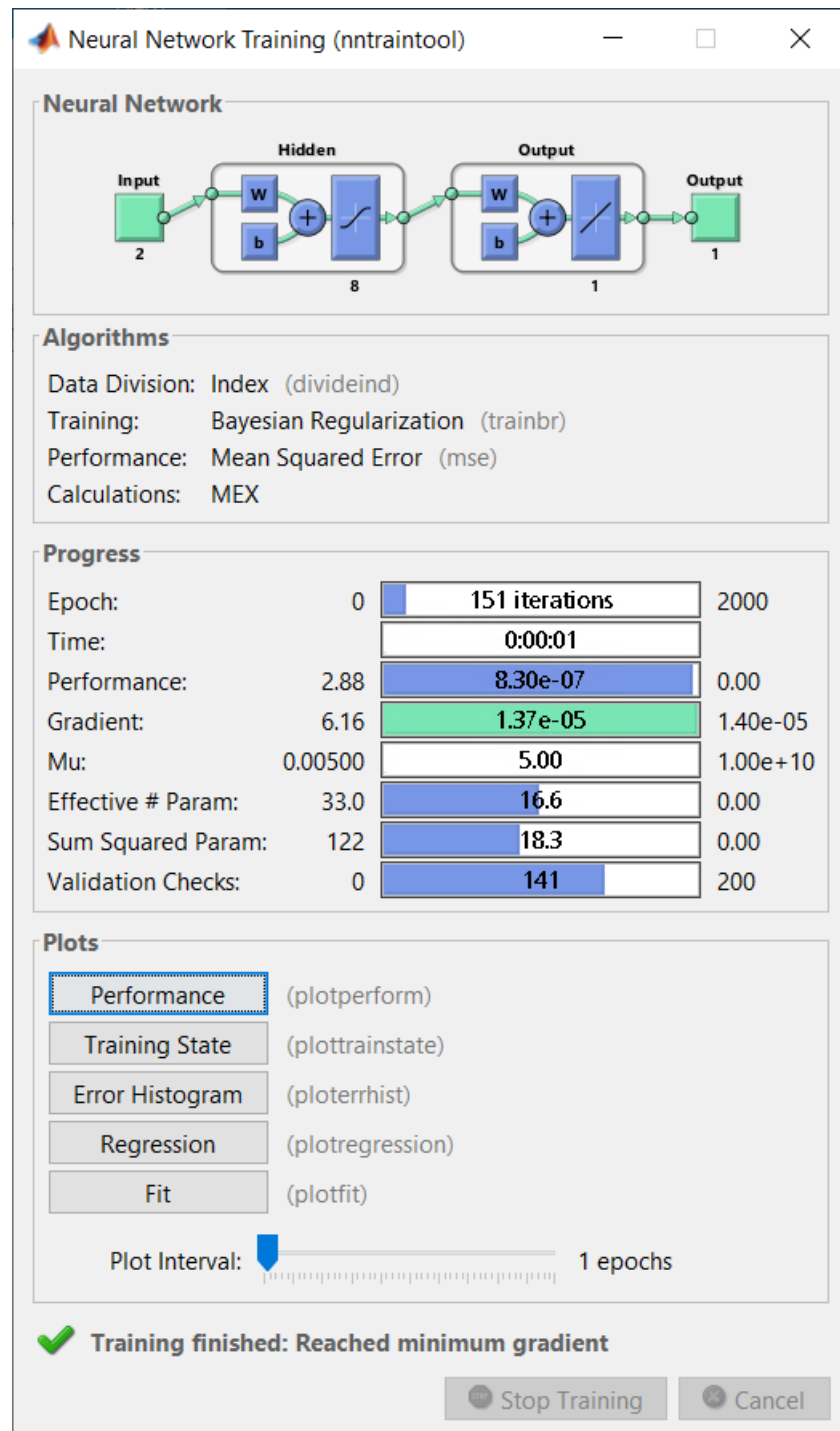


FIGURE 5.9 Outil d'entraînement de réseau de neurones (**nntraintool**) du “*Deep Learning Toolbox*” sur MATLAB.

L'entraînement des réseaux de neurones est terminé lorsque le critère sur le gradient de la performance a été atteint. Ce critère arrête l'entraînement lorsque la variation de performance entre deux époques est faible. Cet entraînement est répété pour les 7 autres réseaux de neurones de l'ensemble. Après avoir été entraîné sur les 25 premières données, le modèle présente des valeurs de racine carrée de l'erreur quadratique moyenne (RMSE) pour les jeux d'entraînement, de validation et de test, respectivement égales à 1.39 cm/s, à 3.74 cm/s et à 5.70 cm/s. Puisque l'erreur sur le jeu de test est supérieure à 2 cm/s, il faut ajouter des données en utilisant la requête par comité. Avec l'ensemble de modèles entraînés et un jeu de données d'entrée (h_{Ag}, i) avec une distribution uniforme de 900 données non étiquetées, le désaccord de comité (l'erreur généralisée) est calculé. La distribution du désaccord du comité est calculée sur 900 données, soit sur une distribution uniforme de 30 valeurs de h_{Ag} et 30 valeurs de i . Plus il a de données pour calculé le désaccord du comité, plus les courbes de celui-ci seront lisses. Les courbes du désaccord du comité après l'entraînement sur les 25 données sont présentées à la figure 5.10.

À partir de la distribution du désaccord du comité, cinq données sont ajoutées à la base d'entraînement, tel que montré à la figure 5.10. Cette figure montre que le désaccord est plus grand lorsque le courant est près de 1 ou lorsque l'épaisseur d'argent est près de 1 μm . Le jeu d'entraînement passe de 25 données à 30 données. Ces données sont simulées et ajoutées au jeu d'entraînement. Ce processus est répété à quatre reprises pour obtenir une RMSE test satisfaisante. L'évolution de la RMSE des jeux d'entraînement, de validation et de test sont

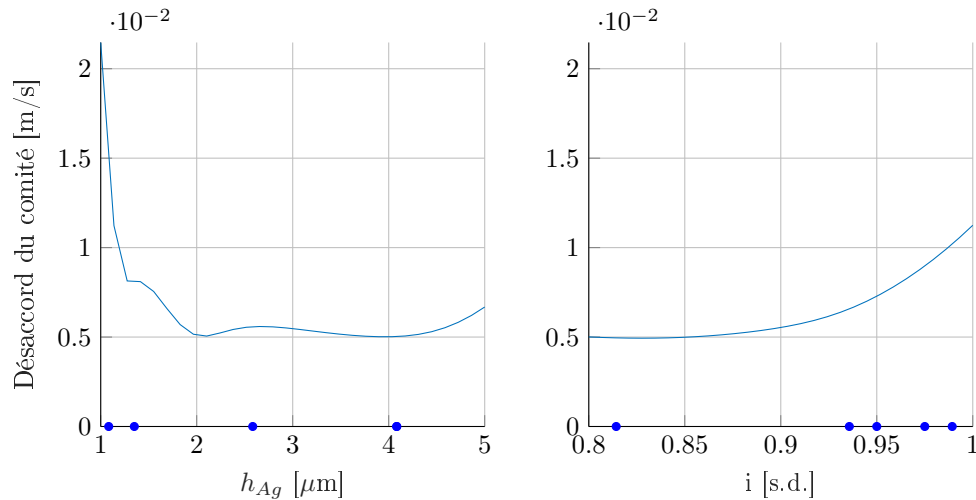


FIGURE 5.10 Désaccord du comité en fonction de l'épaisseur de stabilisateur d'argent et en fonction du ratio du courant appliqué sur le courant critique pour un ruban régulier. La courbe montre le désaccord du comité, et les points bleus sont les valeurs des 5 paires de points à ajouter pour le prochain entraînement.

montrées à la figure 5.11.

La figure 5.11 montre que le critère d'erreur sur le jeu de données tests de 2 cm/s est atteint après la troisième itération, soit lorsque le jeu d'entraînement contient 40 données. La RMSE sur la prédiction des jeux d'entraînement, de validation et de test est alors respectivement égale à 0.4 cm/s, à 0.94 cm/s et à 1.89 cm/s. Une itération supplémentaire a été faite pour améliorer la robustesse du modèle. La racine de l'erreur quadratique moyenne est un bon indicateur de la qualité du modèle, mais il est important de savoir comment les erreurs sont distribuées. La figure 5.12 montre la régression entre la NZPV calculée par éléments finis (É.F.) et la NZPV prédite par l'ensemble de modèles.

Les NZPV des rubans réguliers calculées par éléments finis et les prédictions sont très similaires. En effet, la figure 5.12 montre que les prédictions se situent toutes près de l'axe $y = x$, ce qui implique que la valeur de NZPV prédite est très similaire à la valeur trouvée par simulation. Les prédictions sur l'ensemble d'entraînement sont meilleures que celles sur les jeux de validation et de test puisque la RMSE est plus faible, mais c'est acceptable puisque la différence est moins d'un ordre de grandeur. Il y a donc un peu de sur-ajustement des poids sur les données d'entraînement. Ce sur-ajustement n'est pas un problème puisque les performances sur le jeu de test sont satisfaisantes.

La figure 5.13 montre les jeux de données initiaux, les données ajoutées par apprentissages actifs ainsi que les données de test. Cette figure montre que la densité de données est plus élevée lorsque l'épaisseur d'argent est plus faible et lorsque le ratio du courant est plus près de 1, ce qui est attendu. En effet, la figure 5.8 b) montre la disparité entre les données de la NZPV entre les données avec une épaisseur d'argent de 1 μm et les autres épaisseurs. Il faut donc plus de données dans cette zone pour pouvoir reproduire les variations rapides de la NZPV.

Pour l'entraînement final de l'ensemble, il y a 36 données pour l'entraînement, 9 données pour la validation et 10 données pour le jeu de test. Les prédictions sont présentées en trois dimensions à la figure 5.14. Cette surface montre comment les données d'entraînement, de validation et de test se situent par rapport aux prédictions du modèle.

La surface de prédiction de la figure 5.14 montre que l'ensemble de modèles est capable de prédire adéquatement la dynamique de la NZPV d'un ruban régulier sur une certaine plage de l'épaisseur de stabilisateur et du courant. L'ensemble de modèles peut être utilisé comme outil d'assistance à la conception de ruban supraconducteur. Avec 55 simulations par éléments finis, on peut maintenant prédire la NZPV de ce type de ruban sur l'intervalle de courant réduit $[0.8, 1]$ et l'intervalle d'épaisseur de stabilisateur $[1, 5]$ μm avec une erreur moyenne inférieure à 2 cm/s. L'avantage d'utiliser un ensemble de modèles au lieu de faire

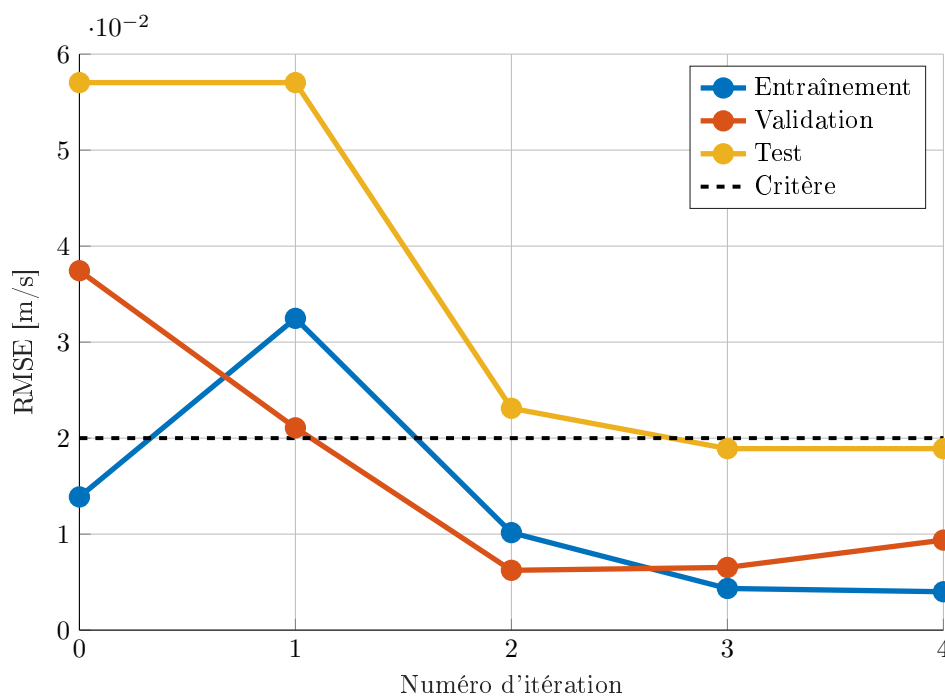


FIGURE 5.11 Évolution de l'erreur moyenne de la prédiction de la NZPV d'un ruban régulier par un ensemble de modèles en fonction du nombre d'itérations d'apprentissage actif.

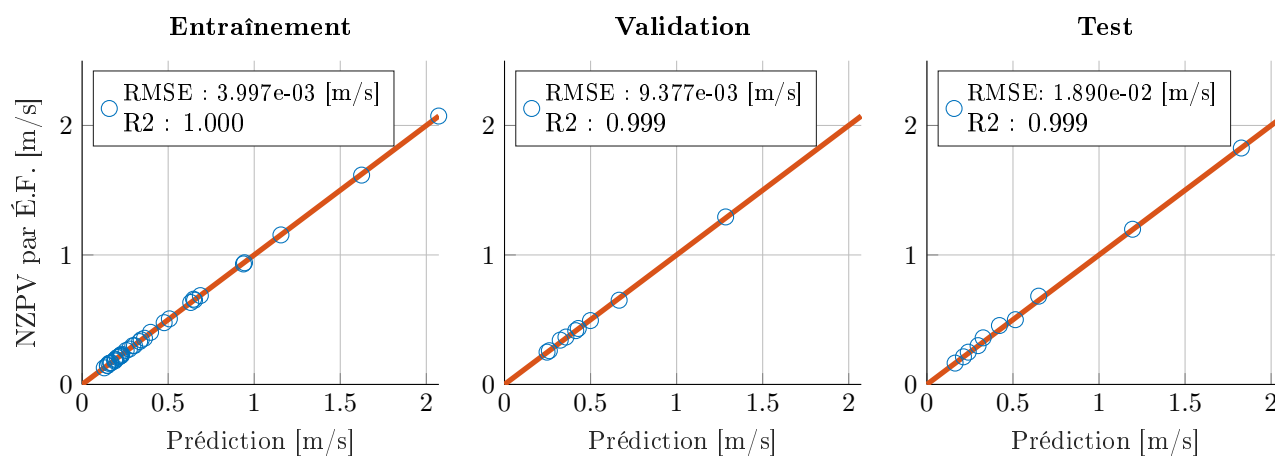


FIGURE 5.12 Résultats de prédiction en fonction des valeurs de NZPV calculées par éléments finis pour les rubans réguliers. Les modèles sont évalués en utilisant le coefficient de détermination R2 et la RMSE sur les jeux de données d'entraînement, de validation et de test.

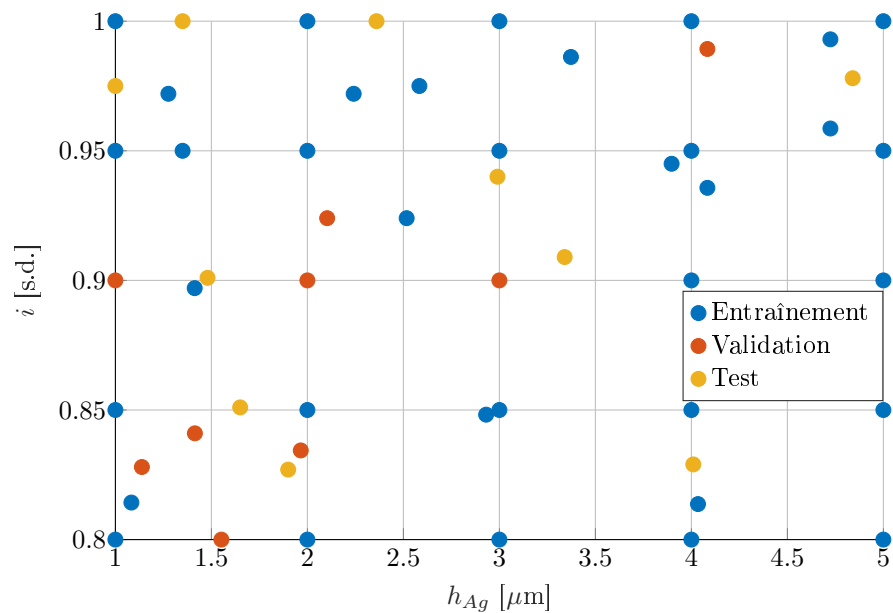


FIGURE 5.13 Distribution des données d'entraînement, de validation et de test dans le plan des entrées.

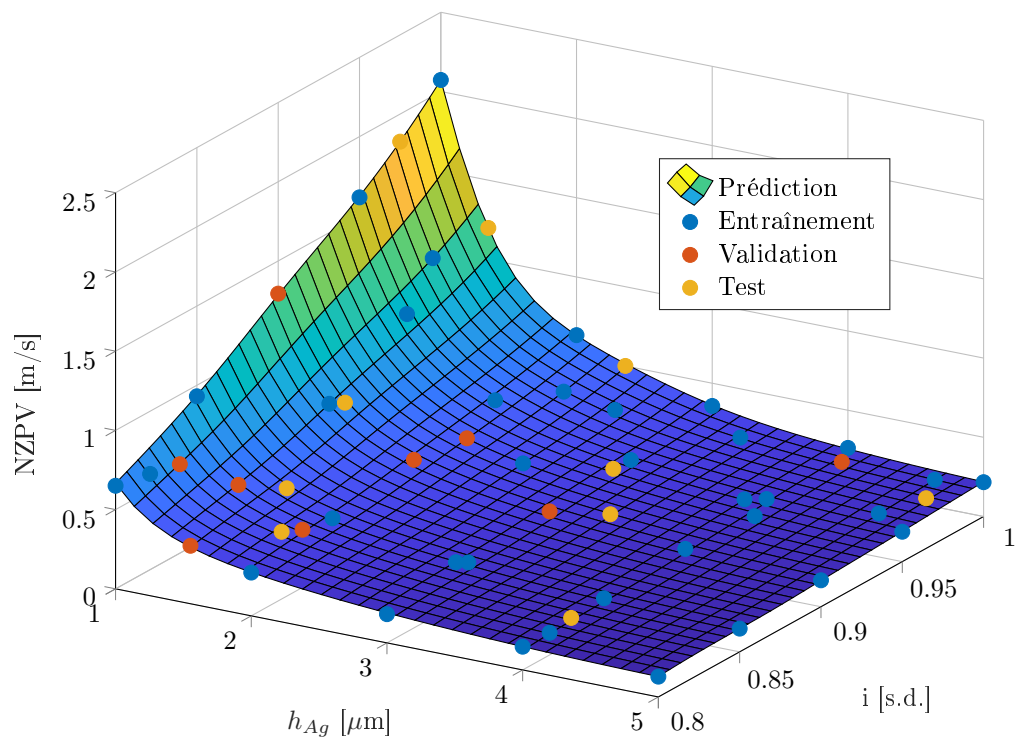


FIGURE 5.14 Surface de prédiction de la NZPV d'un ruban régulier par un ensemble de 8 réseaux de neurones en fonction de h_{Ag} et i .

une simulation par éléments finis en trois dimensions est que la prédiction de la NZPV prend $\sim 100 \mu\text{s}$ de temps de calcul, au lieu de plusieurs heures dans le cas d'une simulation par éléments finis.

5.3 Performance d'un modèle entraîné sur une simulation d'un ruban SuperPower CFD hybride

L'objectif de cette section est de montrer que la méthodologie fonctionne pour une simulation 3D plus complexe que pour le ruban régulier ou le modèle de Dresner. Parmi les paramètres d'entrées, on retrouve encore l'épaisseur d'argent et le courant, mais on ajoute un nouveau paramètre, soit la largeur de l'interface résistive divisé par la largeur du ruban f que l'on nomme le ratio de couverture. La figure 5.1 montre la géométrie du ruban appelé hybride. En pratique, les rubans de la compagnie SuperPower sont modifiés pour obtenir l'architecture hybride dans le but d'accélérer la NZPV.

Dans le cas du ruban hybride, l'épaisseur de stabilisation qui est variable est située au bas du ruban au lieu du dessus. Le stabilisateur sur le dessus et sur le côté sont d'une épaisseur fixe de $1 \mu\text{m}$ et celle du bas varie dans l'intervalle $[1, 5] \mu\text{m}$. De plus une résistance d'interface élevée est introduite sur une partie de l'interface entre le supraconducteur et l'argent, qui a pour objectif d'accélérer la NZPV. Les paramètres du ruban sont les mêmes que le ruban régulier présenté au tableau 5.1, à l'exception de l'épaisseur du substrat, qui est ici de $50 \mu\text{m}$. De plus,

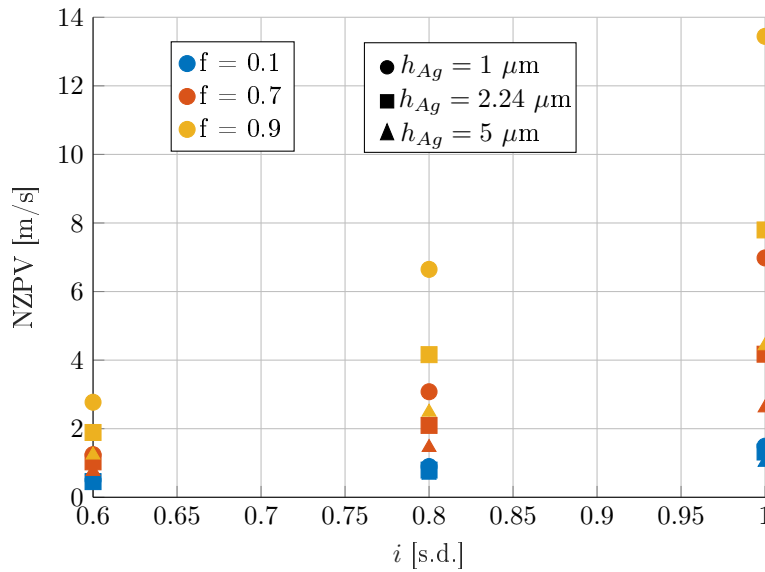


FIGURE 5.15 Distribution du jeu de données (h_{Ag}, i, f) initial servant au calcul de la NZPV d'un ruban CFD hybride.

le courant varie dans un intervalle différent, soit entre 60 et 100 A. Le paramètre du courant réduit i varie donc dans l'intervalle $[0.6, 1]$. Le ratio de couverture f varie dans l'intervalle $[0.1, 0.9]$. Lorsque f vaut 0, il n'y a pas de résistance d'interface entre le supraconducteur et l'argent. Lorsque f tend vers 1, alors la zone de faible résistance au centre du ruban devient de moins en moins large. Lorsque le courant fait la transition du supraconducteur vers l'argent le courant se concentre dans cette bande. Notons que la NZPV augmente rapidement lorsque f est près de 1.

Similairement à la simulation du ruban régulier, il y a deux plans de symétrie permettant de simuler uniquement le quart du ruban. La figure 2.5 a) montre la moitié du ruban selon la longueur, et la partie b) montre la coupe transversale du ruban, la moitié de celle-ci était simulée. Un jeu de données initial est généré avec 27 données. Il est constitué de toutes les combinaisons de h_{Ag} , i et f qui prennent chacune trois valeurs. L'épaisseur du stabilisateur du bas du ruban prend une des valeurs suivantes $h_{Ag} = [1.00, 2.24, 5.00] \mu\text{m}$, $i = [0.6, 0.8, 1.0]$ et f prend l'une des trois valeurs suivantes $[0.1, 0.7, 0.9]$. La figure 5.15 montre la distribution des données d'entraînement ainsi que la NZPV calculée à partir des simulations par éléments finis.

La NZPV de ce type de ruban atteint des valeurs plus grandes que pour le ruban régulier. La valeur maximale est de 13.443 m/s ou ~ 6.5 fois plus grande que la plus grande valeur de

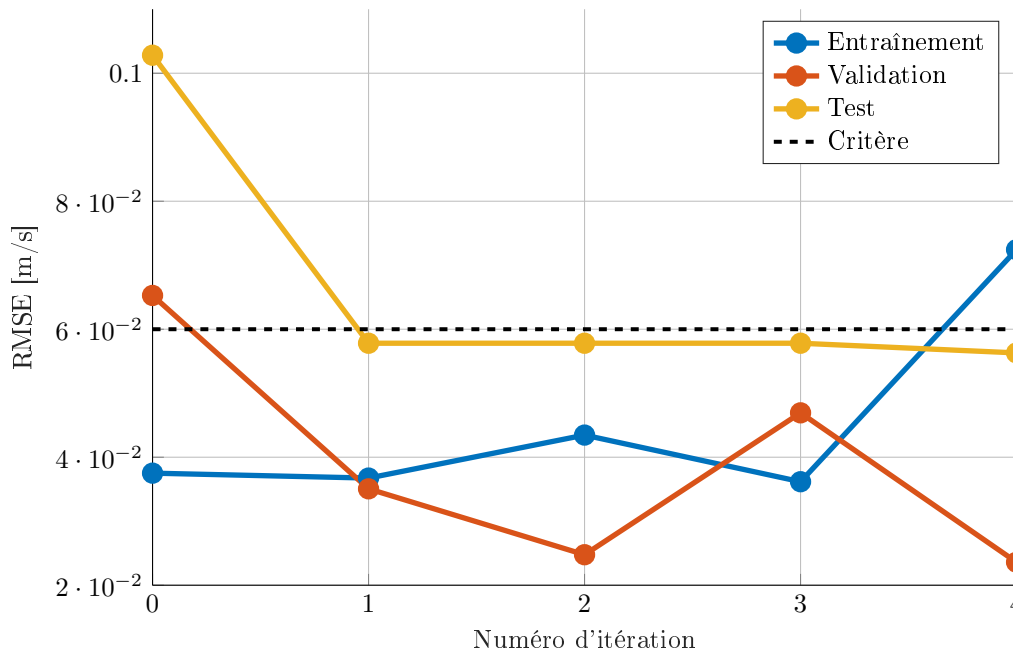


FIGURE 5.16 Évolution de l'erreur moyenne de la prédiction de la NZPV d'un ruban Hybride par un ensemble de modèles à chaque itération d'apprentissage actif.

NZPV du jeu de données des rubans réguliers. La valeur minimale de la vitesse de propagation de la zone normale est de 0.3737 m/s, ce qui est ~ 3 fois plus grande que la valeur minimale du jeu de données pour les rubans réguliers. Le critère d'arrêt est encore sur la RMSE du jeu de données test, mais il est adapté aux données extrêmes de ce jeu de données. Le critère pour arrêter la boucle d'apprentissage actif est donc $RMSE_{test} < 0.06$ m/s. Un critère plus restrictif peut être mis en place, mais cela implique qu'il faudrait générer plus de données et donc il faudrait faire plus de simulations par éléments finis.

5.3.1 Entraînement des modèles

Un ensemble de 8 réseaux de neurones est entraîné sur les données d'entraînement, 20% des données sont utilisées pour la validation et 80 % pour l'entraînement. Comme pour le problème du ruban régulier, il y a 10 autres données formant un jeu de test. À chaque itération, cinq données sont ajoutées jusqu'à ce que le critère sur l'erreur test soit atteint. La figure 5.16 montre que le critère est atteint après une seule itération et la racine de l'erreur quadratique moyenne change très peu après avoir atteint son critère.

La figure 5.17 montre la régression entre la NZPV calculée avec la simulation par éléments finis et la prédiction de l'ensemble de modèles après la 4e itération. Le coefficient de détermination sur chaque ensemble vaut pratiquement 1 pour les trois jeux de données. Les valeurs prédites et les valeurs calculées avec une simulation par éléments finis sont très similaires, cela explique la valeur de $R^2 = 1$. Les données utilisées pour l'entraînement, pour la validation et pour le test sont montrées à la figure 5.18. Les RMSE de la prédiction de la NZPV ne varient pas

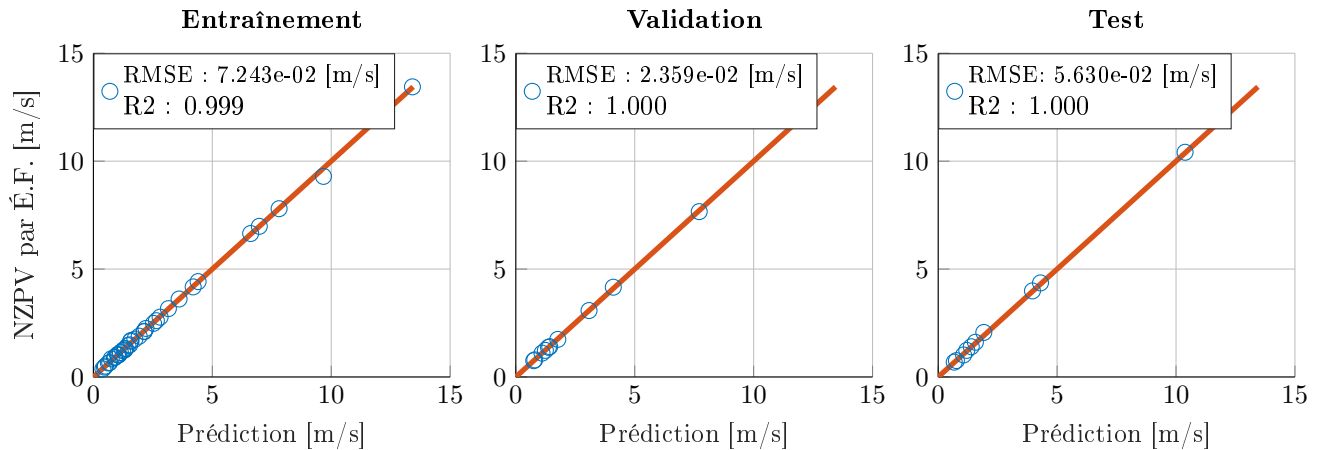


FIGURE 5.17 Résultats de prédiction en fonction des valeurs de NZPV calculées par éléments finis. Les modèles sont évalués en utilisant le coefficient de détermination R^2 et la RMSE sur les jeux de données d'entraînement, de validation et de test.

beaucoup entre les différents jeux de données, le modèle n'est donc pas sur-ajusté. Puisque la RMSE est faible et que le coefficient de détermination est près de 1, on peut affirmer que l'ensemble de modèles est capable de prédire la NZPV d'un ruban hybride selon les trois paramètres h_{Ag} , i et f dans la plage de valeurs présentée plus tôt.

La figure 5.19 montre trois surfaces de prédiction dans l'espace. Chaque surface est associée à une valeur fixe de ratio de couverture (f). Puisque la NZPV dépend de trois paramètres, l'affichage est plus difficile que pour le ruban régulier. Pour faciliter l'affichage, les surfaces affichées sont celles associées aux trois valeurs initiales du paramètre f . Les prédictions selon ce paramètre sont continues, seulement l'affichage est discret. En effet, les données ajoutées par apprentissage actif ne sont pas limitées aux valeurs initiales de f , soit 0.1, 0.7 et 0.9. Le ratio de couverture peut prendre n'importe quelle valeur sur l'intervalle $[0.1, 0.9]$. Cela implique que certains points ne sont pas sur l'une des trois surfaces de la figure 5.19, car les valeurs de f associée à ces données ne sont pas celles des trois surfaces affichées. La figure 5.19 montre que les surfaces de prédictions sont lisses et qu'elles ont un comportement similaire, mais c'est la figure 5.17 qui montre le mieux la qualité des prédictions.

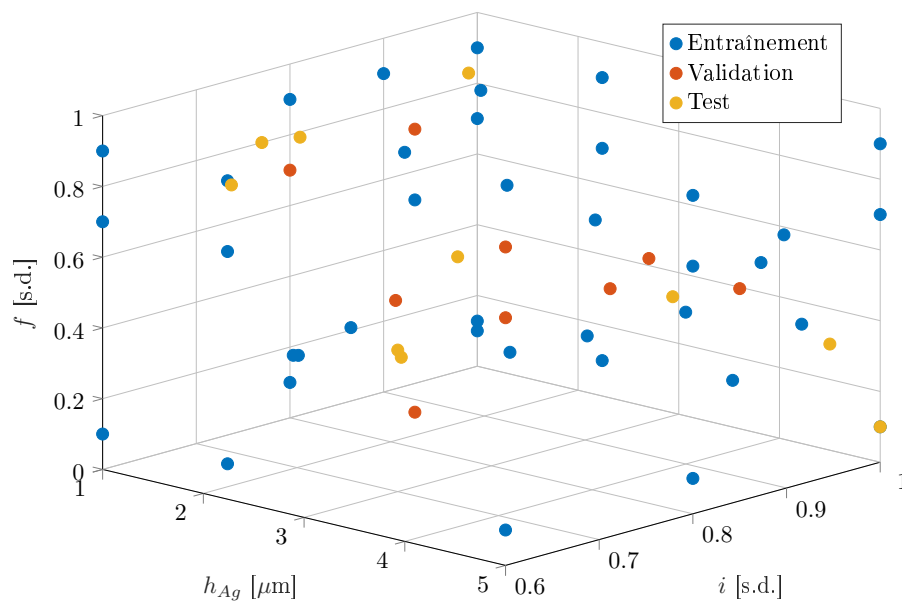


FIGURE 5.18 Distribution des données d'entraînement, de validation et de test dans l'espace des entrées.

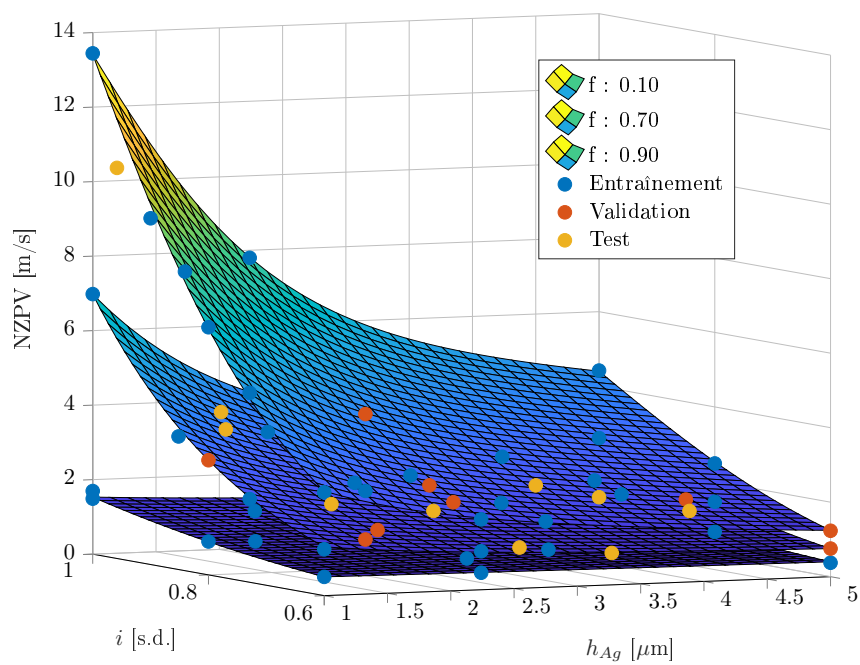


FIGURE 5.19 Surface de prédiction de la NZPV d'un ruban hybride par un ensemble de 8 réseaux de neurones en fonction de h_{Ag} , i et f .

CHAPITRE 6 CONCLUSION

6.1 Synthèse des travaux

Le but de ce projet de maîtrise était de développer une méthodologie pour construire un jeu de données fournissant suffisamment d'information pour entraîner un modèle de régression avec une erreur de prédiction donnée et de trouver quel modèle de régression est le plus adapté pour prédire la vitesse de propagation de la zone normale d'un ruban 2G HTS CC. La méthodologie a été développée sur des données générées avec le modèle analytique de la NZPV développée par Lawrence Dresner. L'évaluation de la qualité des jeux de données a été réalisée en utilisant deux méthodes : le test Gamma et la performance de modèles entraînés sur ces données. Ces deux méthodes ont permis de déterminer le résultat suivant : les données trouvées par apprentissage actif donnent de meilleures performances aux modèles entraînés sur elles que les données distribuées uniformément. Avec le modèle de Dresner, nous avons montré la corrélation entre la distribution de l'erreur sur un jeu de données test et la distribution du désaccord du comité. Cela a permis d'utiliser l'apprentissage actif avec la méthode de requête par comité, qui utilise la différence des prédictions des modèles d'un ensemble afin de choisir les données à étiqueter. On a déterminé que les modèles les plus adaptés pour cette problématique sont les réseaux de neurones, pour deux raisons. Premièrement, ces modèles ont eu de meilleures performances que les autres modèles de régressions classiques. Deuxièmement, les réseaux de neurones présentent une grande flexibilité lorsqu'il s'agit d'incorporer des paramètres d'entrée supplémentaires. Le modèle de prédiction final est un ensemble de 8 réseaux de neurones.

Après avoir déterminé la méthodologie d'acquisition de données et sélectionné le modèle de prédiction, on a testé la méthodologie sur le problème de la prédiction de la vitesse de propagation de la zone normale sur des jeux de données générés à partir de simulation par éléments finis de ruban régulier et pour des rubans de type CFD hybrides. Le jeu de données initial pour le problème avec le ruban régulier était constitué de 25 données, où chaque donnée était l'épaisseur de stabilisateur sur la surface du ruban et le ratio du courant appliqué divisé par le courant critique. Après quatre itérations d'apprentissages actif où 5 données étaient ajoutées à chaque itération avant que le modèle prédise la NZPV où la racine carrée de l'erreur quadratique moyenne a atteint une RMSE de 1.89 cm/s sur le jeu de données test contenant 10 données. Le jeu d'entraînement finale contenait 45 données. Pour le problème avec le ruban CFD hybride, il y avait trois paramètres d'entrées, soit l'épaisseur de stabilisateur sous le ruban, le ratio du courant appliqué et de courant critique et le ratio entre la largeur

de l'interface résistive entre le supraconducteur et le stabilisateur du dessus divisé par la largeur du ruban. Il y avait 27 données dans le jeu d'entraînement initial et 10 données dans le jeu de test. Un critère de 6 cm/s a été atteint après une seule itération, soit lorsque le jeu de données d'entraînement avait 32 données.

6.2 Limitations de la solution proposée

La principale limite de la méthodologie est que deux modèles entraînés sur le même jeu de donnée peuvent offrir des prédictions et des performances différentes. L'ajustement des paramètres des réseaux de neurones utilise une méthode stochastique, il arrive donc que certain entraînement forme des modèles qui ont une faible performance. Les modèles entraînés avec peu de données sont moins robustes que ceux entraînés avec un grand nombre de données. L'utilisation d'ensemble de modèles réduit l'impact d'un modèle moins performant, mais il peut être préférable de recommencer l'entraînement du début.

Un autre limite de la méthodologie est qu'elle a été développée sur trois jeux de données de vitesse de propagation de la zone normale de ruban supraconducteur. Il est possible que pour un autre type de problème, le modèle offrant les meilleurs performance ne soit pas un réseau de neurones.

6.3 Améliorations futures

La prochaine étape de ce projet serait de vérifier que la méthodologie fonctionne pour des problèmes autres que la prédiction de la vitesse de la zone normale. Cette méthodologie peut être utilisée par exemple pour la prédiction du champ magnétique d'un aimant en fonction de la géométrie de celui-ci et de la position d'un senseur.

La méthodologie pourrait également être améliorée en incluant l'optimisation du modèle dans la boucle d'entraînement. Les réseaux de neurones ont une limite sur leurs capacités de représenter la dynamique d'un problème. Si le nombre de neurones est trop petit, le modèle peut manquer de flexibilité pour représenter la dynamique en dessous d'un certain critère d'erreur. Ajoutez plus de données ne permettraient pas d'atteindre le critère. Il serait nécessaire d'augmenter le nombre de neurones ou le nombre de couches lors de l'entraînement. Cela est intéressant puisque les petits réseaux de neurones ont généralement de meilleures performances lorsqu'il y a peu de données que les gros réseaux de neurones, car les réseaux avec beaucoup de neurones peuvent sur-ajuster les données d'entraînement plus facilement. Cette flexibilité supplémentaire permettrait d'utiliser un petit réseau de neurones lorsqu'il y a peu de données disponible et d'ajuster celui-ci à mesure que le nombre de données augmente.

Il serait également intéressant d'étudier le problème d'apprentissage par fenêtres (*Patches learning*), où plusieurs modèles sont entraînés sur différents intervalles de domaine d'intérêt. Chaque fenêtre ferait des prédictions sur son intervalle, ce qui pourrait permettre d'obtenir une meilleure précision.

RÉFÉRENCES

- [1] D. S. Holmes, A. L. Ripple et M. A. Manheimer, “Energy-efficient superconducting computing—power budgets and requirements,” *IEEE Transactions on Applied Superconductivity*, vol. 23, n^o. 3, p. 1 701 610–1 701 610, jun 2013.
- [2] H. K. Onnes, “The resistance of pure mercury at helium temperatures,” *Commun. Phys. Lab. University. Leiden*, vol. 120, 1911.
- [3] D. van Delft et P. Kes, “The discovery of superconductivity,” p. 6.
- [4] M. Tinkham, *Introduction to superconductivity*, 2^e éd. Harvard, US : McGraw-Hill, Inc., 1996.
- [5] V. L. Ginzburg, “On the theory of superconductivity,” *Il Nuovo Cimento*, vol. 2, n^o. 6, p. 1234–1250, dec 1955.
- [6] J. Bardeen, “Theory of the meissner effect in superconductors,” *Physical Review*, vol. 97, n^o. 6, p. 1724–1725, mar 1955.
- [7] J. Bardeen, L. N. Cooper et J. R. Schrieffer, “Microscopic theory of superconductivity,” *Physical Review*, vol. 106, n^o. 1, p. 162–164, apr 1957.
- [8] —, “Theory of superconductivity,” *Physical Review*, vol. 108, n^o. 5, p. 1175–1204, dec 1957.
- [9] L. N. Cooper, “Bound electron pairs in a degenerate fermi gas,” *Physical Review*, vol. 104, n^o. 4, p. 1189–1190, nov 1956.
- [10] Z. Stekly, J. Zar *et al.*, “Stable superconducting coils,” *IEEE Transactions on Nuclear Science*, vol. 12, n^o. 3, p. 367–372, 1965.
- [11] X. Wang, U. P. Trociewitz et J. Schwartz, “Critical current degradation of short YBa₂Cu₃O_{7-δ} coated conductor due to an unprotected quench,” *Superconductor Science and Technology*, vol. 24, n^o. 3, p. 035006, dec 2010.
- [12] D. Zhang, M. D. Sumption, M. Majoros, C. Kovacs, E. W. Collings, D. Panik, M. Rindfleisch, D. Doll, M. Tomsic, C. Poole et M. Martens, “Quench, normal zone propagation velocity, and the development of an active protection scheme for a conduction cooled, react-and-wind, mgb2 mri coil segment,” *Superconductor Science and Technology*, vol. 32, n^o. 12, p. 125003, oct 2019.
- [13] C. Poole, T. Baig, R. J. Deissler, D. Doll, M. Tomsic et M. Martens, “Numerical study on the quench propagation in a 1.5 t mgb2 mri magnet design with varied wire compositions,” *Superconductor Science and Technology*, vol. 29, n^o. 4, p. 044003, mar 2016.

- [14] J. H. Durrell et N. A. Rutter, "Importance of low-angle grain boundaries in YBaCuO₂/subcoated conductors," *Superconductor Science and Technology*, vol. 22, n^o. 1, p. 013001, nov 2008.
- [15] D. C. Larbalestier, J. Jiang, U. P. Trociewitz, F. Kametani, C. Scheuerlein, M. Dalban-Canassy, M. Matras, P. Chen, N. C. Craig, P. J. Lee et E. E. Hellstrom, "Isotropic round-wire multifilament cuprate superconductor for generation of magnetic fields above 30 T," *Nature Materials*, vol. 13, n^o. 4, p. 375–381, mar 2014.
- [16] G. A. Levin, K. A. Novak et P. N. Barnes, "The effects of superconductor stabilizer interfacial resistance on the quench of a current-carrying coated conductor," *Superconductor Science and Technology*, vol. 23, n^o. 1, p. 014021, dec 2009.
- [17] C. Lacroix et F. Sirois, "Concept of a current flow diverter for accelerating the normal zone propagation velocity in 2G HTS coated conductors," *Superconductor Science and Technology*, vol. 27, n^o. 3, p. 035003, jan 2014.
- [18] J. C. e. F. S. C. Lacroix, Y. Lapierre, "High normal zone propagation velocity in second generation high-temperature superconductor coated conductors with a current flow diverter architecture," *Superconductor Science and technology*, vol. 27, n^o. 3, p. 055013, 2014.
- [19] J. Giguère, C. Lacroix, F. Dupuis-Desloges, J.-H. Fournier-Lupien et F. Sirois, "High normal zone propagation velocity in copper-stabilized 2G HTS coated conductors," *Superconductor Science and Technology*, vol. 34, n^o. 4, p. 045010, feb 2021.
- [20] C. Lacroix et F. Sirois, "Impact of current flow diverter on innovative HTS tape architectures for DC fault current limitation at electric fields up to 150 V/m," *IEEE Transactions on Applied Superconductivity*, vol. 29, n^o. 5, p. 1–7, aug 2019.
- [21] J.-H. Fournier-Lupien, C. Lacroix, S. Hellmann, J. Huh, K. Pfeiffer et F. Sirois, "Use of the buffer layers as a current flow diverter in 2G HTS coated conductors," *Superconductor Science and Technology*, vol. 31, n^o. 12, p. 125019, nov 2018.
- [22] G. Levin, W. Jones, K. Novak et P. Barnes, "The effects of superconductor–stabilizer interfacial resistance on quenching of a pancake coil made out of coated conductor," *Superconductor Science and Technology*, vol. 24, n^o. 3, p. 035015, 2011.
- [23] A. M. Miri, S. Fink et W. H. Fietz, "Transient behaviour of superconducting magnet systems of fusion reactor ITER during safety discharge," *Modelling and Simulation in Engineering*, vol. 2008, p. 1–10, 2008.
- [24] M. Ristic, J. V. M. McGinley et F. Lorenzoni, "Numerical study of quench protection schemes for a MgB₂ superconducting magnet," *IEEE Transactions on Applied Superconductivity*, vol. 21, n^o. 5, p. 3501–3508, oct 2011.

- [25] X. Du, M. Jin, Z. Zhang, Z. Xu, L. Ye, D. Zhang, X. Li, G. Zhang et L. Xiao, “Numerical analysis on the quench process and protection of conduction cooled mgb2 magnet,” *IEEE Transactions on Applied Superconductivity*, vol. 20, n^o. 3, p. 2102–2106, jun 2010.
- [26] L. Wang, H. Pan, X. L. Guo, H. Wu, S. X. Zheng et M. A. Green, “Study on the mechanical instability of MICE coupling magnets,” *IEEE Transactions on Applied Superconductivity*, vol. 21, n^o. 3, p. 2363–2366, jun 2011.
- [27] W.-K. Chan, P. J. Masson, C. A. Luongo et J. Schwartz, “Influence of inter-layer contact resistances on quench propagation in YBa₂Cu₃O_x coated conductors,” *IEEE transactions on applied superconductivity*, vol. 19, n^o. 3, p. 2490–2495, 2009.
- [28] W. K. Chan et J. Schwartz, “A hierarchical three-dimensional multiscale electro-thermal model of quenching in REBa₂Cu₃O_{7-δ} coated-conductor-based coils,” *IEEE Transactions on Applied Superconductivity*, vol. 22, n^o. 5, p. 4 706 010–4 706 010, oct 2012.
- [29] L. Dresner, “Propagation of normal zones in thermally insulated superconductors,” dans *Advances in Cryogenic Engineering Materials*. Springer US, 1980, p. 647–653.
- [30] B. Maddock, G. James et W. Norris, “Superconductive composites : Heat transfer and steady state stabilization,” *Cryogenics*, vol. 9, n^o. 4, p. 261–273, aug 1969.
- [31] V. Altov, M. Kremlev, V. Sytchev et V. Zenkevitch, “Calculation of propagation velocity of normal and superconducting regions in composite conductors,” *Cryogenics*, vol. 13, n^o. 7, p. 420–422, jul 1973.
- [32] L. Dresner, “Analytic solution for the propagation velocity in superconducting composites,” *IEEE Transactions on Magnetics*, vol. 15, n^o. 1, p. 328–330, jan 1979.
- [33] K. R. Müller, A. J. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen et V. Vapnik, “Predicting time series with support vector machines,” p. 999–1004, 1997.
- [34] C. K. I. W. Carl Edward Rasmussen, *Gaussian Processes for Machine Learning*, T. Dietterich, édit. Massachusetts Institute of Technology, 2006.
- [35] K. Hornik, M. Stinchcombe et H. White, “Multilayer feedforward networks are universal approximators,” *Neural Networks*, vol. 2, n^o. 5, p. 359–366, jan 1989.
- [36] C. Schran, K. Brezina et O. Marsalek, “Committee neural network potentials control generalization errors and enable active learning,” *The Journal of Chemical Physics*, vol. 153, n^o. 10, p. 104105, sep 2020.
- [37] L. Breiman, “Bagging predictors,” *Machine Learning*, vol. 24, n^o. 2, p. 123–140, aug 1996.
- [38] J. Han, J. Pei et M. Kamber, *Data Mining : Concepts and Techniques*, ser. The Morgan Kaufmann Series in Data Management Systems. Elsevier Science, 2011. [En ligne]. Disponible : <https://books.google.ca/books?id=pQws07tdpjoC>

- [39] B. Settles, “Active learning literature survey,” University of Wisconsin–Madison, Computer Sciences Technical Report 1648, 2009.
- [40] H. S. Seung, M. Oppen et H. Sompolinsky, “Query by committee,” dans *Proceedings of the fifth annual workshop on Computational learning theory*, 1992, p. 287–294.
- [41] R. M. Balabin et S. V. Smirnov, “Interpolation and extrapolation problems of multivariate regression in analytical chemistry : benchmarking the robustness on near-infrared (NIR) spectroscopy data,” *The Analyst*, vol. 137, n^o. 7, p. 1604, 2012.
- [42] A. Stefánsson, N. Končar et A. J. Jones, “A note on the gamma test,” *Neural Computing & Applications*, vol. 5, n^o. 3, p. 131–133, sep 1997.
- [43] D. Evans et A. J. Jones, “A proof of the gamma test,” *Proceedings : Mathematical, Physical and Engineering Sciences*, vol. 458, n^o. 2027, p. 2759–2799, 2002. [En ligne]. Disponible : <http://www.jstor.org/stable/3560031>
- [44] M. H. Center. (2023) Train and apply multilayer shallow neural networks. [En ligne]. Disponible : <https://www.mathworks.com/help/deeplearning/ug/train-and-apply-multilayer-neural-networks.html#bss3311-17>
- [45] “Alliance de recherche numérique du canada,” juill. 2023.