



**HAL**  
open science

# Traitement de la mission et des variables environnementales et intégration au processus de conception systémique

Amine Jaafar

► **To cite this version:**

Amine Jaafar. Traitement de la mission et des variables environnementales et intégration au processus de conception systémique. Autre. Institut National Polytechnique de Toulouse - INPT, 2011. Français. NNT : 2011INPT0070 . tel-00646708v2

**HAL Id: tel-00646708**

**<https://theses.hal.science/tel-00646708v2>**

Submitted on 13 Oct 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Université  
de Toulouse

# THÈSE

En vue de l'obtention du  
**DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE**

**Délivré par :**  
Institut National Polytechnique de Toulouse (INP Toulouse)

**Discipline ou spécialité :**

Génie Electrique

---

**Présentée et soutenue par :**

Amine Jaafar

**le :** mercredi 28 septembre 2011

**Titre :**

Traitement de la mission et des variables environnementales et  
intégration au processus de conception systémique

---

**JURY**

Jean-Yves Tourneret, IRIT, Toulouse (Président)  
Eric Bideaux, AMPERE, Lyon et Stéphane Brisset, L2EP, Lille (Rapporteurs)  
Rochdi Trigui, IFSTTAR, Bron (Examineur)  
Julien Pouget, SNCF, Paris (Invité)  
Xavier Roboam et Bruno Sareni, LAPLACE, Toulouse (Directeurs de thèse)

---

**Ecole doctorale :**

Génie Electrique, Electronique et Télécommunications (GEET)

**Unité de recherche :**

LAPLACE, UMR CNRS-INPT-UPS n°5213

**Directeur(s) de Thèse :**

Xavier Roboam

Bruno Sareni

**Rapporteurs :**

Eric Bideaux, AMPERE, Lyon

Stéphane Brisset, L2EP, Lille



# Résumé

Ce travail présente une démarche méthodologique visant le « traitement de profils » de « mission » et plus généralement de « variables environnementales » (mission, gisement, conditions aux limites), démarche constituant la phase amont essentielle d'un processus de conception systémique. La « classification » et la « synthèse » des profils relatifs aux variables d'environnement du système constituent en effet une première étape inévitable permettant de garantir, dans une large mesure, la qualité du dispositif conçu et ce à condition de se baser sur des « indicateurs » pertinents au sens des critères et contraintes de conception. Cette approche s'inscrit donc comme un outil d'aide à la décision dans un contexte de conception systémique. Nous mettons en particulier l'accent dans cette thèse sur l'apport de notre approche dans le contexte de la conception par optimisation qui, nécessitant un grand nombre d'itérations (évaluation de solutions de conception), exige l'utilisation de « profils compacts » au niveau informationnel (temps, fréquence,...).

Nous proposons dans une première phase d'étude, une démarche de « classification » et de « segmentation » des profils basée sur des critères de partitionnement. Cette étape permet de guider le concepteur vers le choix du nombre de dispositifs à concevoir pour sectionner les produits créés dans une gamme.

Dans une deuxième phase d'étude, nous proposons un processus de « synthèse de profil compact », représentatif des données relatives aux variables environnementales étudiées et dont les indicateurs de caractérisation correspondent aux caractéristiques de référence des données réelles. Ce signal de durée réduite est obtenu par la résolution d'un problème inverse à l'aide d'un algorithme évolutionnaire en agrégeant des motifs élémentaires paramétrés (sinusoïde, segments, sinus cardinaux). Ce processus de « synthèse compacte » est appliqué ensuite sur des exemples de profils de missions ferroviaires puis sur des gisements éoliens (vitesse du vent) associés à la conception de chaînes éoliennes. Nous prouvons enfin que la démarche de synthèse de profil représentatif et compact accroît notablement l'efficacité de l'optimisation en minimisant le coût de calcul facilitant dès lors une approche de conception par optimisation.

## **Mots-clés**

Conception systémique, profils, variables environnementales, missions, classification, synthèse de profil, optimisation, algorithme évolutionnaire, gisement éolien, locomotive hybride, stockage.



# Abstract

This work presents a methodological approach aiming at analyzing and processing mission profiles and more generally environmental variables (e.g. solar or wind energy potential, temperature, boundary conditions) in the context of system design. This process constitutes a key issue in order to ensure system effectiveness with regards to design constraints and objectives. In this thesis, we pay a particular attention on the use of compact profiles for environmental variables in the frame of system level integrated optimal design, which requires a wide number of system simulations.

In a first part, we propose a clustering approach based on partition criteria with the aim of analyzing mission profiles. This phase can help designers to identify different system configurations in compliance with the corresponding clusters: it may guide suppliers towards “market segmentation” not only fulfilling economic constraints but also technical design objectives.

The second stage of the study proposes a synthesis process of a compact profile which represents the corresponding data of the studied environmental variable. This compact profile is generated by combining parameters and number of elementary patterns (segment, sine or cardinal sine) with regards to design indicators. These latter are established with respect to the main objectives and constraints associated to the designed system. All pattern parameters are obtained by solving the corresponding inverse problem with evolutionary algorithms. Finally, this synthesis process is applied to two different case studies. The first consists in the simplification of wind data issued from measurements in two geographic sites of Guadeloupe and Tunisia. The second case deals with the reduction of a set of railway mission profiles relative to a hybrid locomotive devoted to shunting and switching missions. It is shown from those examples that our approach leads to a wide reduction of the profiles associated with environmental variables which allows a significant decrease of the computational time in the context of an integrated optimal design process.

## **Keywords**

System design, profiles, environmental variables, missions, clustering, profile synthesis, optimization, evolutionary algorithms, wind cycles, hybrid locomotive, storage.



# Avant-propos

Les travaux de recherche présentés dans ce mémoire ont été effectués à l'Institut National Polytechnique de Toulouse, dans le laboratoire Plasma et Conversion d'Energie LAPLACE, au sein de du Groupe de Recherche en Energie Electrique et Systémique.

Tout d'abord, je voudrais remercier Monsieur *Maurice Fadel*, directeur du LAPLACE site ENSEEIHT pour l'accueil qu'il m'a réservé au sein du laboratoire.

Je voudrais également remercier tous les membres du jury :

Monsieur *Jean-Yves Tourneret*, Professeur des universités à l'Institut National Polytechnique de Toulouse, laboratoire IRIT, pour le grand plaisir qu'il m'a fait en présidant ce jury.

Monsieur *Eric Bideaux*, Professeur des universités à l'Institut National des Sciences Appliquées de Lyon, laboratoire AMPERE, pour l'honneur qui m'a fait par sa participation à l'évaluation scientifique de ce travail en tant que rapporteur.

Monsieur *Stéphane Brisset*, Maître de conférences-HDR à l'École Centrale de Lille, laboratoire L2EP, pour l'intérêt qu'il a porté à ce travail en acceptant de le juger comme rapporteur.

Monsieur *Rochdi Trigui*, Chercheur à Institut Français des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux IFSTTAR, pour s'être intéressé à ce travail et avoir accepté de siéger à ce jury.

Monsieur *Julien Pouget*, Chef de projet en Énergétique à la Direction de l'Innovation et de la Recherche de la SNCF, Paris, pour son intérêt à ce travail ainsi que les échanges que nous avons pu avoir en dehors du cadre de cette thèse.

Monsieur *Xavier Roboam*, Directeur de recherche au CNRS, d'avoir accepté de diriger ces travaux de thèse. Je voudrais lui exprimer toute ma gratitude pour la grande confiance qu'il m'a accordée durant ces trois années. Merci de m'avoir proposé un sujet très enrichissant et de m'avoir encouragé vers de nouvelles directions. Qu'il trouve dans ce travail le témoignage de ma reconnaissance, de mon profond respect et de mes vifs remerciements.

Monsieur *Bruno Sareni*, Maître de conférences-HDR à l'Institut National Polytechnique de Toulouse, laboratoire Laplace, pour l'encadrement de cette thèse. Je voudrais le remercier pour sa gentillesse et sa disponibilité. C'était un réel plaisir de travailler ensemble pendant ces trois ans. Qu'il soit assuré de ma gratitude, de ma grande estime et de mon profond respect.



Mes remerciements s'adressent aussi à :

L'ensemble du personnel de l'ENSEEIH et du laboratoire LAPLACE, enseignants chercheurs, doctorants, personnels administratifs et techniques pour la chaleureuse ambiance qu'ils ont su créer.

*Pascal Maussion*, Professeur à l'Institut National Polytechnique de Toulouse pour avoir accepté d'être mon tuteur pédagogique durant mes années de monitorat.

*Stéphane Astier*, Professeur à l'Institut National Polytechnique de Toulouse pour l'aide et les explications qu'il m'a apportées dans le domaine des énergies renouvelables.

*Huber Piquet*, Professeur à l'Institut National Polytechnique de Toulouse pour l'aide qu'il m'a apportée lors de mes premières années à l'ENSEEIH.

*Cristophe Turpin*, Directeur de Recherche CNRS pour l'aide qu'il m'a apportée lors de ma première année en tant qu'ATER à l'ENSEEIH.

*Kamel Charrada*, Professeur à l'Institut Préparatoire aux études d'ingénieur de Monastir, pour son aide précieuse, ses conseils et ses encouragements.

*Sami Hassine*, Maître assistant à l'École Nationale d'Ingénieurs de Monastir, pour m'avoir communiqué sa passion pour le génie électrique lors de mes études universitaires.

Ma gratitude va aussi à tous mes professeurs de l'École Nationale d'Ingénieurs de Monastir.

Je voudrais également remercier tous mes amis qui m'ont encouragé et suivi depuis longtemps : *Wafa, Fares, Zeineb, Ader, Madiha, Ahmed, Marwene, Moez, Rym, Paquita,...*

Je voudrais saluer et remercier chaleureusement mon oncle *Khaled*, sa femme *Dhekra* et mes petites sœurs *Dorra, Donia* et *Syrine* pour leur amour, leur aide et leur accueil.

J'associe à ces remerciements mon ami *Jacques* et sa femme *Michèle* pour leur encouragement, leur soutien, leur accueil et leur aide en relisant ce mémoire.

Mes pensées vont également à toute la famille *Jaafar, Khachlouf, Wali, Hdhiri, Aoun*.

Je remercie chaleureusement mes beaux parents et toute la famille *Mili*. Je ne saurais oublier ma fiancée *Meriem* qui partage ma vie au quotidien, qui m'a soutenu et supporté durant cette dernière période d'une manière indéfectible. J'admire sa patience et sa bonté.

Je voudrais saluer et remercier du fond du cœur mes deux chères sœurs *Ines* et *Imen*, leurs maris *Mongi* et *Taha* ainsi que leurs enfants, qui m'ont toujours soutenu et qui m'ont permis de faire ce bout de chemin dans les meilleurs conditions. Je vous souhaite une vie couronnée de santé, de prospérité et de bonheur.

Enfin, je dois ce travail à mes deux chers parents, à mes premiers maîtres, à qui je dois tout ce que je suis. Rien de ce que je peux dire ou écrire ne suffira à exprimer mon amour, mon admiration et ma reconnaissance. Vos dévouements, votre patience, vos affections et sacrifices m'ont toujours comblé. Vous êtes ma source de soutien et de réconfort... dans les bons, comme dans les moments difficiles... Merci d'avoir toujours été là pour moi, que Dieu vous protège et vous accorde une longue vie.

**« N'allez pas là où le chemin peut mener. Allez là  
où il n'y a pas de chemin et laissez une trace. »**

Ralph Waldo Emerson (1803-1882)



# Table des matières

<b>Introduction générale</b>	1
<b>Chapitre 1 : Sur l'intégration de la mission et variables environnementales au sein d'un processus de conception systémique</b>	
1.1 Introduction	5
1.2 Les fondements de la conception systémique	6
1.2.1 Approche systémique et notion de système	6
1.2.2 Règles fondamentales relatives à la caractérisation des systèmes	9
1.2.3 La modélisation systémique	9
1.2.4 Approches de conception séquentielles	11
1.2.5 Approche de conception simultanée	12
1.2.6 Conception simultanée par optimisation	13
1.3 Prise en compte de la mission et des variables environnementales : état de l'art	15
1.3.1 Représentation de la mission par un point de fonctionnement nominal	15
1.3.2 Extraction d'un chronogramme temporel dimensionnant	18
1.3.3 Utilisation de cycles tests normalisés	19
1.3.4 Représentation d'une variable environnementale ou d'une mission à l'issue d'une analyse statistique	20
1.4 Nouvelle approche de classification et de synthèse d'une mission représentative dimensionnante	24
1.4.1 Contexte	24
1.4.2 Indicateurs de caractérisation de la mission et des variables environnementales	26
1.4.3 Mission et variables d'environnement au cœur du système : un couplage bidirectionnel	27
1.4.4 La classification : un outil d'aide à la conception	30
1.5 1 <sup>er</sup> exemple de conception systémique : la locomotive Hybride	32
1.5.1 Contexte du projet PLATHEE	32
1.5.2 Traitement de la mission dans le cas d'une simple hybridation	34
1.6 Traitement de la vitesse du vent pour la conception d'une chaîne éolienne	39
1.7 Conclusion	43
<b>Chapitre 2 : Classification des missions et des variables environnementales</b>	
2.1 Introduction	45
2.2 Techniques de classification : état de l'art	47
2.2.1 Agrégation autour des centres mobiles	48

2.2.2	Classification hiérarchique .....	52
2.2.3	Classification mixte .....	56
2.3	Classification sans <i>a priori</i> sur le nombre de classes .....	58
2.3.1	Problématique liée au choix du nombre de classes .....	58
2.3.2	Algorithme de classification dérivé du RTS .....	60
2.3.3	Exemples d'illustration.....	64
2.4	Application aux variables environnementales des systèmes électriques .....	67
2.4.1	Classification des missions pour des systèmes ferroviaires hybrides multi-segments 67	
2.4.2	Classification des profils de vitesse du vent.....	71
2.5	Conclusion .....	74

### **Chapitre 3 : Synthèse de profils de missions et de variables environnementales**

3.1	Introduction.....	75
3.2	Processus de synthèse de profil de mission ou de variable environnementale .....	76
3.2.1	Motifs élémentaires pour la génération de profil .....	78
3.2.2	Codage des chromosomes .....	83
3.3	Génération d'un profil en fonction du type de motif .....	85
3.3.1	Cas de profils « académiques » .....	85
3.3.2	Cas du profil d'une mission ferroviaire.....	90
3.4	Capacité des motifs élémentaires à synthétiser un profil fictif à caractéristiques données .....	95
3.4.1	Résultats avec nombre de motifs fixe et choix de l'algorithme évolutionnaire .....	96
3.4.2	Résultats avec nombre de motifs variable.....	98
3.5	Exemples d'application.....	100
3.5.1	Application à une famille de missions ferroviaires .....	101
3.5.2	Application à un profil de vitesse du vent.....	107
3.6	Conclusion .....	121

### **Chapitre 4 : Application à la conception simultanée par optimisation d'une chaîne de traction hybride**

4.1	Introduction.....	123
4.2	Modélisation de la locomotive hybride.....	124
4.3	Modèle d'optimisation .....	129
4.3.1	Processus de conception simultanée par optimisation .....	130
4.3.2	Paramètres d'optimisation.....	130
4.3.3	Contraintes d'optimisation .....	132
4.3.4	Critère d'optimisation.....	133
4.4	Conception simultanée par optimisation de la locomotive hybride BB63000 .....	133
4.4.1	Analyse de la différence sur le coût annuel de gasoil .....	136
4.4.2	Analyse du coût annuel des batteries .....	138
4.5	De la classification à la conception optimale.....	140
4.5.1	Classification des missions.....	141

---

4.5.2 Synthèse des missions représentatives.....	143
4.5.3 Conception simultanée par optimisation.....	146
4.5.4 Comparaison des résultats de conception .....	149
4.6 Conception par optimisation de la source d'énergie principale .....	150
4.7 Conclusion.....	154
<b>Conclusion générale</b>	<b>157</b>
<b>Annexe A</b> Base de données des missions ferroviaires	163
<b>Annexe B</b> Hybridation des auxiliaires TGV POS	169
<b>Annexe C</b> Comptage des cycles avec la méthode Rainflow	177
<b>Annexe D</b> Algorithmes évolutionnaires	183
<b>Annexe E</b> Tests de paramétrage du motif élémentaire de type segment	193
<b>Annexe F</b> Choix de la puissance nominale du groupe Diesel	205
<b>Bibliographie</b>	<b>209</b>



## Introduction générale

La conception des systèmes énergétiques, électriques en particulier, est aujourd'hui de plus en plus conditionnée par des enjeux sociétaux liés à des politiques d'économie d'énergie et de réduction des impacts environnementaux. A cela, s'ajoutent de nombreuses exigences telles que l'encombrement et la masse, la durée de vie, la fiabilité, la qualité, la stabilité et la recyclabilité. De plus, la fin du vingtième siècle a été marquée par une évolution notable vers une évaluation plus complète des coûts sur la totalité du cycle de vie du système (coût de fabrication, coût de maintenance et coût d'usage, voire coût de démantèlement/recyclage). Face à ces nouvelles considérations, le concepteur est appelé à tenir compte plus finement de l'environnement dans lequel son système va évoluer. Il est alors primordial, voire impératif, d'évoluer vers une approche de « conception systémique », permettant d'une part d'appréhender les couplages entre constituants du système et facilitant d'une autre part l'intégration de la « finalité » et des « variables environnementales » du dispositif à concevoir, et ce dès les premiers stades de la conception.

Classiquement, la plupart des systèmes sont conçus selon une approche de « conception séquentielle », partant du choix d'une « architecture », effectuant, dans un second temps, le « dimensionnement » des éléments du système puis en optimisant dans une étape ultime la gestion du dispositif. Les nouvelles exigences exposées précédemment poussent les concepteurs vers une approche de « conception simultanée », intégrant véritablement l'ensemble des couplages de la conception systémique. Selon l'approche classique séquentielle, les systèmes sont souvent surdimensionnés, leurs composants étant soumis à une réflexion d'optimisation locale qui ne garantit pas l'optimalité de l'ensemble. La conception simultanée est une approche qui tient compte du système dans sa globalité où les caractéristiques fondamentales de ce dernier sont considérées de façon couplée [SAR 06]. Elle peut en particulier être traitée à l'aide de techniques d'optimisation, là où plusieurs niveaux de



difficultés sont conjointement associés au problème de conception. Partant de la formulation du cahier des charges, ces difficultés concernent la modélisation systémique, l'identification des contraintes de conception, les méthodes de résolution et enfin l'intégration de la « mission » (cycles de fonctionnement ou finalité) et des variables d'environnement influençant le système. Cette dernière difficulté est tout particulièrement cruciale vu que les durées des missions et les cycles d'évolution des variables d'environnement (température, vent, ensoleillement,...), sont généralement plus longs que les constantes de temps dominantes du système, conduisant à des temps de traitement (simulation, optimisation) importants qui pénalisent la recherche de solutions optimales. Cette difficulté devient d'autant plus critique dès lors que le système à concevoir doit satisfaire non pas à un profil (de mission ou/et de variable environnementale) unique, mais à une famille ou classe de profils de caractéristiques parfois différentes. En effet, dans une base de données de profils, il n'y a généralement pas un seul profil « dominant » vis-à-vis de l'ensemble des indicateurs de conception : le plus souvent, une partie d'un profil est « dimensionnante » pour un premier indicateur, tandis qu'un autre profil est particulièrement contraignant vis-à-vis d'un second indicateur, etc. Le choix d'un profil représentatif dimensionnant n'est alors pas évident et constitue une étape essentielle, « partie intégrante » du processus de conception systémique.

C'est dans cet esprit que se situe ce travail qui présente une démarche méthodologique visant le « traitement de profils » de mission et, plus généralement, de variables environnementales (mission, gisement, conditions aux limites). Cette démarche constitue la phase amont essentielle d'un processus de conception systémique. La « classification » et la « synthèse » des profils relatifs aux variables d'environnement du système constituent en effet une première étape incontournable permettant de garantir, dans une large mesure, la qualité du dispositif conçu et ce à condition de se baser sur des « indicateurs » pertinents au sens des critères et contraintes de conception. Ainsi, il existe à travers cette démarche un couplage entre environnement et système à concevoir en ce sens, d'une part, que les profils (environnement et mission) influent sur la conception mais que, d'autre part, les indicateurs (relatifs aux critères et contraintes de conception) permettent eux-mêmes de caractériser le profil traité.

Cette approche s'inscrit donc comme un outil d'aide à la décision dans un contexte de conception systémique. Nous mettons en particulier l'accent dans cette thèse sur l'apport de notre approche dans le contexte de la conception par optimisation qui, nécessitant un grand

nombre d'itérations (évaluation de solutions de conception), impose l'utilisation de « profils compacts » au niveau informationnel (temps, fréquence, statistique,...).

Ce manuscrit s'articule autour de quatre chapitres. Avant de détailler les travaux de recherche constituant la contribution personnelle de cette thèse, le premier chapitre présentera un état de l'art de la conception systémique. Dans un premier temps, nous décrivons brièvement la notion de système mettant l'accent sur l'intérêt d'évoluer vers une conception simultanée par optimisation. Dans un deuxième temps, nous effectuons un état des lieux concernant la prise en compte de la mission et des variables d'environnement dans les processus de conception. Ensuite, nous mettons en exergue les concepts de base de la nouvelle démarche méthodologique de traitement de la mission et des variables d'environnement que nous proposons dans ce manuscrit. Enfin, nous exposons les deux exemples d'application qui serviront, en fil rouge au sein de ce mémoire, d'illustration de notre propos : ces exemples concernent le traitement des missions ferroviaires pour la conception d'une locomotive hybride et le traitement d'un gisement éolien. Nous formulons, pour chaque exemple, l'ensemble des indicateurs pertinents vis-à-vis des critères et contraintes de conception. Ces deux exemples, volontairement très distincts, permettront d'illustrer la généricité de l'approche méthodologique proposée.

Le second chapitre est consacré à la classification des missions et des variables environnementales. Après avoir dressé l'état de l'art des techniques de classification les plus répandues, nous présentons un algorithme de classification sans *a priori* sur le nombre de classes. Ce dernier est basé sur un algorithme génétique avec une méthode de nichage permettant de résoudre le problème de classification en optimisant un critère de partitionnement associé à la série de données. Cet algorithme est ensuite validé par trois benchmarks « tests ». Enfin, nous appliquons cet algorithme de classification sur deux séries de données des systèmes réels. Il s'agit, dans le premier cas, de classer un ensemble de missions relatives à trois systèmes ferroviaires distincts et dans le deuxième cas à la classification des mesures de la vitesse du vent sur deux sites géographiques différents.

L'objectif du troisième chapitre est de développer un processus de « synthèse de profil compact », représentatif des données relatives aux missions et aux variables environnementales influençant le système à concevoir. Il s'agit donc de générer un « profil fictif » de durée réduite dont les indicateurs de caractérisation correspondent aux caractéristiques de référence des données réelles. Ce signal simplifié et compact est obtenu

par la résolution d'un problème inverse à l'aide d'un algorithme génétique en agrégeant des motifs élémentaires paramétrés (sinusoïde, segments, sinus cardinaux). Ce processus de « synthèse compacte » est appliqué ensuite sur des exemples de profils de missions ferroviaires puis sur des mesures de la vitesse du vent associées à la conception de chaînes éoliennes.

Enfin, le quatrième et dernier chapitre traite de la « conception simultanée » par optimisation d'une locomotive hybride économe en énergie et respectueuse de l'environnement (projet PLATHEE : PLAtte-forme pour Trains Hybrides Économes en Énergie et Respectueux de l'Environnement). Nous montrons que la classification permet d'évaluer l'intérêt d'un dispositif spécifique, dédié et optimisé par classe de profils par rapport à un dispositif capable de satisfaire l'ensemble des classes. Nous prouvons également que la démarche de synthèse de profil représentatif et compact accroît non seulement la pertinence mais contribue aussi notablement à l'efficacité de la conception par optimisation en minimisant le coût de calcul.

# Chapitre 1

## Sur l'intégration de la mission et variables environnementales au sein d'un processus de conception systémique

### 1.1 Introduction

Dans ce chapitre introductif, nous exposons la problématique d'intégration de la mission (finalité) et des variables environnementales dans un processus de conception systémique. Dans un premier temps, nous introduisons succinctement l'approche systémique qui s'avère bien adaptée à la conception de systèmes hétérogènes complexes issus de multiples disciplines de la physique. Nous mettons l'accent en particulier sur les interactions du système avec sa finalité et son environnement. Nous expliquons aussi l'importance de la modélisation systémique et du choix du bon niveau de modèles à sélectionner dans une base de modèles à niveau de granularité variable. Nous montrons ensuite que la complexité des systèmes ainsi que le fort couplage entre leurs éléments introduit l'intérêt d'une conception simultanée, approche souvent résolue par des outils d'optimisation et très sensible, comme toute démarche systémique, à la finalité et à l'environnement dans lequel le système évolue. Dans un deuxième temps, nous dressons un bref état de l'art relatif à la prise en compte de la mission et des variables environnementales dans la conception des systèmes énergétiques. Dans une troisième partie de ce chapitre, nous présentons les concepts de base relatifs à la nouvelle démarche de traitement de la mission et des variables d'environnement que nous proposons dans ce mémoire, ceci afin de converger vers une modélisation systémique à finalité et environnement intégrés. Il s'agit plus précisément de synthétiser un profil de mission ou de variable d'environnement de durée réduite. Alors que les profils classiquement utilisés sont souvent relatifs à la forme du signal, notre approche « orientée conception » établit un lien direct entre propriétés du profil synthétisé et les principaux indicateurs pertinents vis à vis de

la conception du système : pour cela, nous définirons un jeu d'indicateurs spécifiques du « pedigree » complet du système à concevoir. Une telle démarche met en exergue le fort couplage bidirectionnel entre ce profil « fictif dimensionnant » et le système à concevoir notamment à travers les indicateurs de conception. Enfin, nous présentons deux exemples concrets très différents de traitement de mission et de variables environnementales en vue de la conception. Le premier exemple s'inscrit dans le cadre du projet PLATHEE et concerne le traitement des missions ferroviaires pour la conception d'une locomotive hybride autonome et respectueuse de l'environnement. Le deuxième exemple porte sur le traitement de la vitesse du vent pour la conception d'une chaîne éolienne. Par le biais d'une expertise de chacun de ces deux systèmes dans leur environnement, nous avons défini les jeux d'indicateurs nécessaires et pertinents vis-à-vis des critères et contraintes de conception. Les deux exemples proposés à la fin de ce chapitre seront retenus comme « fils conducteurs » dans la suite de la thèse, pour illustrer la pertinence des méthodes développées.

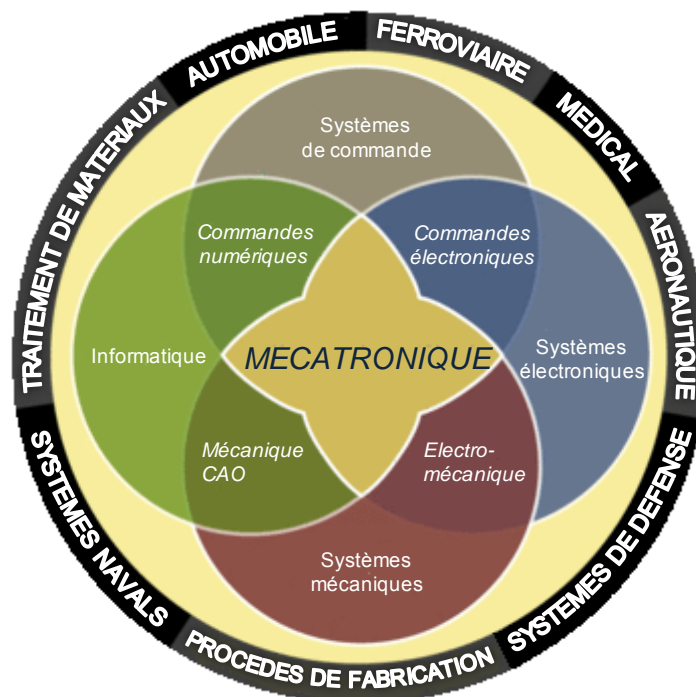
## **1.2 Les fondements de la conception systémique**

### **1.2.1 *Approche systémique et notion de système***

Apparues en 1945, suite à l'étude d'objets complexes dans un contexte multidisciplinaire, les méthodes systémiques n'ont connu leur maturité que vers la fin des années 70. Leur développement a été favorisé par l'arrivée des sciences de l'ingénierie avec la théorie cybernétique de *N. Wiener* (1948) et des sciences de la vie avec la théorie des systèmes ouverts de *L. V. Bertalanffy* (1951 et 1968) [BER 73]. L'approche a été ensuite enrichie par l'apparition des théories de l'information de *Shannon-Weaver* (1950) et a enfin été complétée par la théorie structuraliste de *Piaget* (1968) [ROS 75], [LEM 95]. La théorie cybernétique considère le système comme une « boîte noire » assurant une « finalité » ou « mission » dans un environnement sans nécessité de savoir la manière avec laquelle il réagit tandis que l'approche structuraliste consiste à décrire le comportement d'un système par sa structure, fonctionnant et évoluant [LEM 94]. Contrairement à l'approche analytique classique qualifiée de « mécaniste » et basée sur un traitement des éléments du système d'une manière isolée par champ de discipline, l'approche systémique considère le système dans sa globalité comme « un ensemble d'éléments organisés en fonction d'un but (mission), en interaction entre eux et avec l'environnement dans lequel il évolue » [LEM 96]. Selon l'approche systémique, la cohabitation (couplage) de plusieurs phénomènes physiques multidisciplinaires, spécifiques

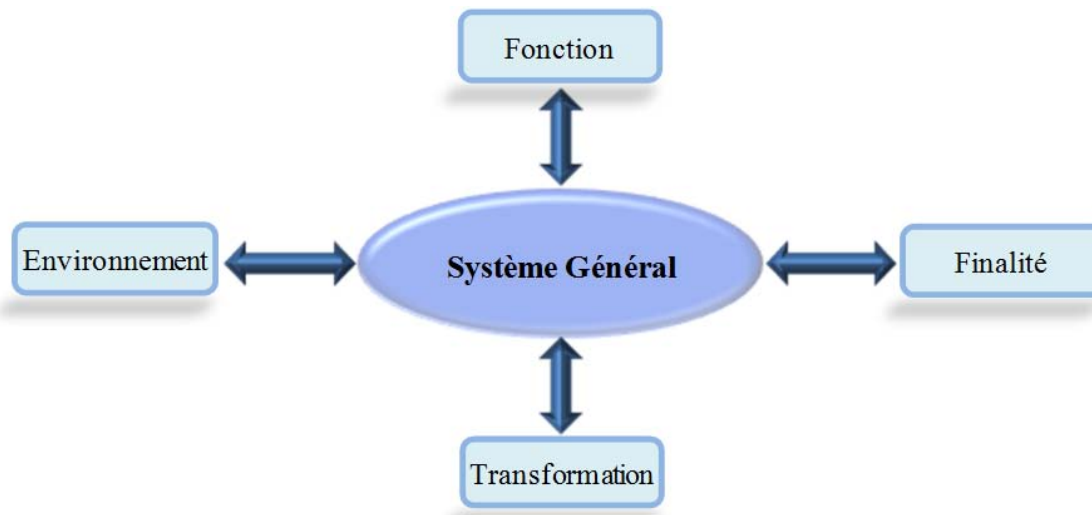
par élément ou objet, confère au système son caractère de complexité. Cette complexité est souvent liée à la caractérisation et à la représentation des sous-systèmes hétérogènes dans un même dispositif. Ainsi, selon *L. V. Bertalanffy*, un système est un « complexe d'éléments en *interaction* ».

Outre ce couplage inter-éléments, nous signalons aussi les fortes interactions entre disciplines, ainsi que le couplage intime entre le système et l'environnement dans lequel il évolue [REG 03]. Ces propriétés sont à la source du concept d'« intégration mécatronique », qui vise justement à coupler les problèmes physiques inhérents aux domaines mécanique et autres (électriques, thermodynamiques,...) avec l'électronique et l'informatique dédiée au contrôle commande et à la supervision (cf. Figure 1.1).



**Figure 1.1.** Illustration des divers couplages dans un concept d'« intégration mécatronique »

Dans sa théorie du système général, *Jean Louis Le Moigne* annonce que : « Le système général se décrit par une *action* (un enchevêtrement d'actions) dans *un environnement* (« tapissé » de processus) pour quelques projets (*Finalités, Téléologie*) fonctionnant (*faisant*) et se transformant (*devenant*) » (cf. Figure 1.2) [LEM 95]. Cette définition met en exergue la « *finalité* » et « *les variables d'environnement* » du système, notions fondamentales dans un contexte de conception systémique auxquelles nous accordons un intérêt particulier dans cette thèse.



**Figure 1.2.** *Forme canonique du système général* « au sens de Le Moigne » [LEM 95]

D'une manière générale, nous distinguons, dans un système donné, l'aspect fonctionnel de l'aspect structurel. Ce dernier est défini par l'architecture et relatif à l'organisation spatiale des éléments du système (limites, réservoirs d'énergie, stockage d'informations, réseau de communication,...) tandis que l'aspect fonctionnel se caractérise par une organisation temporelle (flux d'énergie, délais, boucles d'informations,...). Souvent, une représentation symbolique simple permet de décrire les diverses interactions entre les deux aspects [ROS 75]. Le fonctionnement de la structure est plus complexe que celui des composants du système pris séparément. De nouvelles spécificités émergent du système pris dans son ensemble, propriétés qui n'existent pas forcément au niveau composants : cette caractéristique est inhérente à la notion d'« émergence » [CHA 99].

Dans sa théorie du système général, *Jean Louis Le Moigne* nous signale qu'autour de la notion de système s'articulent des projets différents, principalement la conception, l'analyse et la simulation [LEM 94].

- Pour la conception, nous identifions les *finalités* dans un *environnement* avec une *structure* que nous faisons *fonctionner* puis *évoluer*.
- Pour l'analyse, nous partons des *fonctions* et des *évolutions* que nous observons dans un *environnement* et nous interprétons par rapport aux *finalités* afin de définir la *structure* qui répond aux besoins.
- Pour la simulation nous partons de la *structure* obtenue par conception ou analyse que nous faisons *fonctionner* et *évoluer* dans un *environnement* et nous comparons les résultats simulés aux *finalités* envisagées.

### 1.2.2 Règles fondamentales relatives à la caractérisation des systèmes

Afin de bien décrire le système et de représenter son fonctionnement et son comportement global d'une manière cohérente en tenant compte des différents niveaux de couplages existants, le concepteur doit se référer à certaines règles fondamentales. Ces règles s'articulent autour de l'aspect descriptif au sens de la systémique inhérent aux notions de modélisation que nous reprenons dans la sous-section suivante et de simulation utilisée comme un outil d'aide à la décision [REG 03], [CHA 99] :

- Définir les finalités ou les objectifs du système
- Délimiter le système dans son environnement en précisant les conditions aux limites
- Définir l'architecture du système et indiquer la nature, le rôle et les caractéristiques intrinsèques de ses composants ou sous-systèmes
- Modéliser les sous-systèmes en tenant compte de leurs dynamiques et de la cohabitation des phénomènes multidisciplinaires
- Paramétrer les sous-systèmes et s'assurer de la stabilité de l'ensemble
- Examiner les relations et les niveaux de couplage entre le système et sa finalité, son environnement, ses paramètres et enfin les contraintes et les critères de conception

### 1.2.3 La modélisation systémique

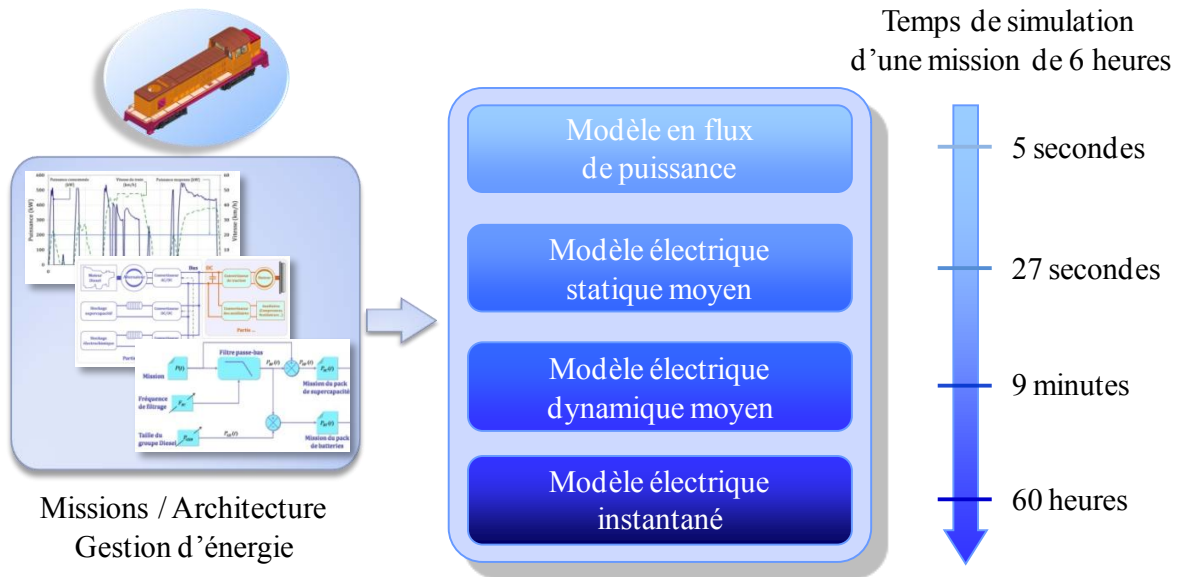
La démarche systémique est fondamentalement basée sur la notion de « modèle ». Selon *Jean Louis Le Moigne*, pour comprendre et donc pour donner du sens à un système complexe, nous devons le modéliser globalement pour *construire son intelligibilité* [LEM 95]. La modélisation permet donc de décrire les diverses interactions entre les composants d'un système à travers des processus spécifiques issus d'une certaine complémentarité entre l'observation empirique et le champ de la théorie. Dans son ouvrage « *Études systémique des actions de formations* », *Jean Beraum* définit la notion de modèle par : « ce fonctionnement est résumé par un modèle, qui peut être une description verbale ou schématique, éventuellement mathématique. L'objectif est de parvenir à une représentation permettant de prévoir les réactions du dispositif à des phénomènes éventuels ».

Compte tenu de la complexité du système et des divers couplages existants entre ses sous-systèmes, environnement et finalité, la modélisation systémique fait souvent appel à plusieurs



types de modèles à niveau de granularité variable. Par exemple, dans un contexte de conception systémique en génie électrique et énergétique, nous distinguons une panoplie de modèles allant de ceux liés au besoin énergétique (« modèles en flux de puissance ») jusqu'aux modèles de contraintes géométriques en tenant compte des modèles de performances dynamiques, de qualité et de fiabilité. Les modèles relatifs au besoin énergétique sont indispensables pour le dimensionnement du système : on distingue notamment les « modèles en flux de puissance » des « modèles effort-flux » statiques, qui permettent de valider la répartition des flux de puissance dans le contexte d'une architecture système (niveau de bus,...), puis dynamiques, relatifs aux boucles de contrôle commande nécessaires à la gestion de l'énergie [AKL 08]. La vérification des performances dynamiques fait appel à des « modèles effort-flux » dynamiques et aux formalismes mathématiques (modèle d'état par exemple) dédiés à la synthèse de commande. Les modèles de qualité relatifs à la forme d'onde dans un réseau d'énergie et à la compatibilité électromagnétique, consistent à vérifier certaines normes dans le plan fréquentiel (harmoniques) ou temporel (gabarit, par exemple pour les modèles de types CEM). Bien que plusieurs travaux de recherche soient actuellement en cours, la maturité complète des modèles de fiabilité n'est pas encore bien établie [RIC 04]. Toutefois, les phénomènes de vieillissement et de durée de vie des éléments de stockage d'énergie sont de plus en plus abordés en termes de modélisation [DUF 08]. Enfin, les modèles utilisés pour les critères géométriques (masse, volume) sont souvent de type analytique et peuvent faire appel aux notions de similitudes. Ils sont complétés par les modèles géométriques à éléments ou différences finies, qui traduisent par éléments la structure géométrique des objets.

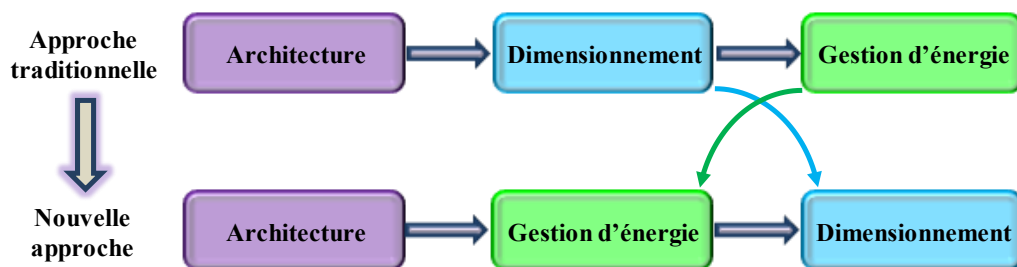
Il faut signaler que ces différents niveaux de modèles visent, bien sûr, différents objectifs de conception, offrent différents niveaux de précision, mais engendrent en retour des coûts de calcul très différents, dont certains (les derniers dans notre liste) peuvent devenir prohibitifs dans une optique de conception systémique. La [Figure 1.4](#) donne une idée sur les temps de calcul nécessaires pour simuler une mission de durée 6 heures avec les différents modèles établis pour la conception systémique d'une locomotive hybride et que nous les présenterons au [chapitre 4](#). Une grande partie de « l'art de l'homme système » consiste donc à choisir, à chaque phase de la conception, le « modèle juste assez précis ».



**Figure 1.3.** Temps de simulation d'une mission de durée 6 heures avec les différents types de modèles utiles à la conception d'une locomotive hybride

#### 1.2.4 Approches de conception séquentielles

De point de vue général, la conception d'un système énergétique s'articule autour de trois composantes fondamentales : l'« architecture », le « dimensionnement » et la « gestion de l'énergie ». Traditionnellement, les approches de conception classiques sont « séquentielles ». Elles consistent à déterminer en premier lieu l'architecture du système, ensuite à dimensionner ses composants et enfin à mettre en place une stratégie de gestion d'énergie afin de satisfaire son cahier des charges et sa finalité, c'est-à-dire la mission à laquelle il est dédié [AST 03] (cf. Figure 1.4). Néanmoins, dans certaines configurations telles que les systèmes multisources, le fait de considérer, conformément à la conception séquentielle, la gestion des flux d'énergie comme une conséquence de la phase de dimensionnement peut éventuellement conduire à un surdimensionnement des sources d'énergie [AKL 08]. Cette réflexion a conduit, dans de nombreux travaux, à adapter la conception en inversant les phases de dimensionnement et de gestion énergétique, comme le montre la Figure 1.4.

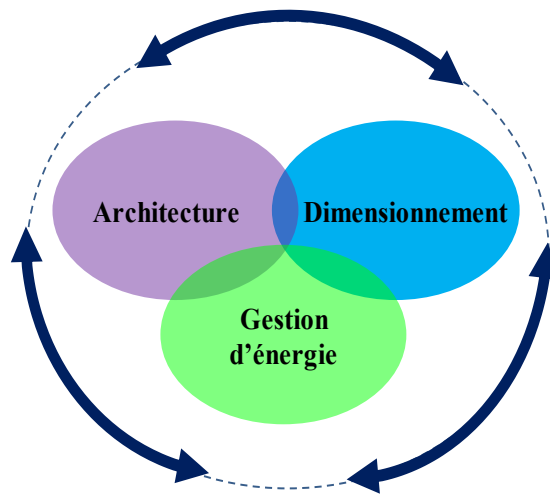


**Figure 1.4.** Approches de conception séquentielles

C'est alors à partir des parts de missions de chaque source d'énergie que ces dernières sont dimensionnées. Ceci permet d'accroître notablement l'efficacité de la conception en intégrant dans la gestion d'énergie les contraintes technologiques des briques énergétiques et les critères de faisabilité, de performance et de durée de vie.

### 1.2.5 Approche de conception simultanée

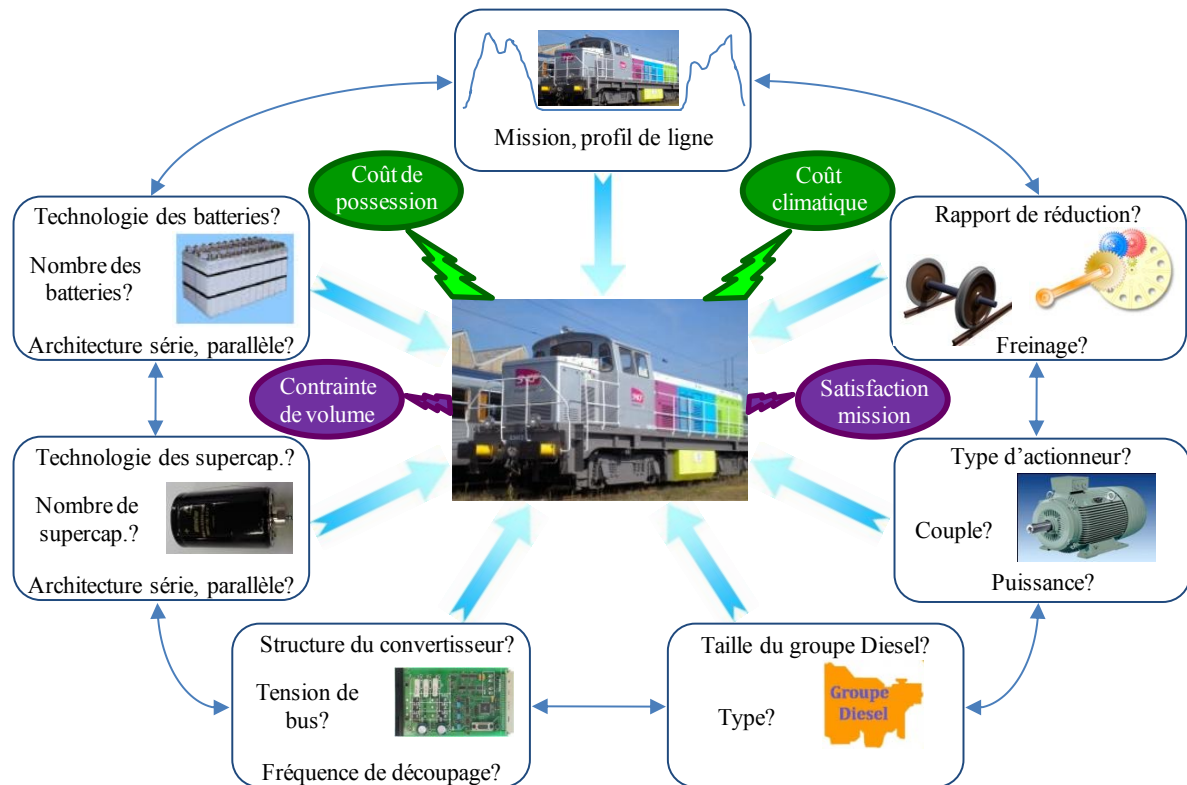
Même si dans la démarche de conception séquentielle, chaque étape est soumise à une réflexion d'optimisation locale, rien ne garantit l'optimalité du système dans son ensemble. En effet, traiter les phases de conception d'une manière séquentielle, ne permet pas de tenir compte des divers couplages existant entre les paramètres relevant de l'architecture, du dimensionnement ou de la gestion du système. Pour ne pas briser de tels couplages, il est souhaitable d'adopter une approche de « *conception simultanée* » [SAR 06] qui vise conjointement et en parallèle la détermination de l'architecture, du dimensionnement et de la gestion énergétique de l'ensemble des éléments comme le montre la [Figure 1.5](#).



**Figure 1.5.** *Approches de conception systémique simultanée*

A titre d'exemple, nous illustrons le cas de la conception d'une locomotive hybride autonome sur lequel nous reviendrons dans les parties suivantes et tout au long de la thèse. Le système présente une complexité élevée et une forte hétérogénéité en termes de couplages (cf. [Figure 1.6](#)). Les différents sous-systèmes (châssis, groupe Diesel, moteur électrique, batterie, supercondensateur, convertisseur statique, transmissions mécaniques, roue, dispositif de freinage,...) sont couplés mutuellement et interagissent fortement avec l'environnement (profil de ligne, signalisation, conducteur, température ambiante) et la finalité (missions ou

cycles de circulation). La conception de cette locomotive doit tenir compte d'un certain nombre de contraintes (masse, volume, satisfaction de la mission,...) et vise à optimiser le « coût de possession » (coûts d'investissement, de maintenance et d'exploitation) et le « coût climatique » (dioxyde de carbone, particules fines). Nous signalons aussi que le caractère multisources de cette locomotive hybride augmente notablement la complexité de sa conception par rapport à une locomotive tout électrique ou tout Diesel.



**Figure 1.6.** Divers couplages au sein d'une locomotive hybride diesel-électrique

### 1.2.6 Conception simultanée par optimisation

Comme nous l'avons signalé précédemment, la complexité du système amène souvent le concepteur à être confronté à des choix à effectuer simultanément sur les trois composantes fondamentales de la conception (architecture, dimensionnement et gestion d'énergie) pour optimiser véritablement les critères en satisfaisant la finalité et en respectant les contraintes de conception. Cette problématique peut être traitée efficacement à l'aide de techniques d'optimisation. Néanmoins, la complexité au niveau du système sera directement repercutée au niveau de l'outil d'optimisation en donnant lieu à des problèmes assez complexes (modèles de conception, convergence d'algorithme, analyse des tendances,...). Il s'agit souvent de

problèmes à grand nombre de variables de conception de nature discrète (paramètres liés aux choix de la structure et de l'architecture,...) ou continue (paramètres de dimensionnement, de gestion d'énergie), soumis à plusieurs contraintes et optimisant des critères hétérogènes (énergétiques, dimensionnels, environnementaux, économiques,...).

Dans [SAR 06], les diverses problématiques associées à la conception simultanée par optimisation ont été exposées. Souvent, le concepteur se trouve dans la difficulté de formaliser le cahier des charges. Il s'agit plus précisément d'identifier les architectures possibles, de définir les paramètres structurels et dimensionnels les plus significatifs au regard des critères de performances et de leur associer des stratégies de gestion d'énergie pertinentes pour satisfaire performance et finalité du système. De très nombreuses configurations sont donc envisageables avec différents degrés de liberté relatifs au paramétrage du système. Ce nombre de degré de libertés est régi par un compromis du fait qu'il doit être d'une part suffisant pour une meilleure exploration des solutions possibles et d'une autre part limité pour réduire le coût de complexité du problème d'optimisation. Une difficulté connexe à la formalisation du cahier des charges est associée à l'identification des contraintes de conception. Ces dernières sont généralement scindées en trois classes : les contraintes liées aux caractéristiques intrinsèques de chaque élément du système, les contraintes liées au système global et enfin les contraintes relatives à la compatibilité de la mise en interaction de divers éléments hétérogènes. Un autre problème générique est lié à la difficulté de la modélisation des sous-systèmes souvent multidisciplinaires et nécessitant donc un champ d'expertise très large. A cela, s'ajoute la prise en compte des diverses dynamiques des éléments du système, particulièrement de type énergétique, afin de bien simuler le fonctionnement du système et de, garantir une évaluation pertinente des critères énergétiques et des critères de performance. Comme nous l'avons signalé dans le [paragraphe 1.2.3](#), il est alors impératif de bien choisir le type et le niveau de granularité des modèles tout en garantissant une compatibilité à la fois d'ordre physique et logicielle dans leur association. La dernière difficulté est liée à l'intégration de la finalité ou de la mission et des variables environnementales du système. Cette problématique, faisant l'objet de notre étude, sera détaillée par la suite de ce chapitre.

On notera enfin que des approches intermédiaires entre la conception séquentielle et la conception simultanée telles que l'optimisation multidisciplinaire ou MDO (Multidisciplinary Design Optimization) offrent aussi des perspectives intéressantes dans un contexte de

conception systémique globale [MOU 09]. Ces approches font encore l'objet d'études notamment dans notre communauté [MES 08], [KRE 09].

### 1.3 Prise en compte de la mission et des variables environnementales : état de l'art

Dans [ROB 00], l'auteur porte une attention particulière sur le « *modèle global à finalité intégrée* ». Le développement d'un tel modèle s'avère indispensable lorsque l'on mène une démarche de conception systémique telle que la conception simultanée. Ce dernier consiste non seulement à modéliser la finalité ou la mission comme un élément du système global mais aussi les variables d'environnement constituant des parties intégrantes du système. Par exemple, la conception d'une chaîne énergétique éolienne ou solaire doit absolument intégrer à la fois les besoins énergétiques (missions relatives aux cycles de consommation) à satisfaire et les ressources climatiques (ensoleillement, température, vitesse du vent) offertes par l'environnement.

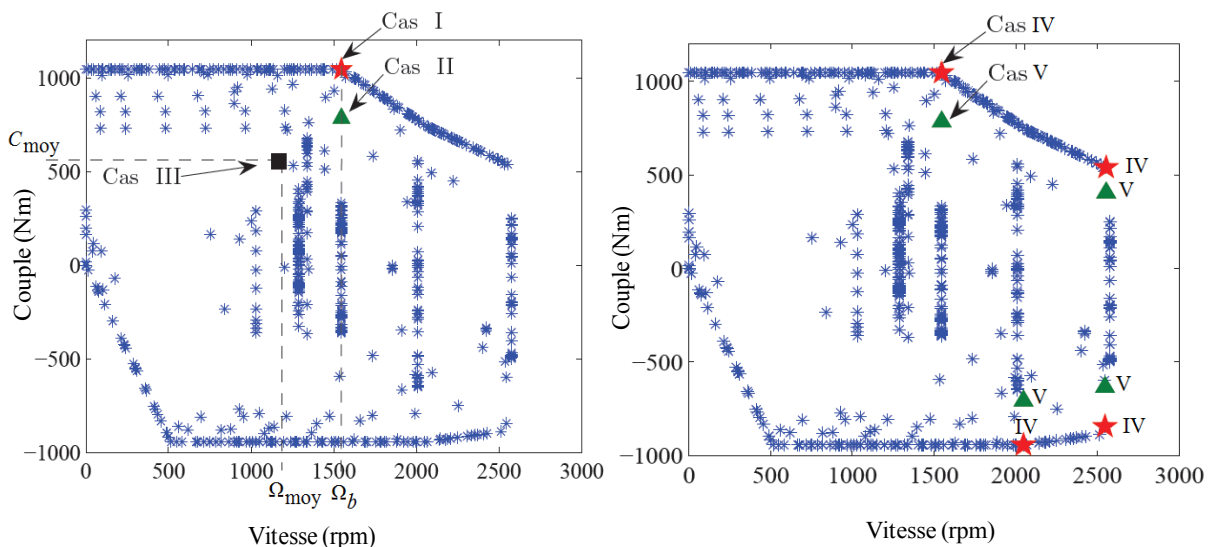
Dans cette section, nous proposons de faire un état de l'art sur l'intégration et la prise en compte des missions et des variables d'environnement dans la conception systémique, en génie électrique en particulier et plus généralement pour les dispositifs énergétiques multiphysiques.

#### 1.3.1 Représentation de la mission par un point de fonctionnement nominal

La conception des moteurs électriques est souvent abordée en considérant un dimensionnement sur un point de fonctionnement nominal [MES 98]. Ce point est facilement identifiable pour des applications industrielles où la charge est considérée comme constante et où la vitesse de « traction » est imposée. Cette approche relativement simple est immédiatement remise en cause dans le cas d'applications de traction électrique en régime variable où les points de fonctionnement suivent des trajectoires dans le plan couple-vitesse. La réduction de l'ensemble de la trajectoire en un seul point de fonctionnement nominal est peu évidente à déterminer *a priori* et parfois peu pertinente. Une simplification possible consiste à considérer pour le point de dimensionnement nominal, un point relatif au couple maximal et à la vitesse maximale des trajectoires, point qui domine énergétiquement l'ensemble des points de la trajectoire. Cette approche permet à coup sûr de répondre au problème du dimensionnement énergétique, mais conduit généralement à un surdimensionnement du système par rapport à l'ensemble des besoins. Par ailleurs, l'efficacité

énergétique du système se trouve optimisée dans une zone de fonctionnement proche du point nominal qui ne correspond pas nécessairement aux points de plus forte occurrence. Il est donc possible, dans une deuxième approche, de considérer pour le point de dimensionnement des points de fonctionnement particuliers (le point de plus forte occurrence par exemple) mais cette solution peut conduire à l'inverse à un dimensionnement qui ne permet pas de satisfaire l'ensemble de la trajectoire. Enfin, l'expérience récente du groupe de recherche GENESYS nous a montré que certains points dominés énergétiquement pouvaient s'avérer dimensionnant vis-à-vis de certains critères ou contraintes de conception : citons par exemple le cas de contraintes normatives de qualité des formes d'onde à l'entrée d'un système d'actionnement violées lors d'un fonctionnement basse vitesse pourtant peu « vorace en puissance ».

Dans [KRE 09], l'auteur donne une comparaison entre différentes solutions optimales de conception de moteurs de traction selon six spécifications du cycle de fonctionnement. Dans les cinq premières spécifications (I, II,...V), le moteur est dimensionné à partir d'un ou de plusieurs points de fonctionnement représentatifs d'un régime permanent (cf. Figure 1.7) alors que la dernière spécification (VI) représente la totalité du cycle de fonctionnement (cf. Figure 1.8). Cette dernière spécification, correspondant au cas du métro ou du tramway où le cycle de fonctionnement est entièrement connu, permet de prendre en considération le comportement thermique transitoire du moteur lors de sa conception [MES 06].

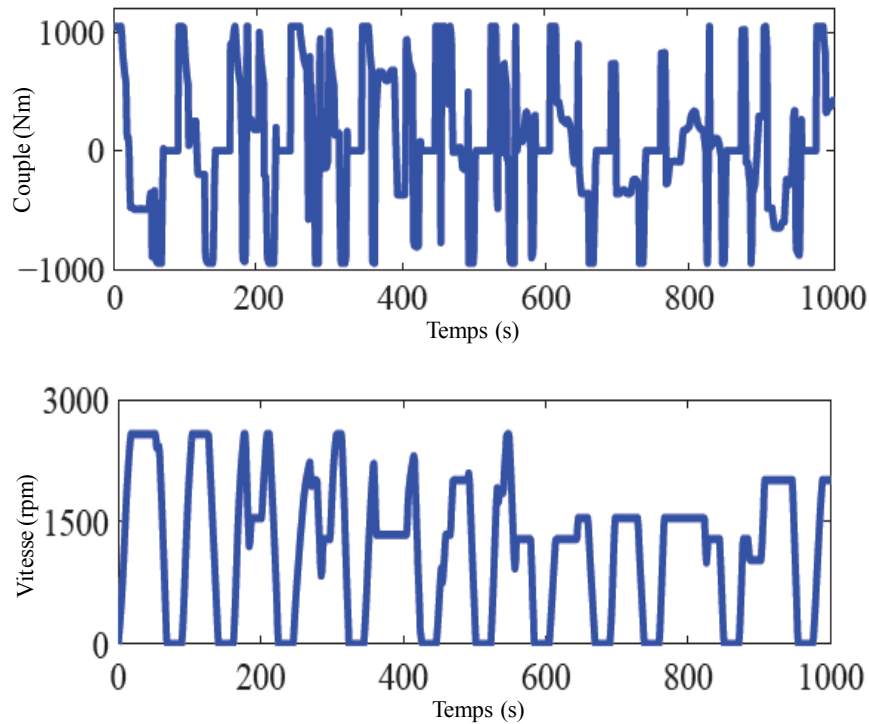


**Figure 1.7.** Spécifications I à V : Couples et vitesses sont imposés par le concepteur

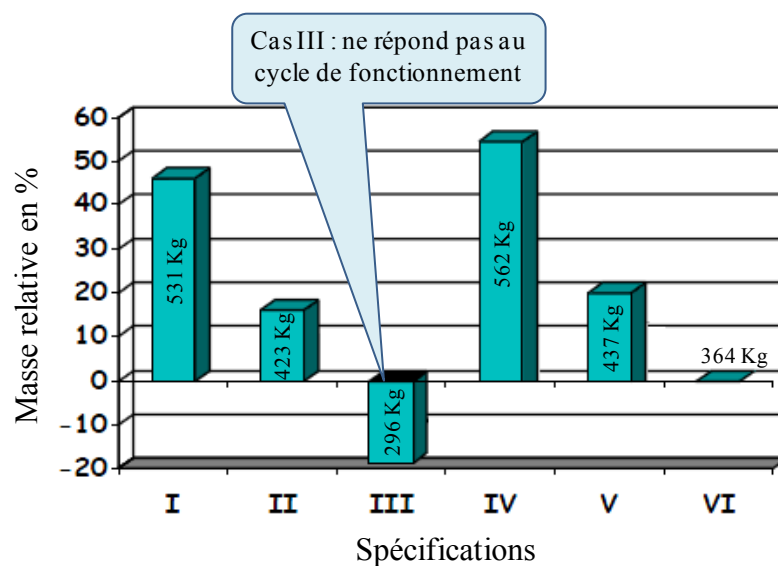
(Cas I : Couple maximal et vitesse de base, Cas II : 75% du couple maximal et vitesse de base,

Cas III : Couple moyen et vitesse moyenne, Cas IV : 4 points caractéristiques de la courbe couple vitesse, Cas V : 4 points caractéristiques à 75% de la courbe couple vitesse) [KRE 09]

Les résultats de dimensionnement du moteur de traction pour chacune de ces spécifications (cf. Figure 1.9) montrent que la solution de masse optimale est obtenue en considérant la totalité du cycle de fonctionnement (spécification VI). Ce résultat prouve l'importance cruciale d'intégrer l'ensemble des caractéristiques de la mission (ou cycle de fonctionnement) du système lors de sa conception.



**Figure 1.8.** Spécification VI : Couple et vitesse imposés par le système via le cycle de fonctionnement [KRE 09]



**Figure 1.9.** Comparaison des solutions optimales obtenues par rapport au cas VI (364 kg) [KRE 09]



### 1.3.2 Extraction d'un chronogramme temporel dimensionnant

Cet exemple s'inscrit dans le cadre de la conception d'un réseau de secours électrique pour l'aéronautique [LAN 06]. Il s'agit de dimensionner un réseau de secours hybride à base d'une turbine éolienne<sup>1</sup> RAT (*Ram Air Turbine*) comme source d'énergie assurant une puissance voisine de la puissance moyenne de la mission et un élément de stockage comme source de puissance afin de fournir la partie fluctuante. Dans ce cas, la mission est définie par la consommation de puissance des charges « essentielles » du réseau en cas de perte totale des moteurs. Un échantillon de mission de secours est ensuite établi indiquant les différentes phases du vol dans le cas d'une perte complète des moteurs (cf. Figure 1.10). Cet échantillon étant spécifié sur un intervalle de temps long (30 minutes environ), il peut s'avérer inutile, et de plus fastidieux, de raisonner sur l'ensemble de la mission car cela engendrerait des temps de simulations prohibitifs tout du moins si l'analyse du comportement dynamique du système est visée avec les niveaux de modélisation adéquats.

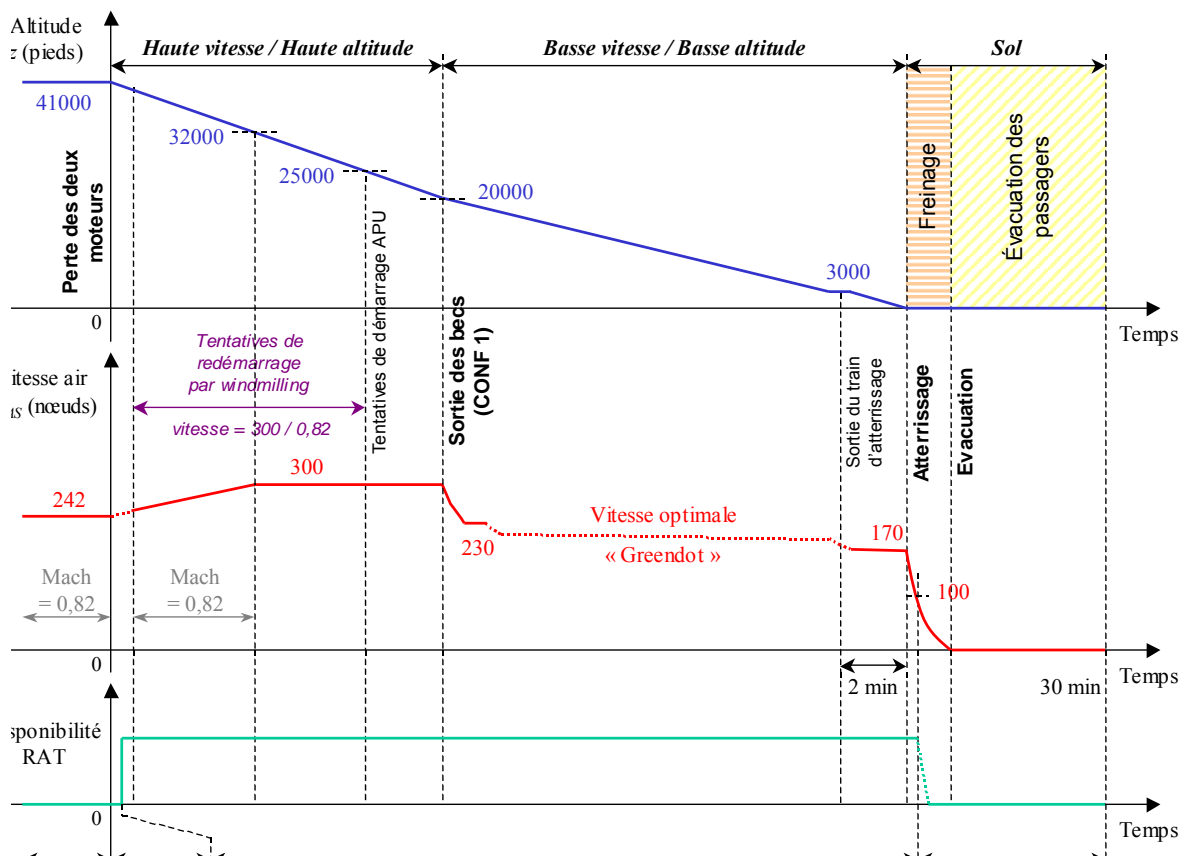
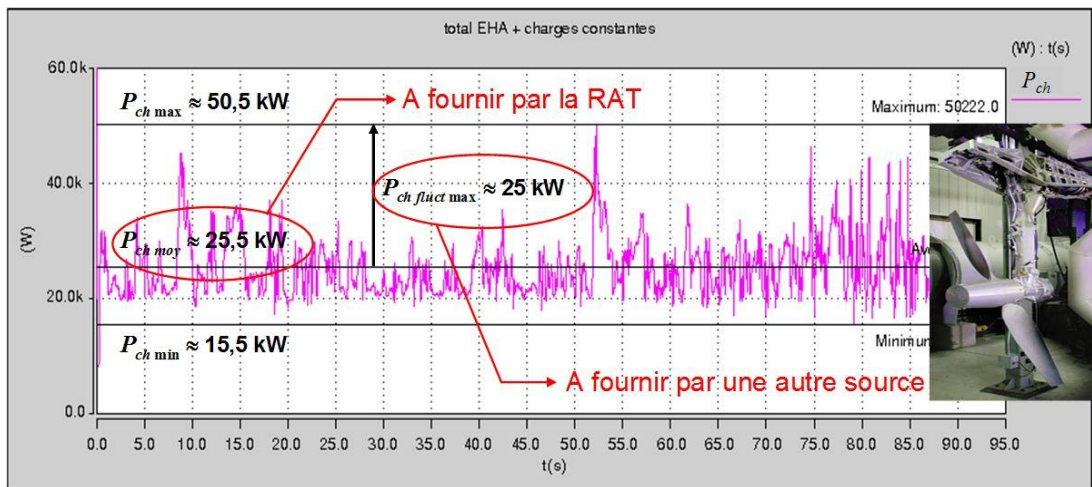


Figure 1.10. Échantillon de mission « secours aéronautique » en cas de perte totale des moteurs

<sup>1</sup> La turbine éolienne entraîne un générateur électrique par l'intermédiaire d'un multiplicateur de vitesse

Dans le cas de cette mission de secours, seul le besoin énergétique le plus critique a été extrait (cf. [Figure 1.11](#)) en localisant, par le biais de l'expertise, une phase considérée comme « dimensionnante », de durée raisonnable (moins de deux minutes) correspondant à une configuration basse altitude et basse vitesse en milieu turbulent. Cette phase particulière constitue un cas critique pour la source de secours. Il s'agit en effet d'une phase au cours de laquelle les charges principales alimentées en mode secours et les commandes de vol font l'objet de manœuvres fréquentes à puissance élevée tandis que le vent relatif vu par la turbine est d'intensité faible.

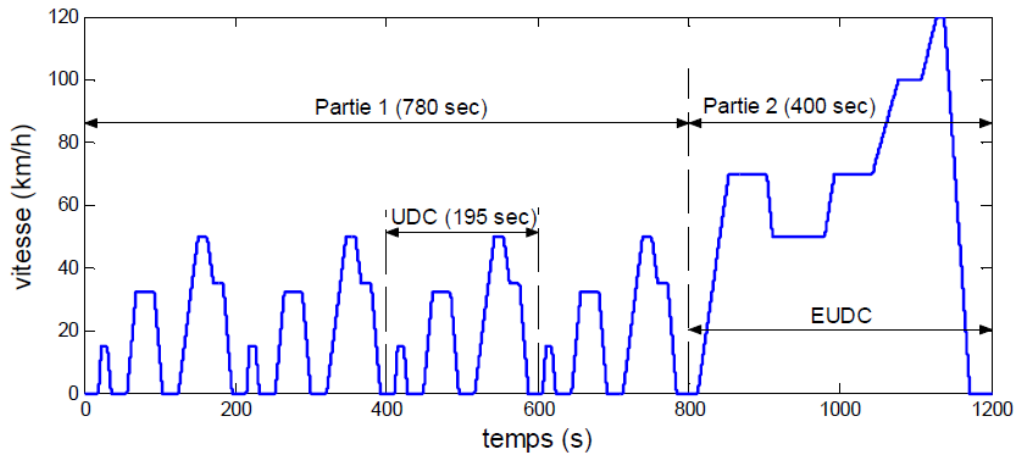


**Figure 1.11.** Extraction d'une mission dimensionnante

### 1.3.3 Utilisation de cycles tests normalisés

Dans un contexte écologique de plus en plus prégnant, l'Europe s'est impliquée récemment dans la normalisation des besoins, notamment dans le domaine des transports. Par exemple, dans le secteur de l'automobile, il existe des cycles normalisés de conduite (profils de vitesse en fonction du temps) afin de comparer, sur une même mission ou cycle de référence, la consommation et l'émission des particules polluantes des divers véhicules. Cette normalisation distingue le cycle urbain UDC (*Urban Drive Cycle*) du cycle extra-urbain EUDC (*Extra Urban Driving Cycle*) tels que définis par la directive européenne [EEC 99]. Le cycle UDC a été introduit pour modéliser ou représenter les conditions de conduite dans une grande ville, avec une faible sollicitation du moteur, une faible température des gaz d'échappement, une vitesse qui n'excède pas 50 km/h et une vitesse moyenne de 18,7 km/h (cf. [Figure 1.12 – UDC \(195 sec\)](#)). Le cycle EUDC caractérise les profils de type routier et correspond à une vitesse moyenne de 62,6 km/h, avec une vitesse limite de 120 km/h (cf. [Figure 1.12 – EUDC \(400 sec\)](#)). Le cycle le plus utilisé en Europe pour l'évaluation des

performances énergétiques et des impacts écologiques des véhicules légers est le NEDC (*New European Drive Cycle*), connu aussi par cycle MVEG-A (*Motor Vehicle Emissions Group*). Il est formé par une succession de quatre cycles UDC et un cycle EUDC comme le montre la Figure 1.12 [BOS 04].



**Figure 1.12.** Nouveaux cycles Européen de conduite mixte NEDC

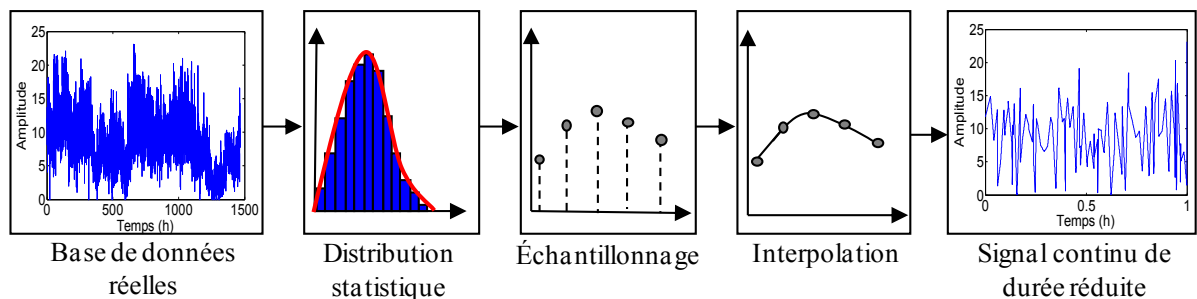
Ces cycles normalisés, acceptés dans le cas de l'automobile par l'ensemble des constructeurs européens, ne sont pas réellement significatifs de l'usage du véhicule, les contraintes d'accélération et les sollicitations du moteur étant plutôt sous évaluées par rapport à la réalité. Les véritables « cycles d'usage » sont souvent plus contraignants, ce qui impacte fortement sur la conception du système, en particulier en ce qui concerne la durée de vie des composants. De plus, ces cycles seront certainement à revoir entièrement avec l'arrivée de nouveaux véhicules électriques et hybrides qui intégreront de nouvelles contraintes liées à l'utilisation des batteries (cyclage, durée de vie), contraintes non nécessairement mises en évidence par les précédents cycles.

### 1.3.4 Représentation d'une variable environnementale ou d'une mission à l'issue d'une analyse statistique

Pour optimiser l'efficacité énergétique d'un système au regard des variables environnementales, la prise en compte des aspects statistiques s'avère indispensable. Nous distinguerons dans cette partie deux approches. La première est basée sur l'analyse de distribution statistique relative à la variable d'environnement considérée (typiquement histogramme de répartition de puissance). La seconde repose sur l'analyse des variables d'environnement à l'aide de méthodes de classification.

- *Exploitation des distributions statistiques des variables environnementales*

La prise en compte des occurrences statistiques relatives aux points de fonctionnement d'un système est primordiale si l'on souhaite que celui-ci soit optimisé au mieux pour l'ensemble de ces opérations. Nous avons déjà signalé au [paragraphe 1.3.1](#) qu'un dimensionnement en un point nominal permettant de respecter des contraintes d'énergie ou de puissance, ne garantissait en aucun cas une efficacité énergétique « optimale » et un dimensionnement optimal sur l'ensemble d'une mission. Pour remédier à ce problème, et généraliser des approches de dimensionnement à un ensemble de points de missions, nous pouvons en première approximation considérer la mission comme une succession de régimes statiques pondérés par leurs occurrences. L'inconvénient de cette méthode est l'absence de prise en compte des aspects dynamiques liés aux transitions entre les différents points de fonctionnement. Toutefois, si ces phénomènes ne sont pas prépondérants dans les bilans globaux, elle offre un intérêt tout particulier vis-à-vis de la minimisation des temps de calculs. On notera par ailleurs qu'il est tout à fait possible de synthétiser des échantillons temporels à partir d'occurrences statistiques si l'on souhaite inclure les aspects dynamiques. Cette approche a notamment été utilisée pour la génération de signaux de vent en éolien. Le gisement éolien associé à un site donné peut être caractérisé en termes d'occurrences statistiques à partir de mesures de vitesses de vent sur de longues périodes. Habituellement, les gisements sont représentés par des lois de Weibull [KRE 09]. Il est alors possible de générer pour la variable de vent un signal temporel de durée réduite à partir de la distribution statistique [ABD 07]. L'approche est basée sur un générateur de nombre aléatoire ayant comme densité de probabilité la distribution statistique ainsi déterminée. La séquence des échantillons obtenus est ensuite interpolée pour construire le signal temporel (cf. [Figure 1.13](#)). Nous reviendrons dans le [chapitre 3](#) sur les détails de cette approche dans un contexte de génération d'un profil représentatif de la vitesse du vent.



**Figure 1.13.** Génération d'un signal temporel continu issu d'une distribution statistique

- *Utilisation de méthodes de classification*

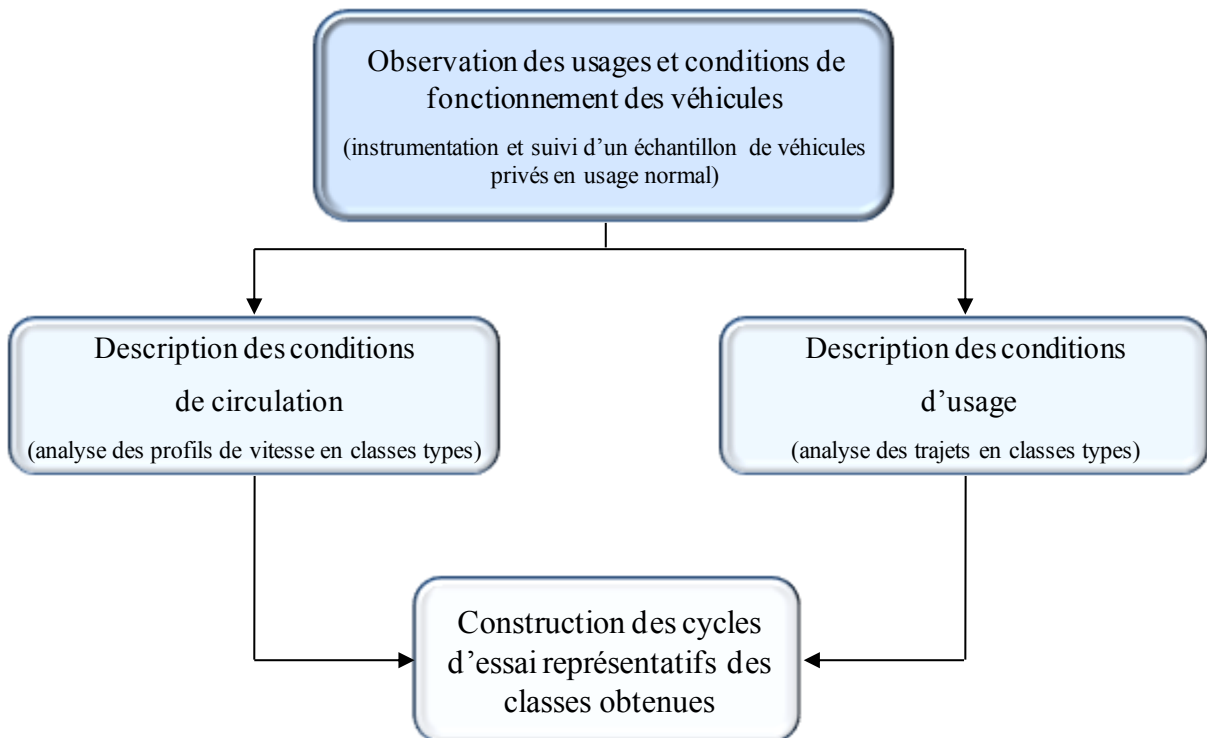
La complexité liée à l'intégration d'un ensemble de points de fonctionnement ou de cycles temporels liés aux variables d'environnement peut être réduite en utilisant des méthodes de classification. Nous illustrons ce type d'approche à l'aide de deux exemples distincts : l'intégration des missions de circulation pour des véhicules automobiles et la prise en compte des caractéristiques d'un gisement éolien relatif à un site géographique particulier.

(a) Analyse de missions de circulation de véhicules automobiles

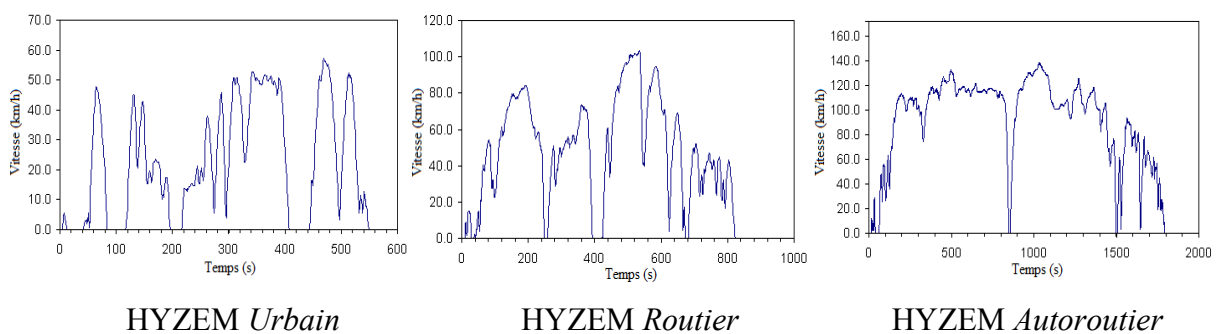
Bien que les cycles normalisés présentés précédemment permettent d'établir une comparaison de la consommation et de la pollution des véhicules dans une même sphère européenne, ces derniers sont loin de décrire les conditions réelles d'utilisation du véhicule. Pour remédier à cette insuffisance, l'INRETS (Institut national de recherche sur les transports et leur sécurité) devenu aujourd'hui l'IFSTTAR, en collaboration avec d'autres partenaires, s'est investi dans la détermination des cycles représentatifs des conditions d'usage et de fonctionnement des véhicules. L'idée est d'extraire des profils de vitesse représentatifs, à partir de l'analyse d'une base de données importante. Il s'agit plus précisément de classer les séquences de circulation réelle enregistrées d'une part selon les conditions de circulation (vitesse, accélération, température, situation du trafic,...) ce qui donne lieu à des classes types et d'une autre part selon les conditions d'usages (trajets, distance, conditions de démarrage, ...) ce qui permet de définir des classes de trajets (urbain, routier et autoroutier) [JOU 99], [AND 04]. Les cycles retenus sont typiquement déterminés selon un enchaînement ou un séquençement des profils représentatifs des classes obtenues (cf. Figure 1.14).

Dans ce contexte, l'INRETS a équipé plusieurs véhicules français par une instrumentation embarquée permettant de relever plusieurs paramètres de fonctionnement (vitesse, accélération, durée, distance, régime moteur, conditions de démarrage, température,...) lors d'un usage quotidien par leurs propriétaires. Dans le cadre du projet européen HYZEM (*European Development of Hybrid Vehicle Technology approaching efficient Zero Emission Mobility*), une campagne d'enregistrement a concerné 77 voitures privées européennes utilisées pendant un mois, en France, Allemagne, Grande-Bretagne et Grèce, sur 10 400 trajets et 90 000 km [AND 97]. L'analyse de ces profils a donné lieu à deux partitions imbriquées en 3 et 8 classes dont chacune est représentée par le profil le plus proche de son centre. La première classification concerne les types de trajets (urbain, routier et autoroutier) tandis que la deuxième permet une description plus fine de ces trois principaux trajets. Il s'agit par

exemple de qualifier un cycle ou un trajet principal (urbain, routier et autoroutier) selon les critères de stabilité de la vitesse, la durée d'arrêt, le niveau de vitesse et la distance du trajet [JOU 99]. Nous donnons à la Figure 1.15 les profils de vitesse des principaux cycles HYZEM.



**Figure 1.14.** Principe de la détermination des cycles représentatifs [JOU 99]



**Figure 1.15.** Cycles représentatifs HYZEM

La base de donnée du projet HYZEM a été ensuite réutilisée dans le cadre du projet européen ARTEMIS (*Assessment and Reliability of Transport Emission Models and Inventory Systems*) avec plusieurs autres enregistrements appartenant à différents laboratoires dont l'objectif est l'amélioration des méthodes d'estimation et d'inventaire des émissions de polluants dans le domaine du transport. Depuis 2005, ce projet a donné lieu à différents cycles

représentatifs des divers trajets et de l'ensemble des conditions de conduite [AND 04], [AND 05].

(b) Analyse de gisements éoliens

Ce deuxième exemple repose sur les mêmes principes que ceux évoqués dans la partie précédente. Il s'agit de regrouper dans une même classe, des séquences temporelles caractérisées par des distributions statistiques similaires. La similarité entre les distributions statistiques des séquences temporelles à classer peut être décrite par les premiers moments statistiques (moyenne, écart type) [CAL 04] ou plus finement par les caractéristiques d'ensemble de la distribution statistique en considérant tous les moments statistiques [CAL 08]. La répartition obtenue propose, pour un nombre de classes donné, la distribution type définissant chaque classe ainsi que la probabilité associée pour qu'une séquence appartienne à une classe. Le chronogramme temporel représentatif de chaque classe est ensuite obtenu à partir des caractéristiques statistiques de sa distribution type.

Cette approche a été appliquée pour la classification de la vitesse du vent dans le cadre des travaux de recherches menés au GRER (Groupe de Recherche en Energie Renouvelable d'Antilles Guyane en collaboration avec le MAPMO (Laboratoire de Mathématiques et Applications Mathématiques et Physiques d'Orléans).  $10^6$  séquences temporelles de 10 minutes de la vitesse du vent ont été mesurées sur le site de « Petit Canal » de la Guadeloupe. Trois classes de vent typiques du gisement ont pu être dégagées. La première classe, regroupant les événements météorologiques les plus probables (90%), est caractérisée par une distribution statistique unimodale symétrique. La deuxième classe, représentative de 9% seulement des événements, est définie par une distribution unimodale asymétrique. Enfin, à la troisième classe sont affectés les événements qui se produisent très rarement (1%) caractérisés par une distribution bimodale.

## **1.4 Nouvelle approche de classification et de synthèse d'une mission représentative dimensionnante**

### **1.4.1 Contexte**

L'état de l'art présenté dans la section précédente a bien mis en exergue l'importance de la notion de « mission dimensionnante » dans le contexte de la conception systémique. Toutefois, les méthodes exposées, même si elles semblent satisfaisantes dans certains cas de

figure, présentent encore quelques limites cruciales. Par exemple, dans le cas d'extraction de chronogramme temporel dimensionnant et de durée réduite (cf. [paragraphe 1.3.2](#)), le concepteur ignore le contenu informationnel de la mission initiale qui contribue pourtant énormément à la spécification des performances du système. De plus, il n'est pas toujours évident de repérer la « part » de mission la plus contraignante (dimensionnante), surtout lorsqu'on dispose d'une famille ou ensemble fourni de missions. Le plus souvent, telle part de mission est dimensionnante (puissance crête par exemple), tandis que telle autre l'est particulièrement vis-à-vis d'une autre contrainte (l'énergie ou la thermique par exemple) et ainsi de suite... Par exemple, en ce qui concerne la détermination des cycles de conduite dans le domaine de l'automobile, le fait de prendre la séquence de vitesse du véhicule la plus proche du centre de la classe identifiée, ne garantit pas la prise en compte des conditions les plus contraignantes (vitesse maximale, accélération maximale, cyclage,...).

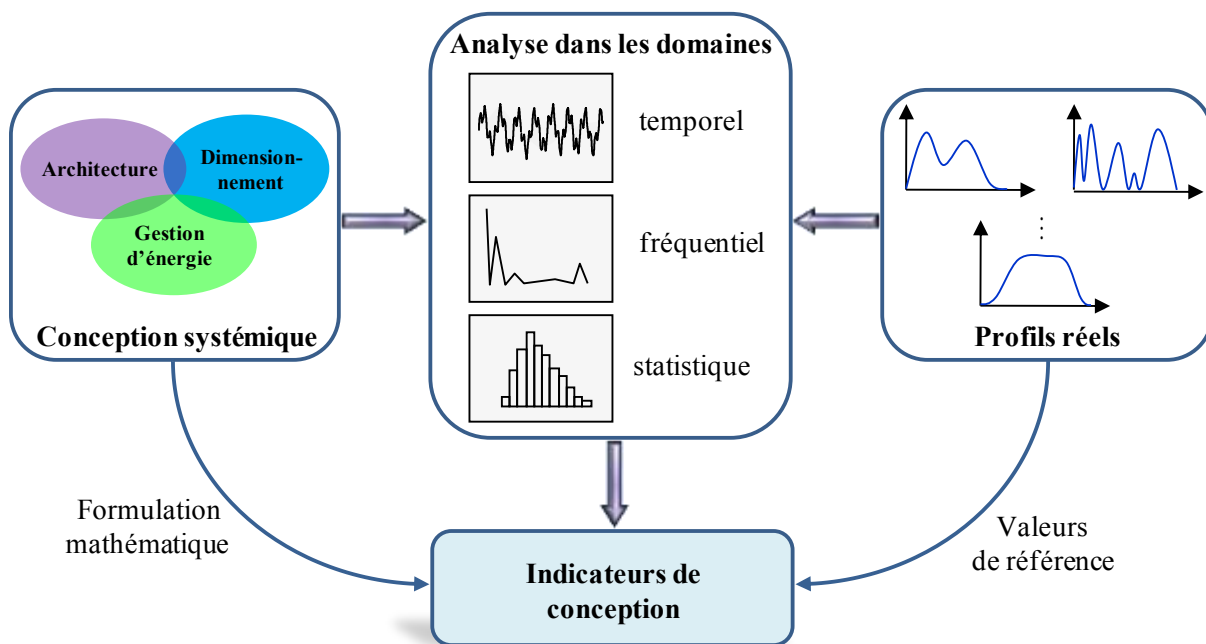
Afin de remédier à certaines limites, une solution consiste à simuler toutes les missions ou les variables environnementales du cahier des charges pour évaluer les performances, les critères et les contraintes de conception du système. Toutefois, cette solution, est rarement acceptable car elle accroît de manière forte les temps de calcul, ce qui s'avère souvent prohibitif dans un contexte de conception par optimisation pour laquelle le profil de mission est « rejoué » un grand nombre de fois au cours de la convergence du processus. L'analyse et la simplification (compactage) des missions et/ou des variables d'environnement apparaît donc comme une étape essentielle pour améliorer l'efficacité de la conception et permettre d'établir des solutions dimensionnées au plus juste tout en étant performantes vis-à-vis de critères hétérogènes (efficacité énergétique, durée de vie, coût...). Il est en fait plus judicieux de considérer dans le processus de conception une seule mission pertinente vis-à-vis des critères et contraintes de conception, tout en étant de faible durée et représentative d'un ensemble de missions (base de données de référence). Ce dernier point essentiel fait l'objet de notre travail, les méthodologies qui seront présentées par la suite constituant la contribution principale de cette thèse.

Pour plus de clarté, nous emploierons le terme « profil » pour désigner la mission (finalité) d'un système ou plus généralement une de ses variables d'environnement. Nous désignons par « profil réel » ( $P_{réel}$ ) la mission ou la variable environnementale réellement mesurée et disponible sous forme de « base de données » et par « profil fictif » ( $P_{fictif}$ ) la mission ou la variable environnementale traitée (profil synthétisé de durée réduite).



### 1.4.2 Indicateurs de caractérisation de la mission et des variables environnementales

L'approche de synthèse des profils représentatifs que nous proposons dans ce manuscrit, repose tout particulièrement sur la définition d'indicateurs de caractérisation pertinents vis-à-vis de la conception systémique. Il s'agit plus précisément de déterminer les caractéristiques essentielles du système à concevoir, caractéristiques se projetant dans les plans temporel (puissance maximale, moyenne, distorsion harmonique, énergie,...), fréquentiel (cyclabilité,...) et statistique (occurrences, densité de probabilité,...). Cette expertise du système nous permet alors d'obtenir une définition ou une formulation mathématique des indicateurs retenus. Les valeurs des *indicateurs de référence* sont issus directement du ou des profils réels (cf. Figure 1.16). Le contenu informationnel du profil fictif compact génère des valeurs de ces mêmes indicateurs de conception : le processus de synthèse converge dès lors que les indicateurs du profil généré correspondent aux indicateurs de référence.



**Figure 1.16.** Extraction des indicateurs de caractérisation de la mission et des variables d'environnement

Généralement, deux familles d'indicateurs sont identifiées : les indicateurs de dimensionnement et les indicateurs de performance. La première famille est relative aux contraintes de conception impactant directement l'architecture, le dimensionnement (taille des différents blocs énergétiques du système, contrainte de masse/volume, ...) et la gestion d'énergie du système. La deuxième famille est relative aux critères de performance (efficacité énergétique, cyclage et durée de vie notamment).

### **1.4.3 Mission et variables d'environnement au cœur du système : un couplage bidirectionnel**

La plupart du temps, l'influence des variables d'environnement (température et pression ambiante, ensoleillement, vitesse du vent,...) sur le système, de même que la satisfaction de la mission, sont représentées de façon unidirectionnelle, marquant les conditions aux limites du dispositif. En d'autres termes, les variables d'environnement ont une influence sur le système et conditionnent son comportement et son évolution, mais on considère que l'évolution du système causée par les variations des variables d'environnement n'a aucun impact sur l'environnement lui-même.

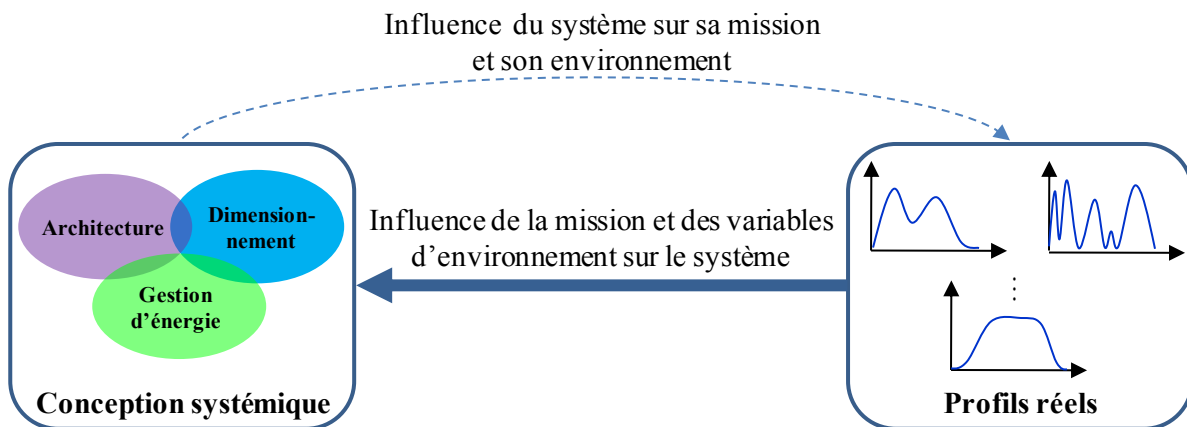
Il faut noter, à l'échelle planétaire, que cette vision unidirectionnelle de l'environnement est aujourd'hui de plus en plus remise en cause, notamment avec la notion d'empreinte écologique. Dans un tel cas, l'environnement (ici l'atmosphère terrestre) devient partie intégrante du système terre et interagit de façon bidirectionnelle.

De façon générale, si nous allons un peu plus loin dans cette réflexion, nous pouvons montrer qu'un couplage bidirectionnel entre système et environnement peut aussi exister au niveau des artefacts. Le système conçu, selon un ensemble de critères et de contraintes, peut d'une certaine manière influencer sa mission et ses variables d'environnement, ou tout au moins la « représentation que l'on se fait de ces variables environnementales ». (cf. [Figure 1.17](#)). Plus précisément, la façon dont le système est conçu où ses caractéristiques de fonctionnement peuvent influencer la mission elle-même ou la représentation de certaines variables d'environnement.

C'est par exemple le cas typique de la conception des véhicules électriques pour laquelle le mode de conduite est clairement différent de celui qu'aurait un conducteur de véhicule thermique classique : on peut alors parler « d'éco-conduite » dès lors que le conducteur de véhicule électrique tiendra compte de sa jauge énergétique (état de charge des batteries) ou modifiera ses habitudes de freinage pour favoriser la récupération électrique : dans ce cas typique, le cycle de « mission éco-conduite » est modifié par les caractéristiques du système conçu. L'exemple de la locomotive hybride PLATHEE que nous reprenons en fil rouge dans ce mémoire est aussi typique dans la mesure où la SNCF établirait des profils de vitesse tenant compte du caractère hybride de la propulsion, et de son dimensionnement : on pourrait même songer à aller jusqu'à optimiser le profil de mission pour profiter pleinement du

stockage pour moins consommer/polluer tout en assurant un minimum de confort et de performance (horaire).

*Nous sommes donc ici face à un couplage systémique environnement (mission) – système à concevoir, certes passionnant, mais plutôt complexe puisque « tout dépend de tout et son contraire »... Le concepteur doit donc faire face à ce nouveau couplage, la méthode de synthèse que nous proposons constituant à ce niveau une aide à la conception.*



**Figure 1.17.** *Couplage bidirectionnel entre le système et son environnement*

Plus particulièrement, selon notre approche de synthèse d'un profil fictif représentatif ( $P_{fictif}$ ), la façon avec laquelle nous proposons de traiter le profil réel ( $P_{réel}$ ) de référence dépend intimement de la conception du système, notamment à travers la définition des indicateurs. Il en résulte alors une forte influence des caractéristiques propres à la conception du système sur le profil fictif à générer qui sera considéré comme « profil dimensionnant ». Nous introduisons donc aussi, à ce niveau, un couplage bidirectionnel entre système et environnement (cf. Figure 1.18). Par exemple, dans le cas d'une chaîne éolienne, la valeur crête du profil fictif ( $P_{fictif}$ ) représentatif, doit impérativement correspondre à la valeur maximale de la vitesse réelle du vent, valeur constituant un indicateur typique du dimensionnement du système éolien.

Mais un problème de couplage supplémentaire se pose aussi. En effet, dans certaines classes de systèmes, il est parfois nécessaire de définir des *a priori* pour spécifier le jeu complet d'indicateurs de conception de référence. Ces *a priori* concernent des caractéristiques de la conception systémique, touchant à l'Architecture ou/et au Dimensionnement ou/et à la Gestion d'énergie. L'exemple du système hybride de la Figure 1.19, comprenant une source d'énergie principale (groupe Diesel) et un élément de stockage, est une bonne illustration. Ces

deux sources d'énergie sont connectées à travers un nœud énergétique pour satisfaire un ou plusieurs profils de charge.

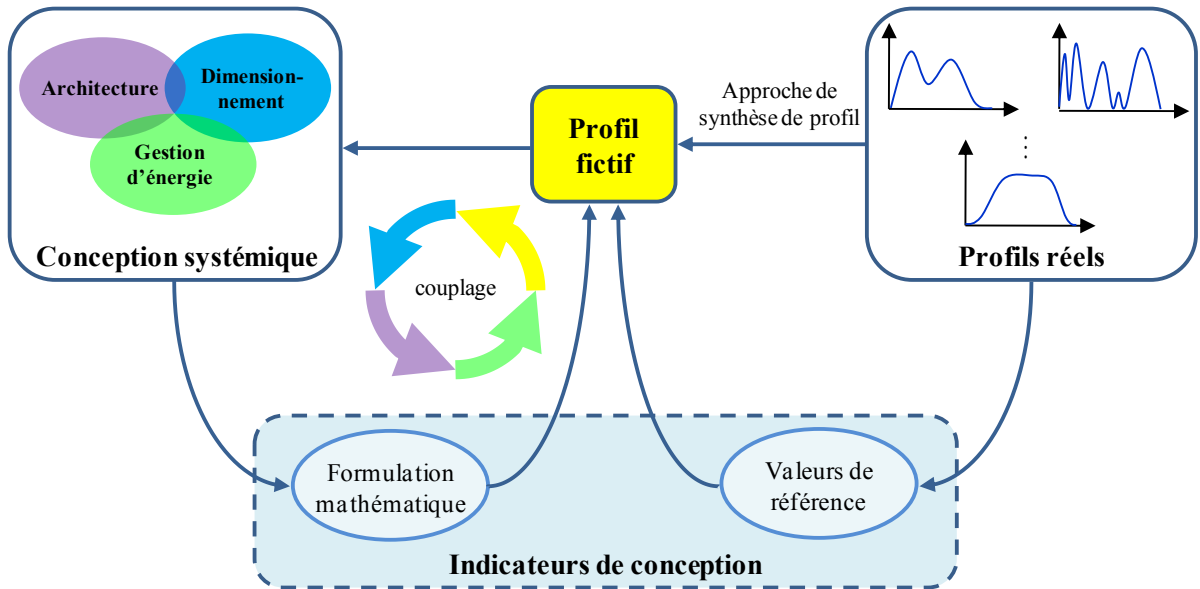


Figure 1.18. Couplage bidirectionnel entre le système et son profil fictif dimensionnant

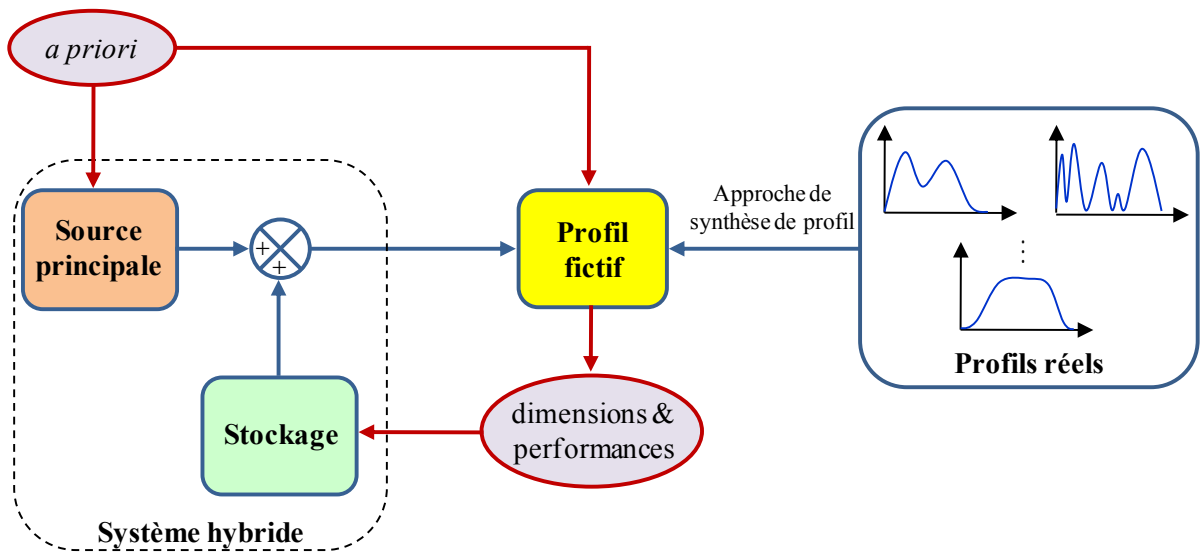


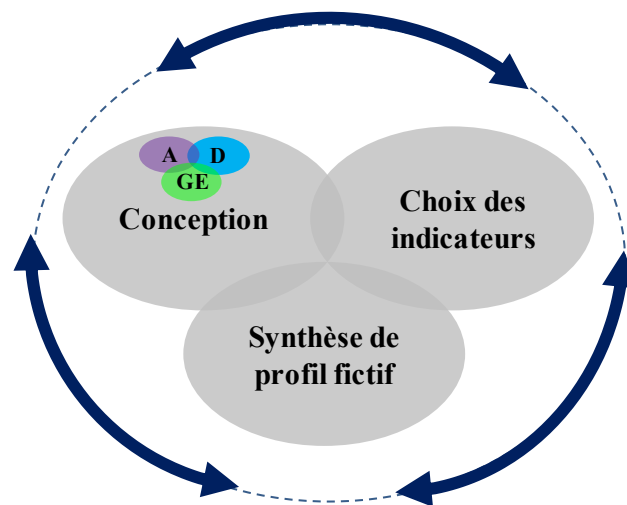
Figure 1.19. Architecture d'un système hybride avec stockage

Pour dimensionner correctement l'élément de stockage, en tenant compte des critères et contraintes pertinents au niveau système (consommation, durée de vie,...), nous sommes obligés d'avoir un *a priori* sur l'approche de dimensionnement de la source d'énergie principale, et ce avant même le processus de synthèse du profil fictif : comme nous le verrons au paragraphe suivant, on suppose « *a priori* » que la source principale assurera la valeur

moyenne de la puissance consommée. Cet *a priori*, touchant ici à la gestion énergétique et au dimensionnement, est nécessaire pour spécifier les indicateurs de référence, tout au moins les indicateurs à caractère énergétique.

Nous sommes donc, là encore, face à un couplage bidirectionnel supplémentaire entre « profil fictif » et « conception systémique », tel qu'illustré sur la [Figure 1.18](#).

L'analyse précédente liée au traitement de la mission et des variables d'environnement dans le contexte d'une approche systémique nous a permis de relever un ensemble de couplages entre approche de synthèse de profil, choix des indicateurs de caractérisation et caractéristiques de la conception systémique (cf. [Figure 1.20](#)). Ces couplages illustrent parfaitement bien les fondements de l'approche systémique, tandis que ses pionniers, tels que *Bertalanffy*, évoquent les notions d'« intégration de la finalité (mission) à la conception » ou encore de « modèles à finalité intégrée ». En bref, « la mission, ou tout au moins la représentation qu'on en fait se situe bien au cœur de la conception systémique ».



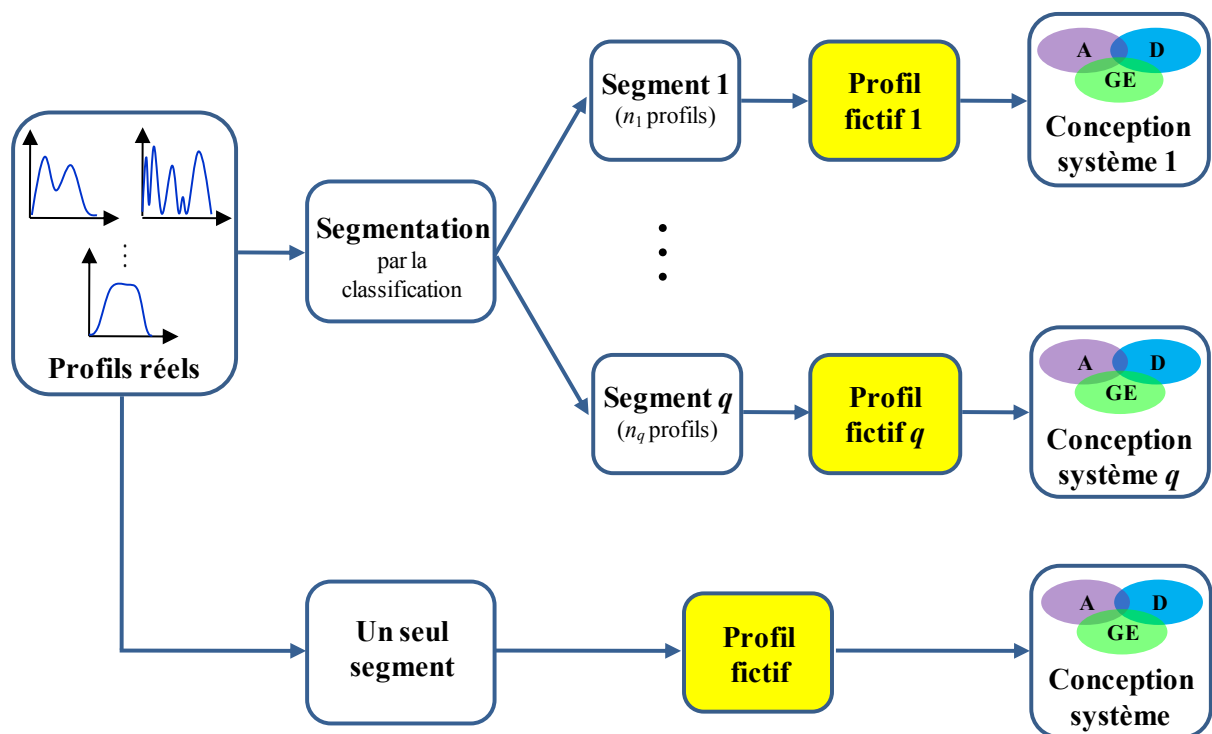
**Figure 1.20.** L'approche de synthèse de profil fictif au cœur de la conception systémique

#### 1.4.4 La classification : un outil d'aide à la conception

Souvent, un système est appelé à satisfaire plusieurs classes de besoins (missions) et doit pouvoir face à diverses conditions environnementales. Dans ce cas, nous parlons de « *segment* » pour définir une certaine classe de profils auxquels le système est confronté. A titre d'exemple, dans le contexte ferroviaire, nous distinguons les segments grande vitesse des segments triage, desserte locale, fret ou transport urbain. Par analogie, dans le secteur de l'automobile, les besoins énergétiques et les performances d'un véhicule varient selon ses

cycles de circulations autoroutiers, routiers, urbains ou mixtes. Certains véhicules sont plutôt dédiés à des segments spécifiques tandis que d'autres permettent de mixer plusieurs classes d'usage. Cette notion de classes s'applique également sur les variables d'environnement. En effet, on peut parler des classes de vitesse du vent : quatre classes de vent sont ainsi définies et normalisées pour les chaînes éoliennes et permettent de spécifier les garanties des produits conçus selon ces classes [WIN 06].

A l'instar de l'étude précédente sur la synthèse de profil de missions et de variables d'environnement, qui fait l'objet du troisième chapitre, les mêmes indicateurs pertinents vis-à-vis de la conception systémique peuvent être utilisés pour la « *segmentation* » des profils. Il s'agit ici de procéder à une classification des profils réels, disponibles dans la « banque de données », selon des indicateurs pertinents vis-à-vis des critères et contraintes de conception (cf. Figure 1.21) : cette approche est détaillée au chapitre suivant.



**Figure 1.21.** La classification est un outil d'aide à la décision

La classification constitue ainsi un outil d'aide à la segmentation de la gamme de produits à concevoir. L'exploitant pourra tirer parti de cette classification pour segmenter « en conscience » sa gamme de produits en fonction de critères d'optimisation technico-économiques et de sa stratégie commerciale.

## 1.5 1<sup>er</sup> exemple de conception systémique : la locomotive Hybride

Ce premier exemple nous permettra, dans la suite de ce mémoire, d'illustrer l'intérêt et la mise en œuvre de notre approche de classification puis de traitement de la mission, en allant jusqu'à l'utilisation de ces approches dans le contexte de la conception systémique par optimisation.

### 1.5.1 Contexte du projet PLATHEE

Suite aux changements climatiques dus à l'accroissement des gaz à effet de serre et à la raréfaction des ressources pétrolières pour les prochaines décennies, le transport ferroviaire français est appelé à trouver des solutions efficaces réduisant les impacts environnementaux de son réseau non électrifié et non électrifiable. Dans une stratégie d'asservissement de son bilan énergétique et écologique et afin d'acquérir le savoir faire en termes de spécification des locomotives hybrides, la SNCF accorde un intérêt particulier à l'hybridation ferroviaire [AKL 07], [THI 08]. C'est dans ce contexte que s'inscrit le projet PLATHEE (PLAte-forme pour Trains Hybrides Économiques en Énergie et Respectueux de l'Environnement) soutenu par l'ADEME et dirigé par la SNCF en collaboration avec plusieurs autres partenaires (LAPLACE, INRETS, Socofer, 2HEnergy, Soprano, Hélion, ERCTEEL et ALSTOM transport). Dans le cadre de ce projet, la SNCF et le groupe GENESYS du laboratoire LAPLACE (Groupe de recherche en Énergie Électrique et Systémique du Laboratoire Plasma et Conversion d'Énergie) s'intéressent à la conception, au dimensionnement et à la gestion d'énergie d'une locomotive hybride autonome et écologiquement propre. Cette locomotive hybride (cf. Figure 1.22.b) doit accomplir des missions de triage, de desserte locale et de secours en l'absence de caténaire, à savoir les mêmes missions que la locomotive Diesel BB63000 (cf. Figure 1.22.a) constituée d'un groupe Diesel de 610 kW et de quatre moteurs à courant continu de puissance nominale 100 kW chacun [AKL 08]. L'objectif du projet PLATHEE est de réduire la consommation du gasoil, et donc l'émission de CO<sub>2</sub>, de 60%. Nous donnons à la Figure 1.23 une comparaison des deux réservoirs de gasoil relatifs aux deux locomotives.

L'hybridation Diesel-électrique consiste alors à remplacer le groupe Diesel par un autre, de taille plus réduite, associé à des batteries de type Ni-Cd (Hoppecke, FNC 1502 HR, 135 Ah/1.2 V) et des supercondensateurs (EPCOS 5000 F/2.5 V) comme éléments de stockage d'énergie. La stratégie de gestion d'énergie fréquentielle qui a fait l'objet de

plusieurs travaux du LAPLACE/GENESYS [CHA 99], [LAN 06], est bien en adéquation avec cette double hybridation : le groupe Diesel fournit la puissance moyenne de la mission, les supercondensateurs assurant la partie haute fréquence et les batteries garantissant le reste de la mission [AKL 08].



(a) Locomotive BB63000



(b) Locomotive PLATHEE (2010)

**Figure 1.22.** Vue de la locomotive Diesel BB63000 et de la locomotive hybride PLATHEE



**Figure 1.23.** Réservoirs de gasoil des deux locomotives BB63000 et PLATHEE

Enfin, nous signalons que le projet PLATHEE a concerné aussi, d'une part, l'étude de l'intégration d'autres sources d'énergie telles que la pile à combustible [PLA 09] ou le volant d'inertie [JAA 09a] et d'une autre part l'hybridation d'autres segments de missions telles que les missions de la locomotive Diesel BB460000 sur la zone FRET de SOTTEVILLE et LE HAVRE [JAA 10d] (cf. Annexe A) et les missions TGV POS (Paris – Est de la France – Allemagne du Sud) (cf. Annexe B). Cette dernière étude (hybridation de la locomotive

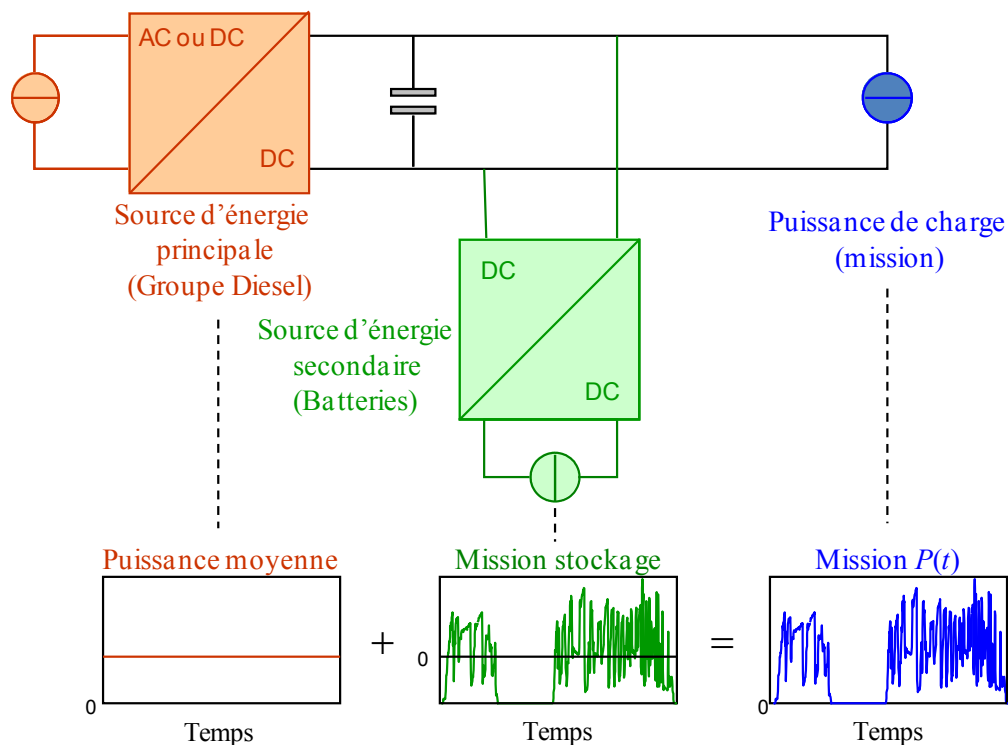


BB460000) a mis en exergue l'importance de la mission au niveau du dimensionnement du système, du choix de l'architecture d'hybridation et du choix de la technologie de stockage.

### 1.5.2 Traitement de la mission dans le cas d'une simple hybridation

- *Architecture d'hybridation et gestion énergétique*

Un premier exemple de traitement de la mission s'inscrit dans le contexte de la conception systémique optimale de cette locomotive hybride. Les missions traitées sont celles utilisées dans le projet PLATHEE et correspondent particulièrement aux missions de circulation réelle de la locomotive BB63000. S'ajoutent à cette base de données de missions celles inhérentes à la locomotive BB460000 et aux systèmes auxiliaires du TGV POS. Dans un souci de simplification, nous nous limitons ici au cas d'une hybridation simple qui consiste à coupler entre elles une source d'énergie principale (groupe Diesel) avec une source d'énergie secondaire (batteries d'accumulateurs). La stratégie de gestion d'énergie fait de sorte que le groupe Diesel fournisse la puissance moyenne de la mission, la puissance fluctuante étant assurée par les batteries (cf. Figure 1.24).



**Figure 1.24.** Architecture et gestion énergétique de la locomotive hybride selon une simple hybridation

- *Indicateurs de caractérisation de la mission pour la conception de la locomotive hybride*

Dans ce paragraphe, nous procédons à la détermination des indicateurs pertinents pour la caractérisation des missions ferroviaires : rappelons que cette pertinence est établie vis-à-vis des critères et contraintes de conception de la chaîne de traction hybride. Ces indicateurs serviront de base pour la synthèse d'un profil fictif représentatif d'une famille de missions. Deux familles d'indicateurs sont définies dans cet exemple : il s'agit de trois critères de dimensionnement permettant de spécifier les caractéristiques énergie/puissance des sources d'énergie et de deux critères de performance et de durée de vie.

Les indicateurs de dimensionnement sont :

– la *puissance maximale à fournir* :  $P_{\max}$

Afin de satisfaire la/les missions du cahier des charges, le système à concevoir doit être capable de fournir la puissance crête relevée dans l'ensemble des données de référence. La puissance maximale est alors un indicateur pertinent vis-à-vis de la conception du système.

$$P_{\max} = \max_{t \in [0, \Delta t_{\text{miss}}]} P(t) \quad (1.1)$$

avec  $P(t)$  désigne la puissance instantanée de la mission de référence de durée  $\Delta t_{\text{miss}}$ .

– *Puissance moyenne à fournir* :  $P_{\text{moy}}$

Selon la stratégie de gestion d'énergie exposée précédemment, la puissance moyenne de la mission est un indicateur primordial au dimensionnement de la source d'énergie principale. En d'autres termes, la puissance nominale du groupe Diesel ( $P_{\text{GDN}}$ ) doit correspondre à la valeur moyenne du signal de mission<sup>2</sup>.

$$P_{\text{moy}} = \frac{1}{\Delta t_{\text{miss}}} \int_0^{\Delta t_{\text{miss}}} P(\tau) d\tau \quad (1.2)$$

Les deux indicateurs de dimensionnement  $P_{\max}$  et  $P_{\text{moy}}$  ont été utilisés dans [AKL 08] pour définir le potentiel d'hybridation de puissance *PHP* :

<sup>2</sup> C'est à ce niveau qu'on voit bien, comme indiqué au [paragraphe 1.4.3](#), [Figure 1.19](#), qu'un *a priori* sur la conception, ici sur la stratégie de gestion d'énergie, est nécessaire afin de définir les indicateurs de caractérisation de la mission, donc avant même de synthétiser le profil fictif. Ce point constitue un problème « épineux » à résoudre lié à l'existence du couplage systémique entre mission synthétisée et dispositif à concevoir.

$$PHP = 1 - \frac{P_{\text{moy}}}{P_{\text{max}}} \quad (1.3)$$

Ce potentiel traduit la réduction potentielle de la source d'énergie principale que permettrait une hybridation par stockage. Plus cet indicateur est proche de 1, plus la mission est favorable à une démarche d'hybridation.

– *Energie utile au système de stockage* :  $E_u$

L'énergie utile au stockage représente la quantité d'énergie nécessaire pour satisfaire la mission du stockage en tenant compte des phases de charge et de décharge. Désignons par  $E_s(t)$  l'évolution instantanée de l'énergie dans le dispositif de stockage :

$$E_s(t) = - \int_0^t (P(\tau) - P_{\text{moy}}) d\tau \quad \text{avec } t \in [0, \Delta t_{\text{miss}}] \quad (1.4)$$

L'énergie utile est définie alors par :

$$E_u = \max_{t \in [0, \Delta t_{\text{miss}}]} E_s(t) - \min_{t \in [0, \Delta t_{\text{miss}}]} E_s(t) \quad (1.5)$$

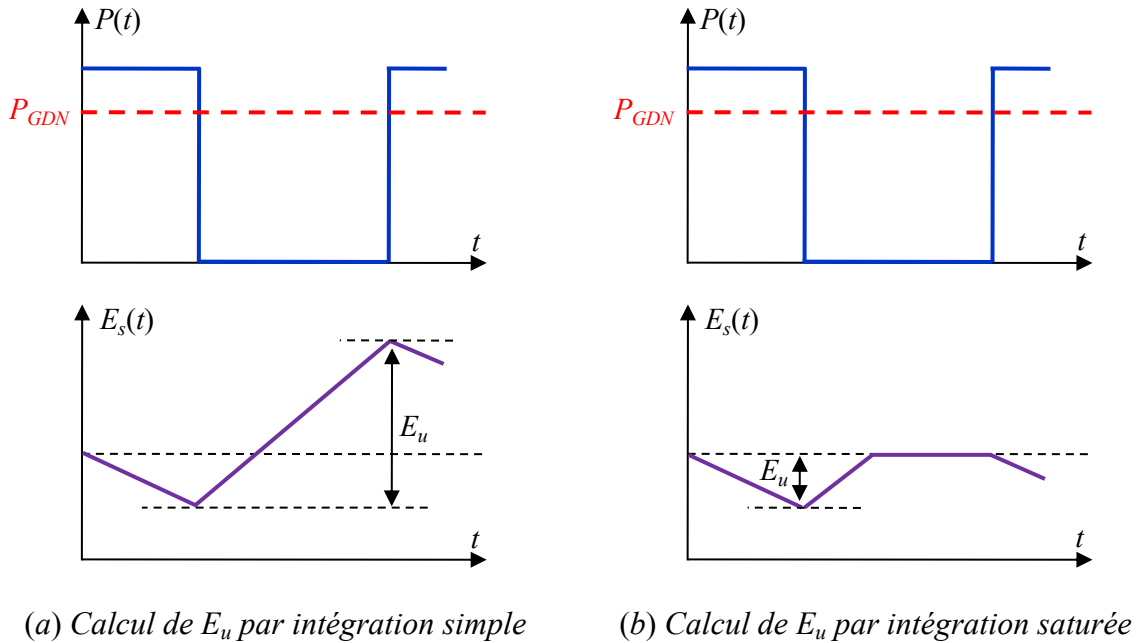
Deux remarques importantes à signaler lors du calcul de l'énergie instantanée  $E_s(t)$  :

- Le signe (-) devant l'intégrale est issu du choix de la convention générateur pour le stockage : une puissance positive correspond à une phase de décharge et donc une consommation d'énergie. A l'opposé, une puissance négative signifie une phase de charge ou de récupération de l'énergie.

- Dans le cas où nous dimensionnons le système pour plusieurs missions de puissances moyennes différentes, le choix de la puissance nominale du groupe Diesel n'est pas évident<sup>3</sup>. Si nous fixons, par exemple, la puissance nominale du groupe Diesel ( $P_{GDN}$ ) à la plus grande valeur des puissances moyennes de toutes les missions, tout en supposant que ce dernier fournit cette puissance d'une manière permanente, le bilan énergétique de l'élément du stockage n'est alors plus nul. Il y aura certainement, pour plusieurs missions, un excès de récupération de l'énergie ce qui conduira à un surdimensionnement énergétique du stockage. Pour remédier à ce phénomène, nous utilisons une intégration saturée qui inhibe toute phase de recharge inutile de l'élément de stockage et permet par la suite de le dimensionner « au plus juste ». Ceci s'explique par une modification de la

<sup>3</sup> Nous reviendrons sur cette problématique au chapitre 4

stratégie de gestion d'énergie en agissant sur l'état de marche/arrêt du groupe Diesel. Nous donnons à la [Figure 1.25](#) une illustration du calcul de l'énergie utile sans et avec intégration saturée. Cette dernière montre qu'une simple intégration tenant compte de toutes les phases de recharge conduit à un surdimensionnement de la taille énergétique du stockage.



**Figure 1.25.** Illustration du calcul de l'énergie utile  $E_u$

La détermination de l'énergie utile  $E_u$  nous permet de calculer la capacité de stockage totale des batteries  $E_{tot}$  en tenant compte de leur profondeur maximale de décharge autorisée. Par exemple, nous choisissons de solliciter les batteries Ni-Cd utilisées dans le projet PLATHEE jusqu'à 80% de leur capacité de stockage totale; il en résulte alors :

$$E_{tot} = \frac{E_u}{0.8} \quad (1.6)$$

Les indicateurs de performance et de durée de vie sont :

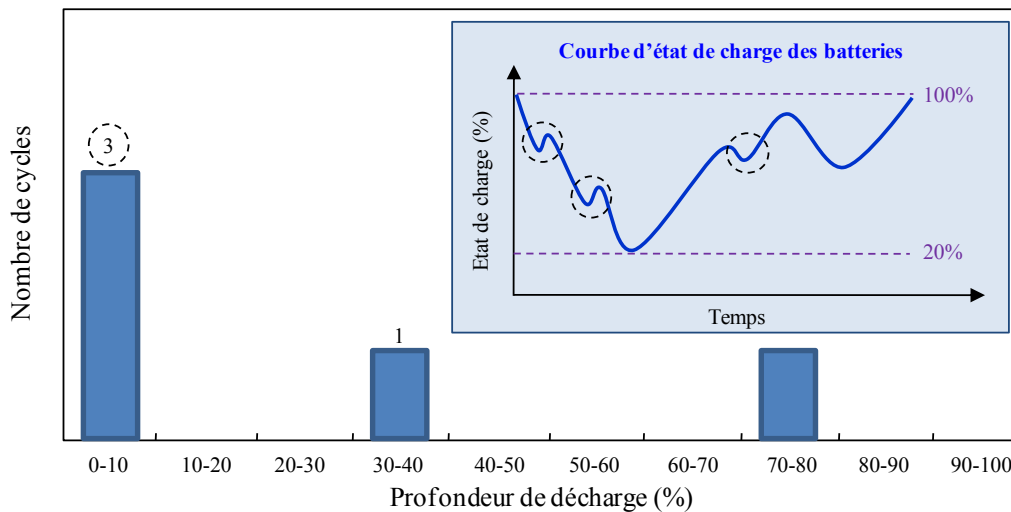
– La distribution statistique de la mission :  $I_{stat}$

L'indicateur statistique traduit l'efficacité énergétique moyenne du système à travers les occurrences des niveaux de puissance au cours de la mission. Il s'agit d'évaluer le rapport entre le temps passé au cours de la mission dans la fourchette de puissance  $[i\%P_{max}, (i+1)\%P_{max}]$  avec  $i = \{0, 1, \dots, 99\}$  (100 intervalles) et la durée de la mission.

Cet indicateur est exprimé alors par la distribution statistique (histogramme des fréquences) de la mission ou également par sa fonction de répartition (statistique des fréquences cumulées).

– *Le nombre de cycles imposés au système de stockage :  $Nb_{cyc}$*

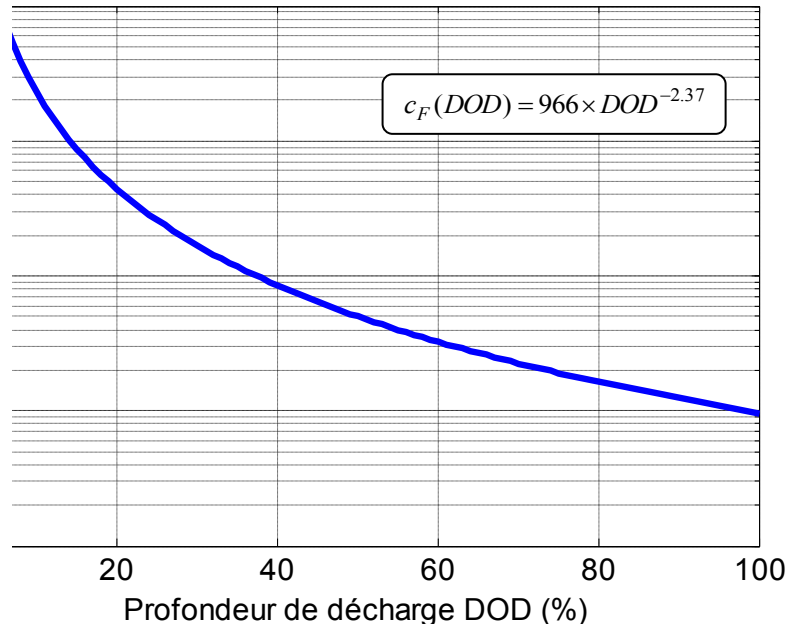
Nous considérons le nombre de cycles consommés par le pack de batteries comme un indicateur typique de la durée de vie. Le comptage du nombre de cycles est basé sur la méthode de « Rainflow » qui permet de déterminer, à partir de la courbe d'état de charge des batteries, le nombre de cycles par profondeur de décharge comme le montre la Figure 1.26 [DRO 97], [BAE 08]. Le principe de cette méthode de comptage de cycles est donné à l'Annexe C.



**Figure 1.26.** *Extraction des cycles en fonction de leur profondeur de décharge*

Le nombre de cycles équivalent ( $Nb_{cyc}$ ) est alors obtenu par la somme pondérée des cycles en fonction de leur profondeur de décharge. Les coefficients de pondération sont issus de la courbe constructeur donnant le nombre de cycles complets que l'élément de stockage est capable de fournir en fonction de sa profondeur de décharge (caractéristique dite de « cycle to failure »). Pour les batteries Ni-Cd Hoppecke FNC 1502 HR, une approximation du nombre de cycles disponibles ( $C_F$ ) en fonction de la profondeur de décharge ( $DOD$ ) pour une température entre 30°C et 40°C, une capacité de charge 1C et une capacité de décharge 2C est représentée par la Figure 1.27 et donnée par la relation suivante [AKL 09] :

$$c_F(DOD) = 966 \times DOD^{-2.37} \quad (1.7)$$



**Figure 1.27.** Nombre de cycles disponibles en fonction de la profondeur de décharge pour les batteries Ni-Cd

En considérant le nombre de cycles fournis par les batteries avec une profondeur de décharge de 100% comme référence, le poids d'un cycle de profondeur  $DOD$  est exprimé par la relation suivante :

$$w_{CYCLE}(DOD) = \frac{c_F(100\%)}{c_F(DOD)} \quad (1.8)$$

Le nombre de cycles équivalent est alors :

$$Nb_{cycles} = \sum_{DOD} w_{CYCLE}(DOD) \times N_{CYCLE}(DOD) \quad (1.9)$$

## 1.6 Traitement de la vitesse du vent pour la conception d'une chaîne éolienne

Ce deuxième exemple vise à illustrer le traitement des variables environnementales. Il s'inscrit dans le cadre de la conception optimale d'une chaîne éolienne. Il s'agit donc ici de traiter la vitesse du vent (vecteur d'énergie primaire) en vue de son intégration dans un processus de conception simultanée par optimisation, ceci afin de réduire le coût de traitement. Nous signalons au lecteur que nous nous limitons, dans cette application, au traitement de la vitesse du vent pour synthétiser un profil représentatif de durée réduite, respectant un ensemble d'indicateurs pertinents à la conception de la chaîne éolienne.

L'intégration de ce profil dans le processus de conception ne fait pas partie de notre étude mais peut s'inscrire dans la suite de ce travail. Nous disposons dans cet exemple de deux mois de mesures successives de la vitesse du vent sur deux sites géographiques différents : le site de « Sidi Daoued » en Tunisie et le site de « Petit Canal » en Guadeloupe (France).

Une phase d'analyse de la vitesse du vent dans les plans temporel et statistique nous a permis de dégager deux familles d'indicateurs établis vis-à-vis des critères de conception typiques du domaine éolien :

Les indicateurs de dimensionnement sont :

– la *Vitesse maximale* :  $V_{\max}$

Cet indicateur traduit la prise en compte des rafales de vent lors du dimensionnement de la chaîne éolienne.

$$V_{\max} = \max_{t \in [0, \Delta t_{profil}]} V(t) \quad (1.10)$$

avec  $V(t)$  désigne la vitesse instantanée du vent sur un horizon temporel  $\Delta t_{profil}$ .

– la *Vitesse cubique moyenne* :  $\langle V^3 \rangle$

Il s'agit d'un indicateur très pertinent énergétiquement car il traduit la puissance utile moyenne captée par l'éolienne. En effet, la puissance utile instantanée de la turbine de l'éolienne est proportionnelle à la vitesse cubique du vent comme nous le montrons plus bas au niveau de l'équation (1.13).

Dans un système de conversion aérodynamique, l'énergie  $dE$  d'une colonne d'air de masse volumique  $\rho$ , de longueur  $dl$ , de section  $S$  égale à la surface balayée par la turbine et évoluant à une vitesse instantanée  $V$  s'écrit [MAR 08] (cf. Figure 1.28) :

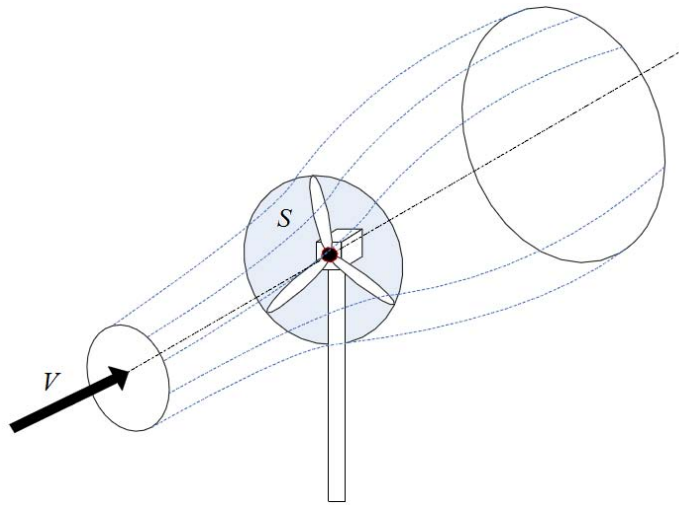
$$dE = \frac{1}{2} \rho S dl V^2 \quad (1.11)$$

Sous l'hypothèse que  $dl = V dt$ , la puissance  $P$  de cette colonne d'air s'écrit :

$$P = \frac{dE}{dt} = \frac{1}{2} \rho S V^3 \quad (1.12)$$

La puissance captée par la turbine  $P_{éol}$  n'est qu'une partie de la puissance  $P$  de la masse d'air. Elle est déterminée en tenant compte d'un coefficient de puissance  $C_p$  dont la valeur dépend du type de la turbine. Pour plus de détails sur ce coefficient de puissance et sur la puissance réellement récupérée par la voilure éolienne, nous invitons le lecteur à consulter [HAU 00] et [JOH 01].

$$P_{éol} = C_p \times P = \frac{1}{2} \rho C_p S V^3 \quad (1.13)$$



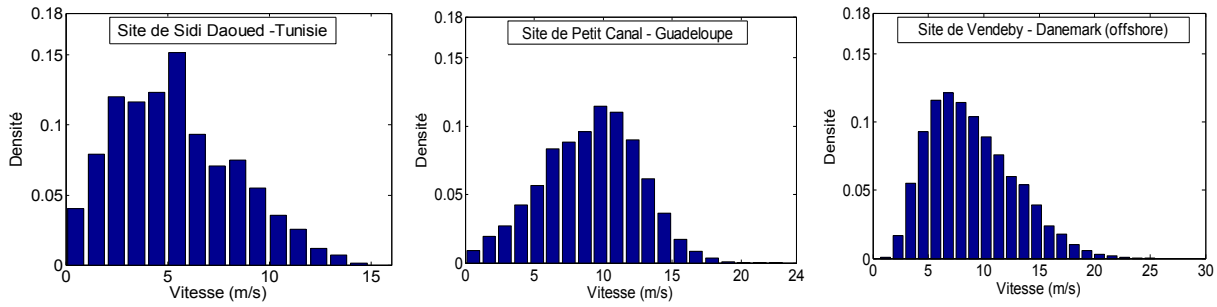
**Figure 1.28.** Conversion aérodynamique

Les indicateurs de performance sont :

– La distribution statistique de la vitesse du vent : *pdf* (probability density function)

Il est important de tenir compte de la distribution statistique de la vitesse du vent (*pdf*) dans la conception de la chaîne éolienne. Le profil fictif que nous proposons de générer doit alors respecter la même distribution statistique de la vitesse réelle du vent mesurée sur un horizon très large. Parmi les lois de probabilité qui peuvent caractériser la distribution statistique de la vitesse du vent, on note la distribution log-normal, la distribution gaussienne bivariante et la distribution de Weibull. Cette dernière est classiquement considérée comme la plus adéquate avec les propriétés statistiques du vent réel [KEL 92]. La Figure 1.29 montre quelques distributions statistiques de la vitesse du vent pour trois sites géographiques différents : le site de « Sidi Daoued » en Tunisie, le site de « Petit Canal » en Guadeloupe (France) et le site de « Vendeby » au Danemark (offshore). Nous vérifions bien que la distribution de la vitesse du vent suit une loi de Weibull pour les trois sites ainsi présentés.

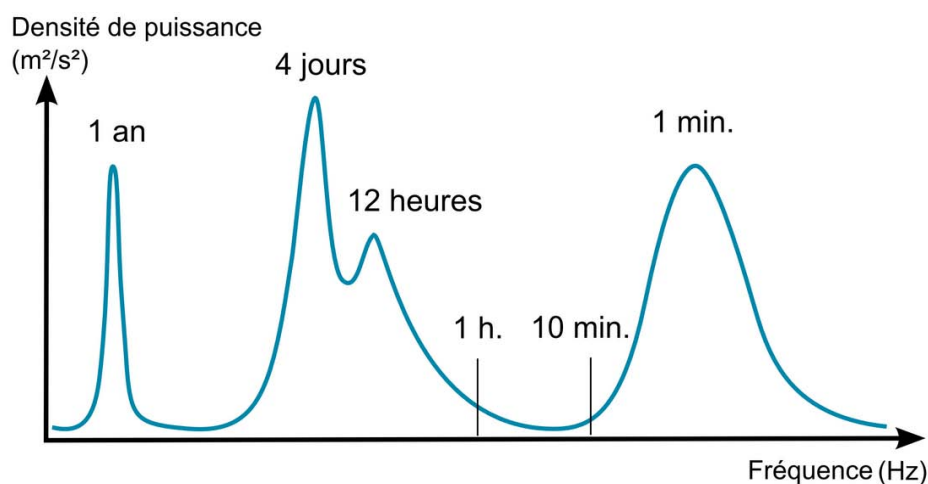




**Figure 1.29.** Distribution statistique de la vitesse du vent pour trois sites géographiques différents

– La turbulence du vent :  $I_{turb}$

La turbulence du vent est un phénomène crucial dans l'estimation de la durée de vie de l'éolienne notamment en évaluant sa fatigue mécanique<sup>4</sup>. Il est alors important, d'intégrer un indicateur de caractérisation de la turbulence dans le signal temporel représentatif. La Figure 1.30 donne le graphique de « Van der Hoven » qui expose une analyse spectrale de la densité de puissance de la vitesse du vent dans la couche limite turbulente (couche caractérisée par une altitude maximale de 1 km) [JOU 10], [JOA 07]. Ce graphique montre l'existence de plusieurs échelles de fluctuation du vent. La turbulence est caractérisée par une période de l'ordre de la minute. Elle est souvent liée aux conditions météorologiques, au gradient thermique vertical, à la distribution des nuages et à la rugosité du sol (profil local, végétation, bâtiments). Le « trou » entre une heure et dix minutes signifie l'existence de périodes de « calme » à très faible énergie de fluctuation [HIR 89], [JOU 10].



**Figure 1.30.** Spectre de densité de puissance de la vitesse horizontale du vent à 100 mètres « graphique de Van der Hoven »

<sup>4</sup> Cet indicateur permet, au même titre que les vitesses maximale et moyenne de spécifier les 4 classes typiques du domaine éolien.

En se référant au graphique de « *Van der Hoven* », nous pouvons considérer que le choix d'une période d'échantillonnage  $T_e = 10$  minutes pour la modélisation du vent hors turbulence garantit un bon découplage entre la dynamique lente et la dynamique rapide, caractéristique du phénomène de turbulence. Souvent, la dynamique rapide de la vitesse du vent est modélisée, sur chaque période d'échantillonnage  $T_e$ , par une distribution gaussienne de moyenne nulle et d'écart type  $\sigma$  [STR 90] alors que la dynamique lente suit, dans la plupart des cas, une loi de Weibull [KEL 92]. Nous caractérisons alors la turbulence du vent, sur chaque intervalle  $i$  de durée  $T_e = 10$  minutes, par l'écart type  $\sigma_i$  de la distribution gaussienne correspondante. Nous définissons par la suite l'indicateur de turbulence  $I_{turb}$  sur tout le profil du vent par la valeur moyenne des écarts types  $\sigma_i$ .

$$I_{turb} = \langle \sigma_i \rangle \quad \text{avec} \quad i = [1, \text{partie entière} (\Delta t_{profil} / T_e)] \quad (1.14)$$

## 1.7 Conclusion

Au niveau de ce chapitre introductif, nous avons tout d'abord présenté les notions de base relatives au système et au contexte spécifique à l'approche de conception systémique. Nous avons montré que la conception des systèmes complexes, particulièrement les systèmes multisources, est plus appropriée selon une approche de conception simultanée qu'avec une approche de conception séquentielle, même si cette dernière facilite la tâche des concepteurs par une rupture des couplages entre architecture, dimensionnement et gestion du système. Le caractère simultané de cette approche de conception permet de garantir l'optimalité globale du système. Afin de tenir compte des couplages de conception, cette approche simultanée est le plus souvent résolue à l'aide d'un outil d'optimisation, couplé à un modèle adéquat (systémique) du dispositif : cette approche permet alors de déterminer, d'une manière simultanée, les paramètres relatifs à l'architecture, au dimensionnement et à la gestion d'énergie. A travers un état de l'art, nous avons mis en exergue la problématique d'intégration de la mission et des variables environnementales dans une démarche de conception systémique ainsi que certaines démarches plus ou moins classiques. Cette problématique est d'autant plus critique lorsqu'il s'agit d'intégrer un ensemble de missions ou de variables d'environnement dans un processus de conception simultanée par optimisation, processus dans lequel le temps de calcul est primordial pour le traitement de systèmes complexes.

Ensuite, nous avons présenté les concepts de base d'une nouvelle approche de synthèse d'une mission ou d'une variable environnementale, dimensionnante et compacte, approche

que nous proposons d'étudier dans ce rapport. Cette démarche est initialement basée sur la détermination d'un jeu d'indicateurs pertinents vis-à-vis des critères et contraintes de conception. Nous avons également montré l'intérêt de ces indicateurs pour une démarche de classification des missions du cahier des charges dans une phase amont de la conception permettant ainsi de guider le choix du concepteur sur le nombre de dispositifs (segments) à concevoir. En fin de ce chapitre, nous avons spécifié les indicateurs de conception relatifs à deux exemples d'applications qui nous permettront dans les chapitres suivants d'illustrer l'intérêt de notre nouvelle approche de traitement de la mission et des variables d'environnement. Le premier exemple s'inscrit dans le contexte du projet PLATHEE pour lequel les missions ferroviaires font partie intégrante du processus de conception simultanée par optimisation d'une locomotive hybride. Le deuxième exemple concerne le traitement de la vitesse du vent pour la conception d'une chaîne éolienne. Ces deux exemples, volontairement très différents en termes de nature et de domaine montrent bien que les démarches de classification et de synthèse proposées dans ce mémoire ont un caractère très générique, pouvant en particulier dépasser le domaine de la conception en génie électrique.

## Chapitre 2

# Classification des missions et des variables environnementales

### 2.1 Introduction

La classification d'objets ou d'organismes et êtres vivants est une problématique fréquemment rencontrée dans de nombreuses applications : reconnaissance de forme, analyse de données spatiales, segmentation d'images, fouille de données, analyse de marché, bioinformatique (séquençement d'ADN ou de protéines). Du point de vue de la systémique, la classification de profils de mission ou, plus généralement, de variables environnementales selon des indicateurs pertinents vis-à-vis des critères et contraintes de conception, fait partie d'un ensemble d'outils d'aide à la décision et constitue une phase amont essentielle d'un processus de conception. En effet, elle doit permettre au concepteur d'évaluer l'intérêt d'un dispositif spécifiquement optimisé pour une classe donnée, relative à des profils de mission et/ou de variable environnementale, ceci par rapport à un dispositif capable de satisfaire simultanément un ensemble de classes. Dit autrement, sur l'exemple bien connu de l'automobile, « que gagne-t-on à *classer, segmenter* une gamme de véhicule (urbain, routier, tout terrain, etc) d'un point de vue du ratio coût/performance et ce en regard de l'usage et de ses occurrences » ?

La classification ou la segmentation peut ainsi permettre d'améliorer notablement certains critères tels que l'efficacité énergétique du système. C'est par exemple le cas lorsque, à l'issue d'une classification selon des indicateurs à caractère énergétique, les segments ou les classes obtenus sont différents de la segmentation commerciale ou d'usage du système.

La classification consiste à regrouper des éléments ayant des caractéristiques ou des *indicateurs similaires*. Cette similarité est souvent traduite par un critère mathématique

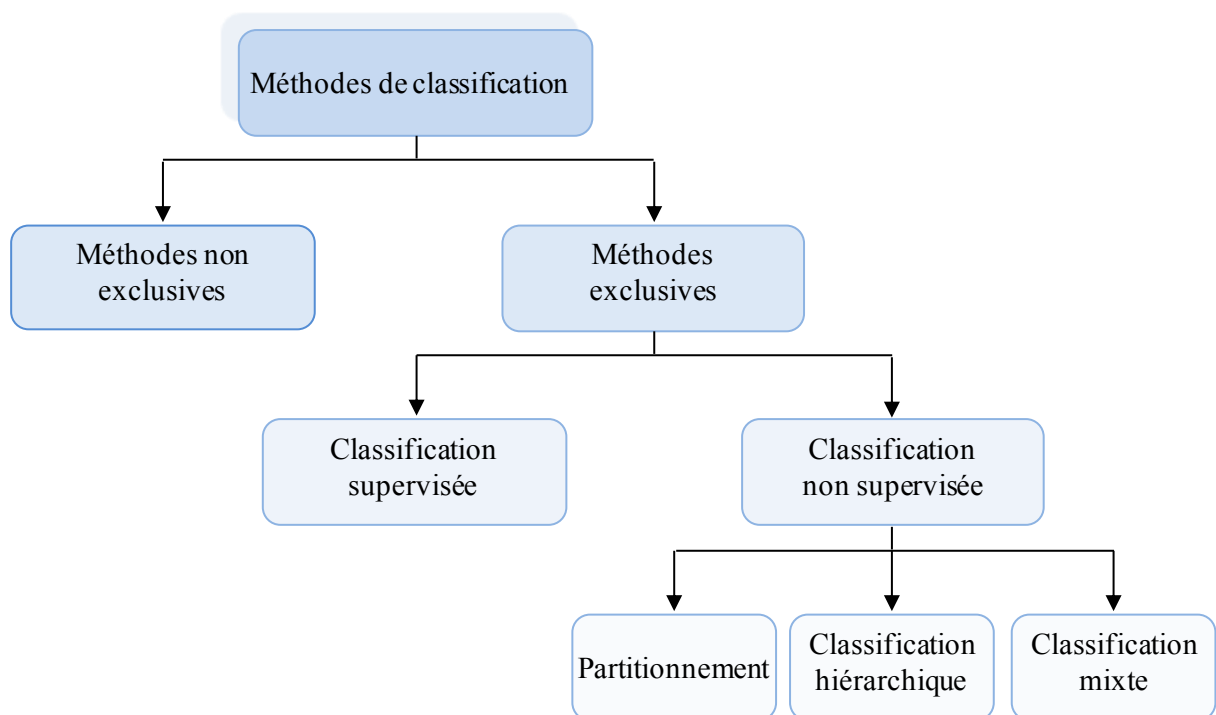
(distance) qualifié de critère d'agrégation. Les techniques de classification sont engendrées par une démarche algorithmique et non par des calculs formalisés usuels. Elles ne nécessitent alors que des notions mathématiques élémentaires. Elles sont nombreuses et variées et le choix d'une méthode dépend à la fois du problème traité, de sa dimension et de la nature des données. Les méthodes de classification les plus répandues sont les méthodes « *exclusives non supervisées* », telles que la technique d'*agrégation autour des centres mobiles*, la *classification ascendante hiérarchique* et la *classification mixte* [JAI 88], [LEB 00]. Ces méthodes nécessitent de connaître ou d'imposer le nombre de classes de la série de données à analyser pour pouvoir classer l'ensemble des éléments. Or, fixer *a priori* le nombre de classes optimal avant d'avoir classé les éléments n'est pas forcément évident. Pour solutionner ce problème, il est possible d'utiliser des critères de partitionnement qui expriment un compromis entre la minimisation des distances intra-cluster et la maximisation des distances inter-clusters. L'optimisation de ces critères permet de déterminer simultanément le nombre optimal de classes d'un problème et le partitionnement des données. Cette approche permet ainsi de formuler le problème de classification en un problème d'optimisation multimodal que l'on peut résoudre avec des méthodes appropriées telles que les *algorithmes génétiques de nichage* (*niching genetic algorithms*) [SHE 05].

Dans ce chapitre, nous dressons brièvement l'état de l'art des techniques de classification. Nous portons notre attention, dans un premier temps, sur les méthodes les plus populaires : la technique d'*agrégation autour des centres mobiles*, la *classification ascendante hiérarchique* et la *classification mixte*. Dans un deuxième temps, nous présentons le principe de la classification sans *a priori* sur le nombre de classes, basé sur l'optimisation de critères de partitionnement. En particulier, nous présentons un algorithme génétique dérivé de la méthode RTS (*Restricted Tournament Selection*) [HAR 95] capable de résoudre le problème de classification en optimisant l'indice *silhouette* associé à la série de données. L'algorithme développé est tout d'abord appliqué à 3 benchmarks « tests » puis à l'analyse de missions et variables environnementales de systèmes énergétiques typiques. Nous nous intéressons dans un premier temps à la classification de missions ferroviaires utilisées pour la conception de systèmes d'alimentation pour locomotive hybride avec stockage. Puis, nous analysons à l'aide de l'outil de classification, différents profils de vitesse de vent issus de deux sites géographiques différents, ces profils étant utilisables pour des applications éoliennes.

## 2.2 Techniques de classification : état de l'art

Les techniques de classification sont définies comme des processus de rassemblement d'un certain nombre d'objets en sous-groupes selon un critère de dissimilarité. Cette dissimilarité est souvent traduite par une distance afin de quantifier la relation entre les objets ou les éléments à classer. En effet, lorsque ces éléments sont représentés par des points dans un espace à  $q$  dimensions, le critère mathématique de dissimilarité entre deux éléments est défini par la distance entre leurs points représentatifs. A titre d'exemple, dans une population de 100 individus caractérisé chacun par son poids et sa taille (espace à 2 dimensions), les deux individus les plus semblables sont ceux dont la distance entre *points représentatifs* est la plus faible. A ce niveau, il est plus pratique de définir une *matrice de dissimilarité*  $\mathbf{D}$  qui regroupe toutes les distances inter-éléments. Les lignes ainsi que les colonnes de cette matrice correspondent donc aux éléments à classer.

Nous présentons, à la [Figure 2.1](#) les différentes variantes des méthodes de classification.



**Figure 2.1.** Les méthodes de classification [JAI 88]

Les méthodes de classification « *non exclusives* » permettent qu'un élément ou un objet appartienne, en même temps, à plusieurs classes [SHE 05]. Cette affectation multi-classes conduit souvent à un recouvrement entre les classes obtenues. A contrario, les méthodes

« *exclusives* » n'autorisent qu'une unique affectation d'un élément dans une classe. Ces dernières sont les plus efficaces en termes d'aide à la décision et elles sont scindées en deux catégories : la classification « *supervisée* » et la classification « *non supervisée* ».

– La famille des méthodes de classification « *non supervisée* » est basée uniquement sur la matrice de dissimilarité **D**. Les éléments sont alors dits « *non étiquetés* » car l'algorithme de classification ne tient compte d'aucune information sur la classe d'un élément.

– Contrairement au cas précédent, les méthodes de classification « *supervisée* » (dites aussi méthodes inductives) utilisent non seulement la matrice de dissimilarité **D** mais aussi d'autres informations utiles sur les classes. Elles ont pour objectif d'apprendre par « l'exemple ». Il s'agit plus précisément de classer de nouveaux éléments « *étiquetés* » dont la classe n'est pas connue. Dans un tel cas, les classes sont déjà définies soit au préalable soit par des éléments antérieurs bien classés. Par exemple, pour la reconnaissance de l'écriture manuscrite, l'idée de base est d'affecter chaque lettre d'un mot, selon sa forme, dans l'une des 26 classes représentant les lettres de l'alphabet.

Dans ce chapitre, nous accordons une attention particulière aux méthodes de classification « *exclusive non supervisée* ». Nous différencions la méthode de partitionnement connue sous le nom *d'agrégation autour des centres mobiles* et qui conduit à une seule partition de l'ensemble des données, de la méthode de *classification hiérarchique* donnant lieu à une séquence de partitions imbriquées. Nous parlons également d'une troisième technique de *classification mixte* qui recombine les deux techniques précédentes.

### **2.2.1 Agrégation autour des centres mobiles**

La technique *d'agrégation autour des centres mobiles*, appelée aussi méthode de *K-means*, est la technique de partitionnement utilisant le formalisme le plus simple. Inventée par *Forgy* en 1965 [FOR 65] puis développée par *MacQueen* en 1967 [MAC 67], cette méthode a prouvé son efficacité pour les applications de partitionnement à grand nombre de données. Le critère de dissimilarité souvent utilisé est la distance euclidienne entre deux éléments de l'ensemble des données [LEB 00]. Avant de décrire les différentes étapes de l'algorithme *d'agrégation autour des centres mobiles*, nous présentons au [Tableau 2.1](#) les distances les

plus usuelles dans les problèmes de classification. Nous désignons par  $E_1$  et  $E_2$  deux éléments à classer définis dans un espace de dimension  $q$  ( $E_1 = \{x_1(E_1), \dots, x_q(E_1)\}$ ,  $E_2 = \{x_1(E_2), \dots, x_q(E_2)\}$ ).

Désignons par  $\mathbf{E} = \{E_1, E_2, \dots, E_n\}$  l'ensemble de  $n$  éléments ou objets à classer. Chaque élément est caractérisé par  $q$  indicateurs. Nous supposons que l'espace  $\mathbb{R}^q$  est muni d'une distance  $d$ . Notons par  $N$  le nombre de classes désiré et par  $g_k$  le centre (aussi qualifié de barycentre ou cluster) de la classe  $C_k$  ( $k = \{1, \dots, N\}$ ). Les étapes successives de l'algorithme, données ci-dessous, sont également illustrées à travers un exemple de classification représenté par la [Figure 2.2](#). Il s'agit de regrouper 15 éléments, définis dans un espace de dimension  $q = 2$ , en deux classes selon les indicateurs de classification  $x_1$  et  $x_2$ .

Distance	Expression
Minkowski	$d(E_1, E_2) = \sqrt[n]{\sum_{i=1}^q  x_i(E_1) - x_i(E_2) ^n}$
Euclidienne	$d(E_1, E_2) = \sqrt{\sum_{i=1}^q (x_i(E_1) - x_i(E_2))^2}$
Manhattan	$d(E_1, E_2) = \sum_{i=1}^q  x_i(E_1) - x_i(E_2) $
Tchebychev	$d(E_1, E_2) = \max_{i=\{1, \dots, q\}}  x_i(E_1) - x_i(E_2) $

**Tableau 2.1.** Exemples de mesures de distance utilisées dans la classification

**Étape 0 :** L'étape initiale consiste à effectuer un tirage pseudo-aléatoire de  $N$  centres de classes provisoires parmi les  $n$  éléments de l'ensemble  $\mathbf{E}$ . Les  $N$  centres ainsi obtenus :

$$\{g_1^0, g_2^0, \dots, g_N^0\}$$

donnent lieu à une répartition initiale des  $n$  éléments en  $N$  classes :

$$\{C_1^0, C_2^0, \dots, C_N^0\}$$

de sorte que tous les éléments de la classe  $C_k^0$  sont les plus proches du centre  $g_k^0$ .

**Étape 1 :** Cette étape permet de déterminer  $N$  nouveaux centres définis comme étant les



centres de gravité des classes obtenus suite à l'étape 0. Ces nouveaux centres de classes :

$$\{g_1^1, g_2^1, \dots, g_N^1\}$$

donnent naissance à une nouvelle partition en affectant chaque élément de la population au barycentre le plus proche. Nous obtenons alors les  $N$  classes suivantes :

$$\{C_1^1, C_2^1, \dots, C_N^1\}$$

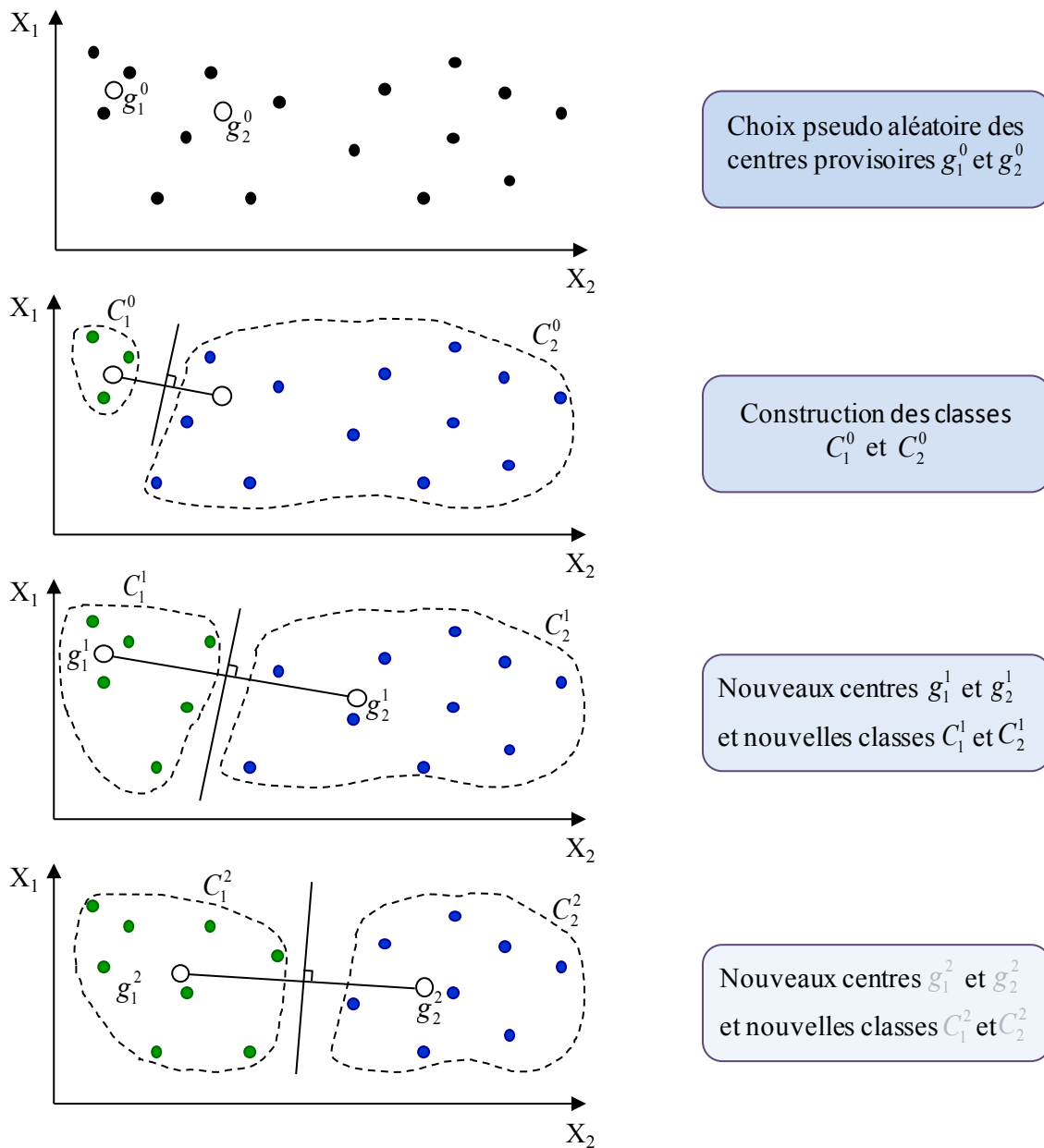


Figure 2.2. Exemple d'illustration de l'algorithme d'agrégation autour des centres mobiles

**Étape  $m$**  : L'étape  $m$  est basée sur le même principe que l'étape 1. Il s'agit de déterminer  $N$  nouveaux centres de classes :

$$\{g_1^m, g_2^m, \dots, g_N^m\}$$

en prenant les centres de gravité des classes obtenues à l'étape  $m - 1$  :

$$\{C_1^{m-1}, C_2^{m-1}, \dots, C_N^{m-1}\}$$

La partition d'ordre  $m$  fait alors apparaître les classes :

$$\{C_1^m, C_2^m, \dots, C_N^m\}$$

Ce processus de partitionnement sera alors réitéré jusqu'à ce que deux itérations successives donnent lieu au même classement ou lorsqu'un nombre maximal d'itérations est atteint. L'algorithme d'*agrégation autour des centres mobiles* relève d'un problème d'optimisation qui tend à minimiser l'erreur quadratique moyenne (*Mean Squared Error* ou *MSE*) définie par la somme des carrés des distances euclidiennes entre chaque élément et le barycentre de sa classe d'affectation. Cette dernière est exprimée par l'équation (2.1).

$$MSE = \frac{1}{n} \sum_{k=1}^N \sum_{E_i \in C_k} \|E_i - g_k\|^2 \quad (2.1)$$

La convergence de l'algorithme est alors certaine car la variance intra-classe (l'erreur quadratique moyenne *MSE*) ne peut que décroître ou rester stable entre deux étapes successives. Pour plus de détail sur la convergence de l'algorithme, nous renvoyons le lecteur à [LEB 00].

Avec pratiquement le même principe, la technique d'*agrégation autour des centres mobiles* est utilisée sous plusieurs variantes. Par exemple, au lieu de calculer les nouvelles positions des centres de gravité de chaque classe après la réaffectation de tous les éléments, une des variantes consiste à modifier les positions des centres de classes après chaque réaffectation d'un élément de la population. Cette alternative peut conduire à des partitions de bonne qualité mais son inconvénient est qu'elle dépend de l'ordre des éléments dans le fichier numérique à l'entrée de l'algorithme. En 1972, la méthode d'*agrégation autour des centres mobiles* a été généralisée par *Diday* sous le nom de méthode des *nuées dynamiques*. Cette dernière consiste à modéliser chaque classe non plus par son centre de gravité mais plutôt par

un certain nombre d'éléments à classer dénommés « étalons ». Ces derniers éléments forment un « noyau » permettant de mieux décrire certaines applications que les centres de gravité [DID 72], [DID 74].

Bien qu'elle semble simple et efficace, la méthode d'*agrégation autour des centres mobiles* présente toutefois quelques limites. Son principal inconvénient est lié au fait que la partition finale dépend fortement de la partition initiale qui consiste à un choix pseudo-aléatoire des centres de classes de l'étape 0. L'algorithme garantit ainsi l'obtention de la meilleure classification à partir de la partition initiale mais rien ne garantit l'obtention d'une partition optimale globale. Le deuxième inconvénient est traduit par le fait que le nombre de classes  $N$ , jusqu'à présent considéré comme une donnée d'entrée de l'algorithme<sup>5</sup>, peut décroître au cours des itérations. Cette diminution est la conséquence de l'apparition d'une ou de plusieurs classes vides notamment lorsque leurs centres de gravité ne sont pas les plus proches au moins d'un élément de l'ensemble  $E$ .

### 2.2.2 Classification hiérarchique

Contrairement à la méthode d'*agrégation autour des centres mobiles*, la technique de *classification hiérarchique* n'aboutit pas à une unique partition en  $N$  classes mais plutôt à une hiérarchie de partitions des  $n$  éléments. Partant du particulier (éléments ou objets pris séparément) pour remonter au général (tous les éléments sont affectés à une seule classe), cette méthode permet de définir un *arbre de classification* ou *dendrogramme* contenant  $n - 1$  partitions et dont chacune est incluse dans la partition suivante. De ce fait, la *classification hiérarchique* est aussi connue sous le nom de « *classification ascendante* » [LEB 00]. Elle consiste à agréger en binôme et d'une manière successive les éléments ou les objets à classer selon un critère de distance approprié. Cependant, non seulement il faut calculer les distances entre deux éléments mais aussi les distances entre un élément et un groupe d'éléments ou parfois entre deux groupes d'éléments. Ceci nécessite alors de définir un critère d'agrégation permettant de quantifier la dissimilarité entre deux entités de nature hétérogène (par exemple un élément et un groupe d'éléments). Selon le critère d'agrégation, la hiérarchie de partitions peut être différente, ce qui rend cette méthode fortement sensible au choix du critère de regroupement. Nous donnons dans le paragraphe suivant quelques variantes du critère d'agrégation à base de calcul de distance.

---

<sup>5</sup> Nous revenons sur cette problématique un peu plus loin

- *Critères d'agrégation*

La première étape de la technique de *classification hiérarchique* est basée sur la matrice de dissimilarité  $\mathbf{D}$ . Ceci suppose que l'espace des éléments soit muni d'une distance  $d$ , par exemple la distance euclidienne, permettant de quantifier la dissimilarité entre deux objets élémentaires de l'ensemble des données. En général, c'est aussi à partir de cette distance que les distances séparant deux groupes d'objets sont calculées. La distance d'un objet élémentaire à un groupe d'objets est un cas particulier qui se produit lorsque l'un des deux groupes contient un seul élément. Les variantes les plus connues sont : le *saut minimal*, le *saut maximal*, la *distance moyenne*, la *distance des centres de gravité* et la *distance de Ward*. Pour toutes les techniques détaillées ci-dessous, nous désignons par  $C_1$  et  $C_2$  deux classes ou deux sous-ensembles non vides respectivement de taille  $n_1$  et  $n_2$ .

La technique de *saut minimal* [SNE 57], [JOH 67] définit la distance entre  $C_1$  et  $C_2$  comme étant la plus petite distance séparant deux objets élémentaires de ces deux classes. Elle est formulée de la manière suivante :

$$\delta(C_1, C_2) = \min d(E_i, E_j) \text{ avec } E_i \in C_1, E_j \in C_2 \quad (2.2)$$

Contrairement à la technique précédente, la technique de *saut maximal* est définie par :

$$\delta(C_1, C_2) = \max d(E_i, E_j) \text{ avec } E_i \in C_1, E_j \in C_2 \quad (2.3)$$

La troisième variante, souvent très utilisée, est la *distance moyenne* :

$$\delta(C_1, C_2) = \frac{\sum_{E_i \in C_1, E_j \in C_2} d(E_i, E_j)}{n_1 n_2} \quad (2.4)$$

La méthode des *centres de gravité* est aussi intéressante. Désignons par  $g_1$  et  $g_2$  les centres de gravité respectivement des classes  $C_1$  et  $C_2$ . La distance inter-classes est alors :

$$\delta(C_1, C_2) = d(g_1, g_2) \quad (2.5)$$

Enfin, la distance de *Ward* [WAR 63], appelée aussi critère d'agrégation selon l'inertie, traduit la différence de l'inertie intra-classe avant ( $I_1 + I_2$ ) et après ( $I_{12}$ ) l'agrégation des deux classes  $C_1$  et  $C_2$  en une seule classe  $C_{12}$  de centre gravité  $g_{12}$ . Son expression est alors la suivante :

$$\delta(C_1, C_2) = I_{12} - (I_1 + I_2) \quad (2.6)$$

$$\text{avec : } I_{12} = \frac{1}{n_1 + n_2} \sum_{E_j \in C_1 \cup C_2} d(E_j, g_{12}), \quad I_1 = \frac{1}{n_1 + n_2} \sum_{E_j \in C_1} d(E_j, g_1) \quad \text{et} \quad I_2 = \frac{1}{n_1 + n_2} \sum_{E_j \in C_2} d(E_j, g_2)$$

D'après [GRI 84], la distance de *Ward* semble la plus performante pour la catégorisation de documents selon certains descripteurs spécifiques.

Bien que les techniques de calcul du critère d'agrégation soient nombreuses, il n'existe pas de méthode spécifique permettant le choix du critère le plus convenable en fonction du problème traité. Pour plus de détails, nous invitons le lecteur à consulter la référence [XU 05].

- *Algorithme de classification*

Désignons par  $\mathbf{E} = \{E_1, E_2, \dots, E_n\}$  l'ensemble des  $n$  éléments à classer. Soit  $q$  la dimension de l'espace dans lequel sont définis les éléments de  $\mathbf{E}$ . Nous supposons que l'espace  $\mathbb{R}^q$  est muni d'une distance euclidienne. Nous désignons par  $C_k$  ( $k \geq 1$ ) les classes d'éléments construites au cours des différentes étapes de l'algorithme de *classification ascendante hiérarchique* dont le principe est le suivant :

La première étape de l'algorithme « Étape 1 » consiste à considérer chaque élément comme étant une classe, déterminer la matrice de dissimilarité  $\mathbf{D}$  traduisant les distances entre ces  $n$  classes puis regrouper seulement les deux classes les plus proches. Une première partition à  $n - 1$  classes est alors obtenue. Il s'agit plus précisément de  $n - 2$  classes ayant un seul élément et une seule nouvelle classe à deux éléments. Au niveau de l'étape suivante, le calcul de la matrice de dissimilarité  $\mathbf{D}$  devient plus facile. Il s'agit de rectifier la matrice  $\mathbf{D}$  calculée à l'« Étape 1 » en intégrant seulement les distances séparant la nouvelle classe (obtenue suite à l'étape 1) aux restes des classes. Les distances entre les  $n - 2$  classes, non affectées par l'« Étape 1 », restent intactes. Une fois la nouvelle matrice  $\mathbf{D}$  calculée, il suffit de regrouper de nouveau les deux classes les plus proches. Une deuxième partition à  $n - 2$  classes est alors obtenue. Le processus sera ensuite réitéré jusqu'à obtenir une seule classe contenant tous les  $n$  éléments de l'ensemble  $\mathbf{E}$ .

Nous donnons à la Figure 2.3 une illustration du principe de la *classification ascendante hiérarchique*. Nous traitons un exemple à  $n = 7$  éléments ( $\mathbf{E} = \{E_1, E_2, \dots, E_7\}$ ) dont nous présentons l'évolution des partitions suivant toutes les étapes nécessaires à la convergence de

l'algorithme. La Figure 2.4 donne l'arbre hiérarchique, appelé aussi dendrogramme, relatif à l'exemple ainsi traité. L'axe des ordonnées de ce graphe correspond au critère d'agrégation ou de distance considéré. Dans ce contexte de classification hiérarchique, les objets élémentaires  $\{E_1, E_2, \dots, E_7\}$  sont nommés « éléments terminaux ». Les classes  $\{C_8, C_9, \dots, C_{13}\}$  sont appelées « nœuds » de l'arbre. Elles peuvent regrouper deux éléments terminaux, deux nœuds ou enfin un élément terminal et un nœud.

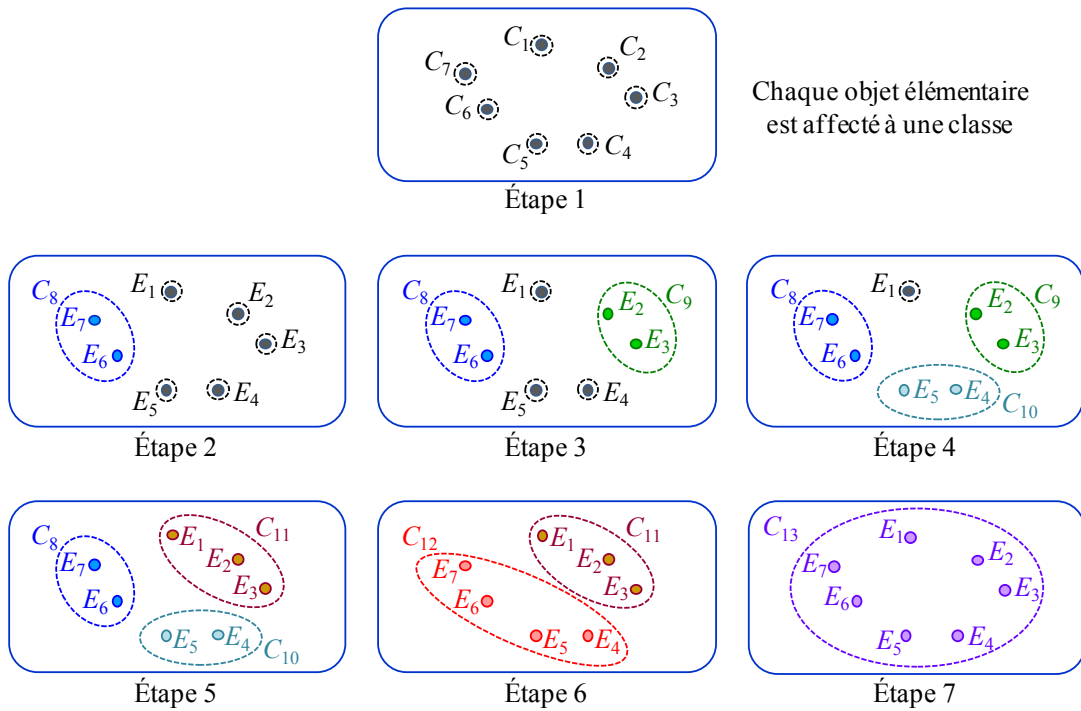


Figure 2.3. Exemple d'illustration de la classification ascendante hiérarchique

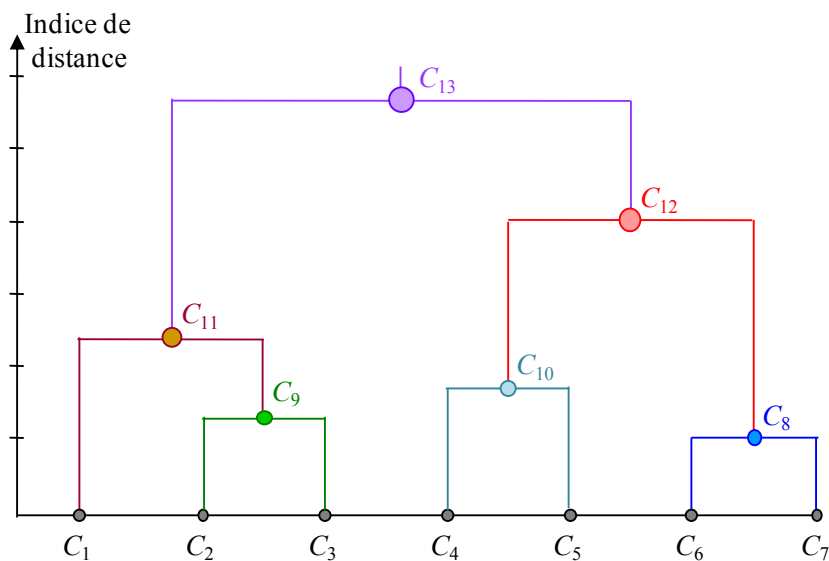


Figure 2.4. Arbre de la classification ascendante hiérarchique

La méthode de *classification ascendante hiérarchique* donne lieu à plusieurs partitions possibles. Elle apporte seulement des indications sur le nombre de classes et ne permet pas de trouver le nombre optimal de classes. Le choix d'une partition particulière demande l'expertise de l'utilisateur. Cette dernière est d'autant plus fine que le niveau de coupure horizontal de l'arbre est au voisinage des *éléments terminaux*. De ce fait, cette méthode est mal adaptée aux problèmes de classification à nombre de données élevé [LEB 00].

### 2.2.3 Classification mixte

Les deux techniques de classification que nous venons de citer peuvent être considérées comme complémentaires. En effet, contrairement à la méthode *d'agrégation autour des centres mobiles*, la technique de *classification hiérarchique* aboutit toujours aux mêmes partitions pour un même ensemble de données. Cependant, cette dernière trouve rapidement ses limites lorsqu'il s'agit de classer un grand nombre de données, configuration pour laquelle la méthode *d'agrégation autour des centres mobiles* est parfaitement adaptée. En 1982, Wong a combiné les avantages de ces deux techniques et a défini la technique de *classification mixte* (*hybrid clustering*) [WON 82] qui paraît, jusqu'à présent, la méthode la plus adaptée à un très vaste assortiment de données [LEB 00]. Cette technique a recours à la méthode *d'agrégation autour des centres mobiles* en deux temps, intercalés par une *classification ascendante hiérarchique*. Les différentes étapes de la méthode de *classification mixte* sont brièvement décrites ci-dessous et illustrées par la Figure 2.5.

#### Étape 1 : *Partitionnement initial*

Cette première étape consiste à obtenir une première partition des  $n$  éléments de l'ensemble  $\mathbf{E}$  en  $k$  classes en utilisant l'algorithme *d'agrégation autour des centres mobiles*. Le choix de la variable  $k$  est arbitraire de sorte que ce nombre de classes est à la fois très faible par rapport à la taille de la population ( $n$ ) et assez important par rapport au nombre de classes désiré à la fin de l'algorithme ( $N$ ). L'objectif de cette phase est de transformer la population initiale de taille  $n$  en une deuxième population composée de  $k$  classes et de taille beaucoup plus réduite ( $k$ ).

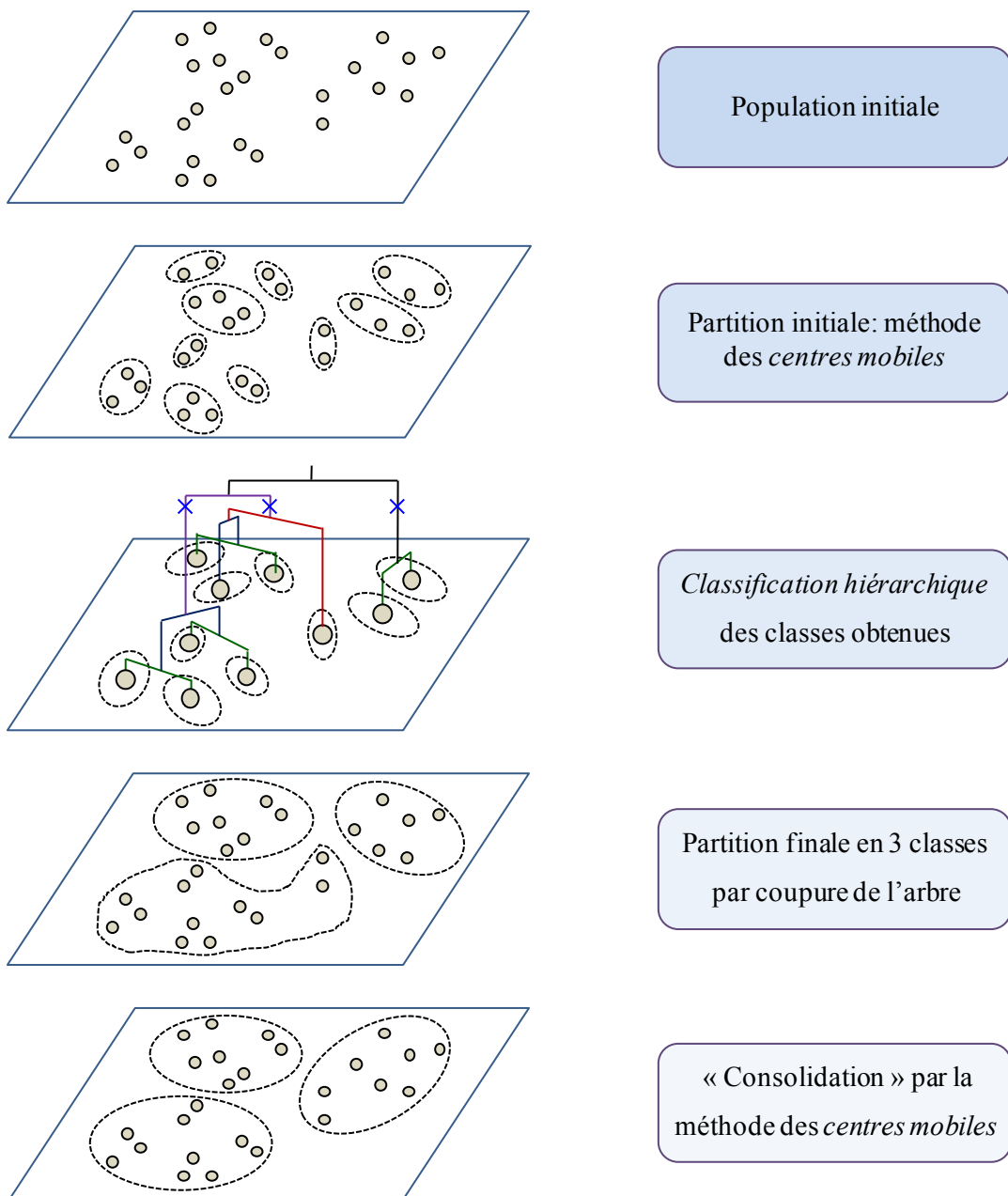
#### Étape 2 : *Classification hiérarchique des classes obtenues*

Les classes obtenues à l'«Étape 1» sont considérées comme des objets élémentaires ou plutôt des éléments terminaux pour l'étape 2 basée sur une *classification ascendante hiérarchique*. Il s'agit de remédier à certains

regroupements prématurés issus du partitionnement initial suite à un mauvais choix du nombre de classes  $k$ .

### Étape 3 : *Partitionnement final*

Le nombre de classes désiré  $N$  permet de définir le niveau de coupure de l'arbre hiérarchique de l'« Étape 2 ». Afin d'aboutir à une partition optimale à  $N$  classes, la technique de *classification mixte* fait appel, encore une fois, à l'algorithme d'*agrégation autour des centres mobiles*. Il s'agit d'une phase de consolidation des classes par réaffectation.



**Figure 2.5.** Illustration du principe de la *classification mixte* [LEB 00]



## 2.3 Classification sans *a priori* sur le nombre de classes

### 2.3.1 Problématique liée au choix du nombre de classes

Les méthodes précédentes ont l'inconvénient de nécessiter la connaissance du nombre de classes. Ce paramètre est fixé *a priori* pour la méthode *d'agrégation autour des centres mobiles* [XU 05] et déterminé par une expertise de l'arbre des partitions dans le cas de la *classification hiérarchique* [LEB 00]. Même dans le cas de la technique de *classification mixte*, le choix du nombre de classes finales ( $N$ ) et le choix du nombre de classes pour la partition initiale ( $k$ ) restent un problème incontournable.

En 1965, *Ball et Hall* ont traité la problématique du choix d'un nombre de classes optimal pour la méthode *d'agrégation autour des centres mobiles* [BAL 65]. Ils ont ainsi proposé l'algorithme ISODATA qui impose à l'algorithme des centres mobiles trois règles de gestion des classes. La première règle consiste à détruire une classe si son effectif est faible. La deuxième règle permet de fusionner deux classes si la distance séparant leurs centres de gravité est inférieure à un certain seuil. Enfin, la troisième règle permet de fragmenter en deux une classe dont les éléments possèdent un taux de dispersion supérieur à un seuil donné. Cependant, l'algorithme ISODATA reste sensible aux choix des seuils à imposer, qui varient selon l'exemple de classification traité [MON 00].

Le problème du choix du nombre de classes a été résolu suite à l'apparition de critères de partitionnement exprimant un compromis entre la maximisation des distances inter-classes et la minimisation des distances intra-classe. De nombreux critères de partitionnement peuvent être utilisés [BOL 03], [MIL 85], [SHE 05]. Les plus répandus sont l'indice de *Davies-Bouldin (DB)* [DAV 79], l'indice de *Calinski et Harabasz (CH)* [CAL 74] et l'indice de *silhouette (SIL)* [KAU 90]. Avant de décrire les expressions mathématiques de ces indices, nous définissons quelques notations génériques. Nous désignons par  $C_i$  une classe de rang  $i$ , de taille  $n_i$  et de barycentre  $g_i$ . Soit  $\mathbf{E}$  l'ensemble des  $n$  éléments à classer  $\{E_1, E_2, \dots, E_n\}$ ,  $g$  le barycentre de tous les éléments et  $N$  le nombre de classes. Nous supposons que l'ensemble des éléments est muni d'une distance  $d$ . L'expression de l'indice de *Davies-Bouldin (DB)* est la suivante :

$$DB = \frac{1}{N} \sum_{i=1}^N R_i \quad (2.7)$$

$$\text{avec } R_i = \max_{j, j \neq i} \{ (S_i + S_j) / d_{ij} \} \quad S_i = \frac{1}{n_i} \sum_{E_k \in C_i} \|E_k - g_i\|^2 \quad d_{ij} = \|g_i - g_j\|^2$$

L'expression de l'indice de *Calinski et Harabasz (CH)* est :

$$CH = \frac{\frac{\text{trace } B}{N-1}}{\frac{\text{trace } W}{n-N}} \quad (2.8)$$

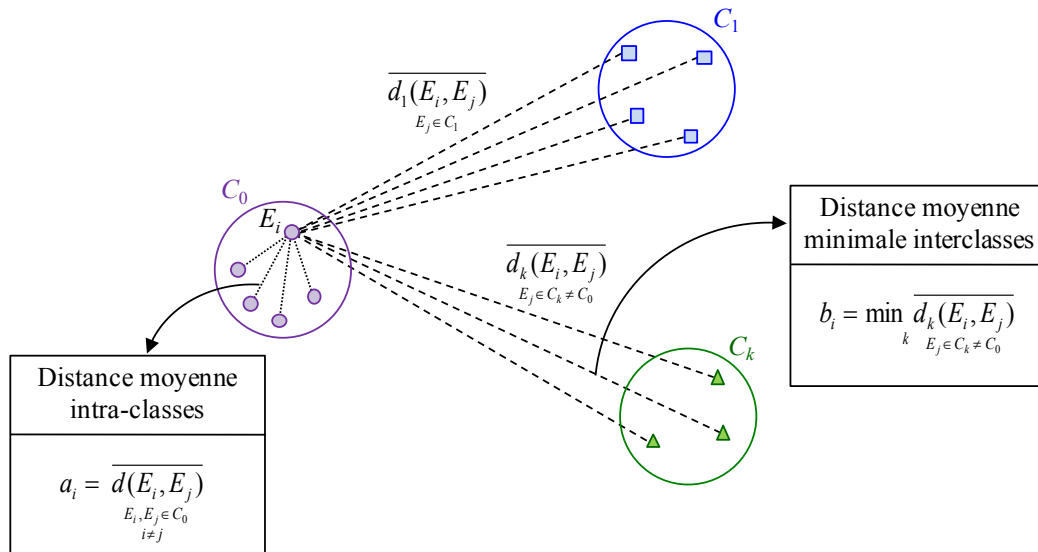
$$\text{avec } \text{trace } B = \sum_{k=1}^N n_k \|g_k - g\|^2 \quad \text{trace } W = \sum_{k=1}^N \sum_{i=1}^{n_k} n_k \|E_i - g_k\|^2$$

Pour l'indice *DB*, la classification est optimale quand ce dernier est minimal. Dans le cas de l'indice *CH*, c'est la maximisation de ce dernier qui aboutit à la classification adéquate.

La méthode de silhouette attribue à chaque élément  $E_i$  de l'ensemble des données un indice de silhouette  $s(i)$  connu sous le nom « *silhouette width* » [ROU 87] et donné par l'expression suivante :

$$s(i) = \frac{b_i - a_i}{\max(b_i, a_i)} \quad (2.9)$$

où  $a_i$  désigne la distance moyenne entre l'élément  $E_i$  et tous les éléments de sa classe d'appartenance et  $b_i$  désigne la distance moyenne minimale entre l'élément  $E_i$  et tous les éléments d'une autre classe (cf. Figure 2.6).



**Figure 2.6.** Illustration des coefficients  $a_i$  et  $b_i$  pour le calcul de l'indice de silhouette

Lorsque  $s(i)$  est proche de 1 cela indique que l'élément  $E_i$  est bien classé. Si  $s(i)$  est proche de 0, cela suggère que l'élément  $E_i$  pourrait aussi bien être affecté à un cluster voisin. Enfin, lorsque  $s(i)$  est proche de  $-1$ , nous pouvons dans ce cas conclure que l'élément est mal classé.

La qualité du partitionnement est alors exprimée par l'indice de *silhouette global* défini par la moyenne des indices de silhouette  $s(i)$  de tous les éléments de l'ensemble des données ( $n$ ) :

$$SIL = \frac{1}{n} \sum_{i=1}^n s(i) \quad (2.10)$$

Plus l'indice de silhouette global ( $SIL$ ) est proche de 1, plus la classification se rapproche de l'optimum.

La méthode de classification sans *a priori* sur le nombre de classes consiste à optimiser un ou plusieurs critères de partitionnement, en intégrant le nombre de classes comme paramètre d'optimisation [JAA 10a]. Ceci permet à la fois de déterminer le nombre de classes optimal d'une série de données et d'affecter les individus aux classes adéquates. D'une manière générale, même si le nombre de classes a été fixé au préalable, les problèmes de classification sont très multimodaux. Cette multi-modalité est d'autant plus accentuée que le nombre de classes devient variable. Face à ce genre de problèmes typiquement multimodaux, les algorithmes génétiques de nichage (*niching GAs*) offrent généralement des performances intéressantes en réduisant le risque de convergence prématurée vers un optimum local [SAR 98]. Dans le paragraphe suivant, nous présentons plus en détail l'algorithme génétique que nous avons développé pour la classification de données. Cet algorithme est basé sur l'utilisation de la sélection par tournoi restreint (*Restricted Tournament Selection* ou *RTS*) [HAR 95], ainsi qu'une technique de recombinaison auto-adaptative [NGU 08] avec comme objectif la maximisation de l'indice silhouette global.

### 2.3.2 Algorithme de classification dérivé du RTS

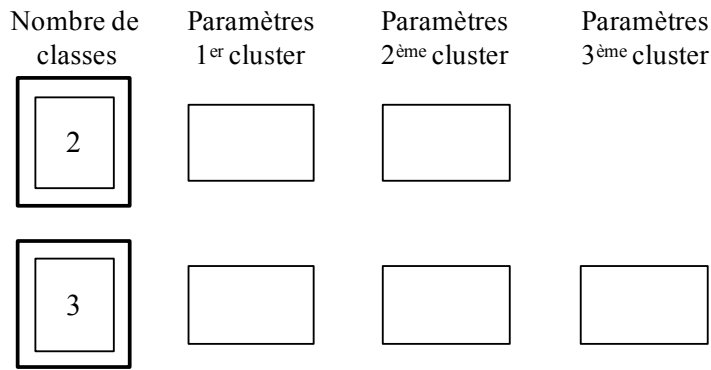
- *Représentation des individus*

La plupart des méthodes de classification à base d'algorithmes génétiques utilisent un codage binaire ou réel des objets (ou des permutations entre objets) à classer [JAI 99]. Cependant, cette représentation trouve rapidement ses limites en raison de la taille excessive du *chromosome* lorsqu'il s'agit de classer un grand nombre de données. Une solution plus

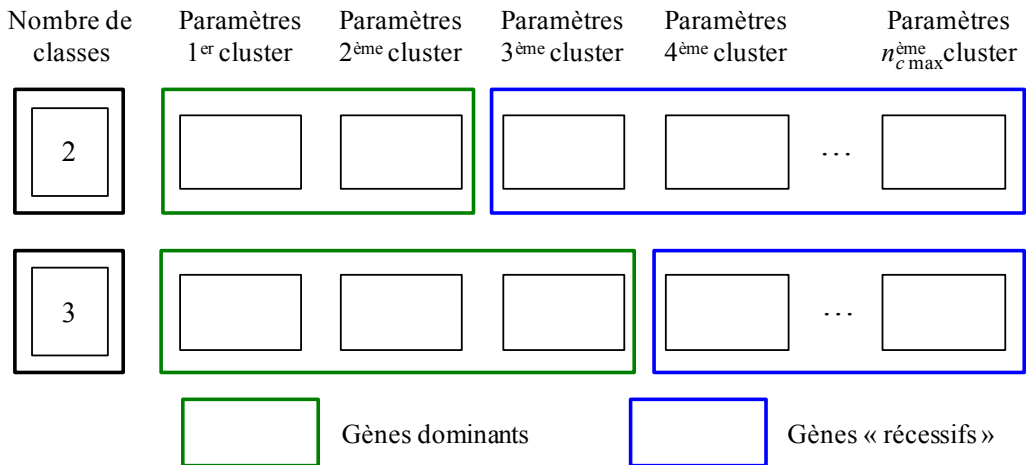
judicieuse consiste à ne coder dans le chromosome que les barycentres des classes [MAU 00]. Le partitionnement des données est alors réalisé en fonction de la position de ces barycentres. Afin d'assurer la variabilité du nombre de classes, un *gène* supplémentaire est ajouté dans le chromosome de chaque individu. Deux stratégies différentes de codage de chromosome peuvent être utilisées :

– *codage avec taille de chromosome variable* :

Après l'initialisation du gène relatif au nombre de classes, le chromosome sera construit par un choix aléatoire des barycentres des classes. Ce type de codage fait apparaître des individus caractérisés par des chromosomes de taille variable (cf. Figure 2.7.a). Il s'avère nécessaire dans ce cas de développer des opérateurs de croisement spécifiques capables d'assurer la recombinaison des chromosomes de différentes tailles.



(a) *Codage avec taille de chromosome variable*



(b) *Codage avec taille de chromosome fixe*

**Figure 2.7.** *Stratégies de codage du chromosome en fonction du nombre de classes*

– *codage avec taille de chromosome fixe* :

La deuxième stratégie consiste à accorder, à tous les individus de la population, la même taille de chromosome. Cette dernière est fixée par le nombre maximal de classes admissibles ( $n_{cmax}$ ), supposé égal à la racine carrée du nombre d'éléments à classer [PAL 95]. Lorsque le nombre de classes ( $n_c$ ), déterminé par le gène spécifique du chromosome, est inférieur au nombre de classes maximal, certains gènes peuvent être considérés comme « récessifs ». Ils ne s'expriment pas dans le décodage des individus mais ils participent aux opérations de croisement et de mutation (cf. Figure 2.7.b). En plus de sa simplicité, cette stratégie présente l'avantage de maintenir la diversité de la population grâce à l'utilisation des gènes « récessifs » avec les opérateurs de nichage.

- *Procédure de réparation*

Considérons un individu obtenu à partir de la phase d'initialisation de la population ou suite aux opérations génétiques. Chaque élément de l'ensemble des données est affecté à une classe dont le barycentre est le plus proche. A l'issue de cette étape, certaines classes dont les barycentres figurent dans le chromosome de l'individu considéré peuvent se trouver vides (car ne contenant aucun élément). Le nombre effectif de classes est alors inférieur à la valeur du gène spécifique au nombre de classes dans le chromosome de l'individu. Une telle situation sera suivie d'une procédure de réparation afin de modifier les caractéristiques du chromosome en question. Elle consiste à remplacer la valeur du gène spécifique au nombre de classes dans le chromosome de l'individu par le nombre effectif de classes. Les barycentres des classes vides seront déplacés vers la fin du chromosome. Nous donnons à la Figure 2.8 un exemple d'illustration de la procédure de réparation du chromosome. Il s'agit de classer 100 éléments avec un nombre de classes maximal  $n_{cmax} = 8$ . Le gène spécifique du nombre de classes dans le chromosome initial est égal à 5 alors que le nombre effectif de classes vaut seulement 4. Les gènes du barycentre  $g_3$  relatif à la classe vide  $C_3$  seront déplacés à la fin du chromosome.

- *Hybridation par K-means*

Dans le but d'accélérer la recherche « locale » et donc la précision de la convergence de l'algorithme génétique, il est recommandé d'appliquer [SHE 05], à chaque génération, une itération de *K-means* pour tous les individus produits par les opérateurs génétiques. Cette étape d'hybridation est introduite lors de l'évaluation de chaque individu créé (y compris lors de l'initialisation). Elle consiste, à affecter chaque élément de l'ensemble des données dans la classe représentée par le barycentre le plus proche au niveau du chromosome correspondant.

Ensuite, les barycentres des classes obtenues sont recalculés et mis à jour dans le chromosome de l'individu. Il convient de noter que l'opérateur de réparation est éventuellement appliqué à la fin de cette étape afin d'éviter l'apparition des classes vides.

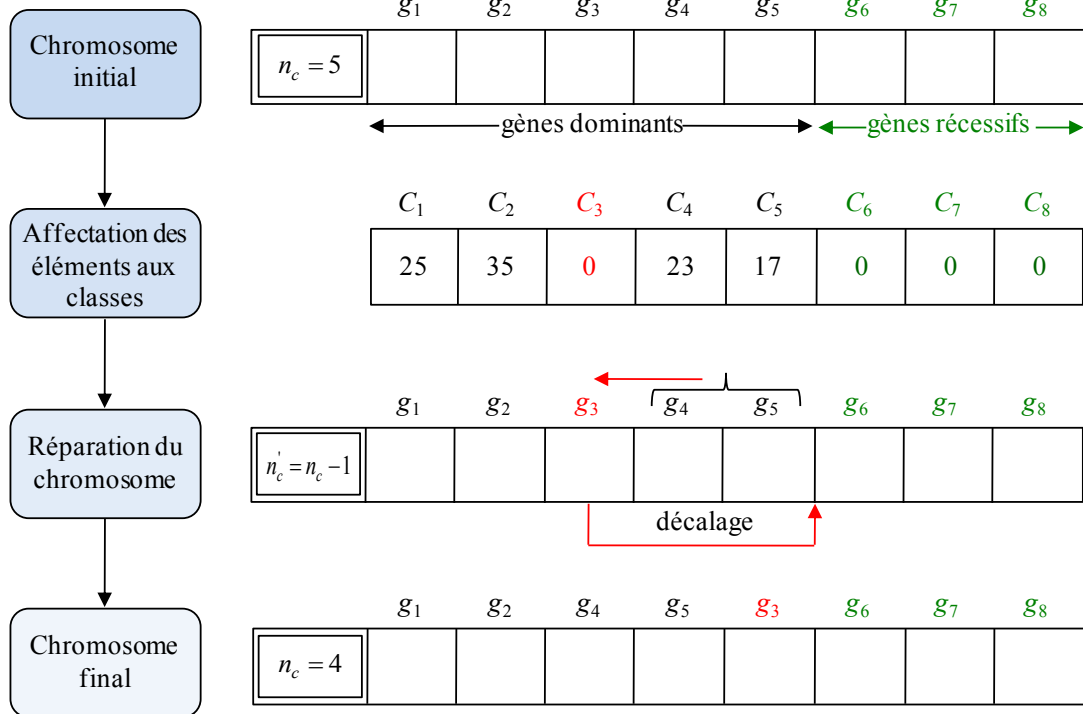


Figure 2.8. Exemple d'illustration de la procédure de réparation du chromosome

- *Fonction objectif*

Nous avons choisi d'utiliser l'indice silhouette comme fonction objectif de l'algorithme génétique, de manière à évaluer la qualité de la classification. D'autres critères de partitionnement, parmi les critères présentés précédemment, peuvent également être employés ou une combinaison de ces différents critères. Cette dernière approche a fait l'objet d'une étude particulière dans [SHE 05]. Elle a l'avantage d'être plus robuste que lorsqu'on utilise chaque critère indépendamment. Toutefois, la difficulté est alors reportée vers la formulation « agrégée » des différents critères, le choix des facteurs de normalisation et de pondération n'étant pas évident à fixer *a priori*. Nous verrons par la suite que nous n'avons pas eu à revenir sur le choix de l'indice silhouette comme critère de partitionnement, celui-ci s'étant révélé suffisamment performant dans le traitement des différents problèmes.

Pour conclure sur cette partie, nous donnons à la Figure 2.9 le pseudo-code détaillé de l'algorithme génétique développé. Nous précisons plus particulièrement la fonction d'évaluation (appliquée à tout individu généré, soit lors de l'initialisation, soit à l'issue des

opérations génétiques) incluant l'étape d'hybridation avec *K-means* et le corps général de l'algorithme.

---

Procédure d'EVALUATION d'un individu

---

1. Affecter les éléments à classer aux barycentres des classes les plus proches parmi les  $n_c$  premiers barycentres du chromosome
  2. Calculer les nouveaux barycentres (itération *K-means*) et modifier le chromosome en conséquence (modification des positions des barycentres et réparation si nécessaire)
  3. Calculer l'indice silhouette global
- 

---

Schéma général de l'algorithme génétique<sup>6</sup>

---

1. INITIALISATION

POUR chaque individu de la population

- générer aléatoirement un nombre  $n_c$  de classes :  $n_c = U(1, n_{cmax})$
- générer aléatoirement les positions des barycentres des  $n_{cmax}$  classes du chromosome
- EVALUER l'individu

2. ITÉRATIONS

POUR  $g = 1 \dots G$  générations

POUR  $i = 1 \dots N/2$  ( $N$  : taille de la population)

- Sélectionner arbitrairement 2 parents dans la population
  - Croiser les deux parents avec la procédure de recombinaison auto-adaptative [NGU 08] pour générer deux enfants
  - Muter les gènes du chromosome des enfants avec une probabilité  $p_m$  (à la fois le gène relatif au nombre de classes et les positions des barycentres des classes)
  - EVALUER les deux enfants
  - Sélectionner dans la population un échantillon de  $CF$  parents de référence
  - Réaliser un tournoi entre les enfants créés à la suite des opérateurs génétiques et les parents les plus proches présents dans l'échantillon de référence (se référer à l'algorithme *RTS* [HAR 95] décrit à l'Annexe D)
  - Réinsérer les vainqueurs dans la population
- 

**Figure 2.9.** Pseudo-code détaillé de l'algorithme génétique développé pour la classification

### 2.3.3 Exemples d'illustration

Afin d'évaluer l'efficacité de l'algorithme de classification développé, nous traitons trois exemples de données bidimensionnelles. Les variables  $X_1$  et  $X_2$  peuvent représenter les indicateurs de caractérisation des variables environnementales d'un système. Les caractéristiques de données relatives aux trois exemples étudiés sont résumées dans le [Tableau 2.2](#). Le premier « benchmark »  $S_0$  est composé de 200 éléments répartis en 12 classes

---

<sup>6</sup> Le taux de mutation  $p_m$  est généralement fixé à  $1/n$  où  $n$  désigne le nombre de paramètres du chromosome. Le facteur de surpeuplement  $CF$  (*Crowding Factor*) est fixé à une valeur de 100% pour une meilleure stabilisation des niches [SAR 98].

sans aucun recouvrement. La densité des classes varie de 4 éléments pour les petites classes à 36 éléments pour les classes les plus denses. Les deux autres « benchmarks »  $S_1$  et  $S_2$  [FRÄ 06] sont composés chacun de 5000 éléments répartis en 15 classes à densités quasi-uniformes avec différents niveaux de recouvrement (cf. Figure 2.11.a et Figure 2.12.a). L'indice silhouette global ( $SIL$ ) associé à la classification initiale de référence des 3 benchmarks est précisé dans la dernière ligne du tableau.

Benchmarks	$S_0$	$S_1$	$S_2$
Nombre d'éléments	200	5000	5000
Nombre de classes	12	15	15
Tailles des classes	{4, 4, 7, 10, 11, 13, 14, 15, 23, 29, 34, 36}	{350, 350, 350, 349, 347, 342, 341, 338, 334, 326, 325, 318, 316, 314, 300}	{350, 350, 350, 350, 346, 345, 340, 334, 333, 329, 321, 320, 317, 315, 300}
Recouvrement	Non	Oui	Oui
$SIL$	0.9488	0.8751	0.7747

**Tableau 2.2.** Caractéristiques des benchmarks bidimensionnels  $S_0$ ,  $S_1$  et  $S_2$

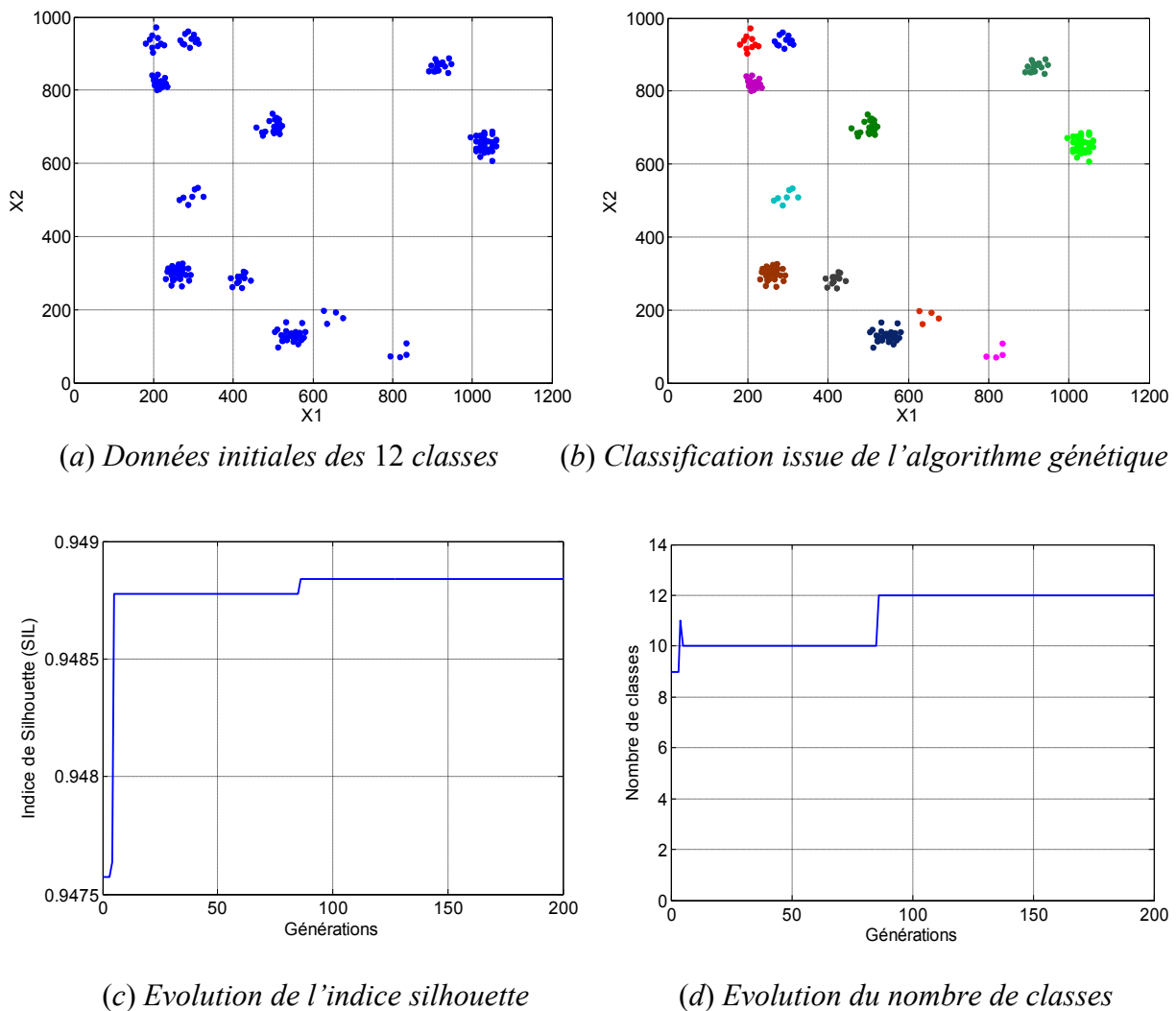
L'algorithme génétique développé est tout d'abord appliqué au premier benchmark. La taille de la population est fixée à 100 individus et le nombre de générations à 200. Les résultats typiques obtenus à la suite d'une exécution de l'algorithme sont illustrés par la Figure 2.10. Nous présentons la série de données initiale  $S_0$  (cf. Figure 2.10.a) et indiquons la classification trouvée par l'algorithme génétique (cf. Figure 2.10.b). Nous montrons aussi l'évolution de l'indice de silhouette global ( $SIL$ ) et du nombre de classes en fonction du nombre de générations (cf. Figure 2.10.c et Figure 2.10.d). Nous constatons à partir de ces figures que l'algorithme retrouve rapidement le nombre de classes adéquat ainsi que la bonne classification.

Les benchmarks suivants sont résolus en prenant une population de taille 100 et un nombre de générations égal à 300. La Figure 2.11 et la Figure 2.12 montrent la répartition initiale des données et le résultat de classification respectivement pour les familles de données  $S_1$  et  $S_2$ . Les barycentres des classes sont représentés par un cercle plein. Nous constatons pour ces deux exemples, que l'algorithme génétique résout globalement bien le problème de classification. Il convient de signaler que les classes de données initiales ne peuvent pas être retrouvées d'une manière exacte car la technique de classification utilisée (notamment l'hybridation avec *K-means*) ne permet pas d'avoir des recouvrements entre classes.

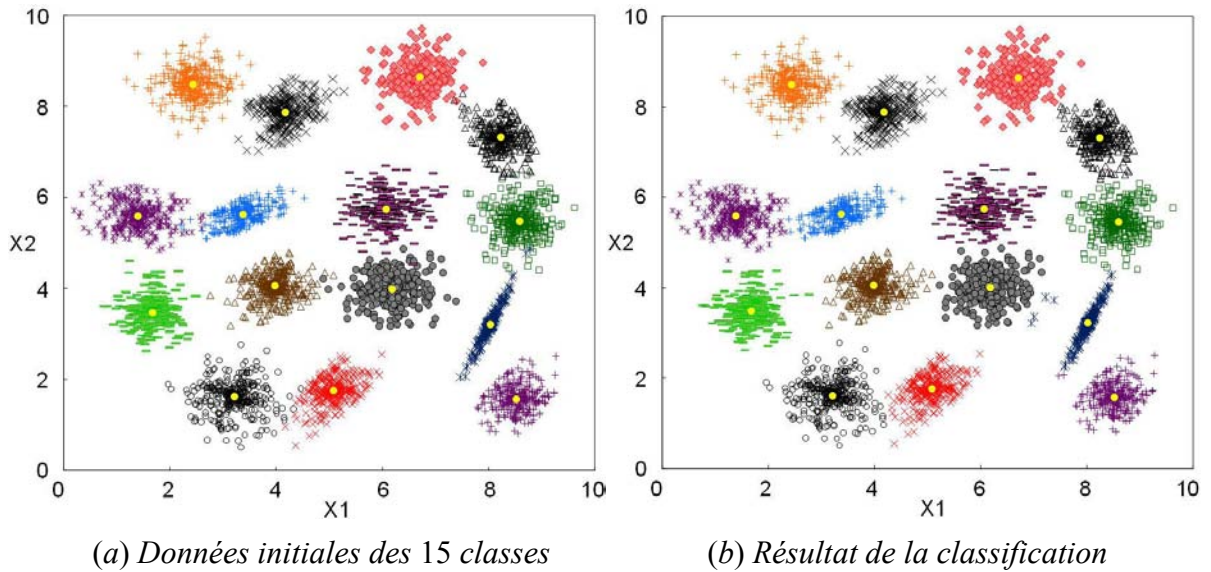


Cependant, grâce à la forte densité des classes et au faible niveau de recouvrement des données initiales, nous retrouvons presque les mêmes positions des barycentres. Le pourcentage des éléments mal classés est de 0.9% pour  $S_1$  et 3% pour  $S_2$ . Nous pouvons donc nous satisfaire du comportement de l'algorithme vis-à-vis de ce type de problèmes.

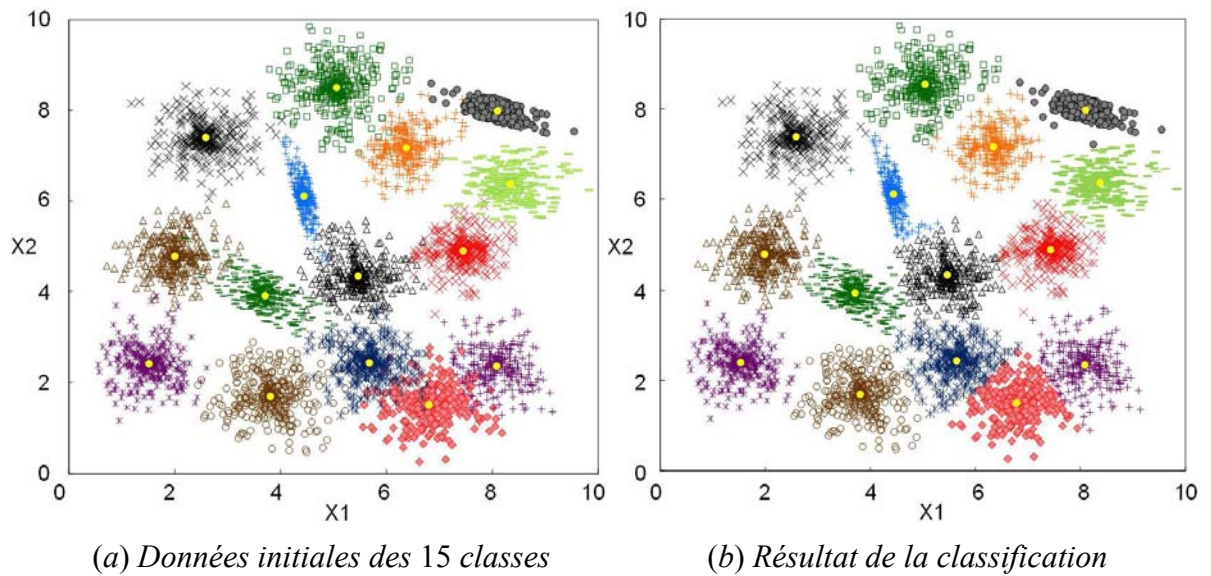
Pour conclure cette partie, nous indiquons le coût de calcul nécessaire à la résolution de ces problèmes. Le temps CPU dépend à la fois de la complexité du problème traité (nombre d'éléments) et du nombre de classes à retrouver. Sur un PC standard (*Core Duo 2 GHz*), ce temps est de l'ordre de 10 minutes pour le « benchmark »  $S_0$  avec une population de 100 individus et 200 générations. Par contre, les « benchmarks »  $S_1$  et  $S_2$  nécessitent un temps de calcul de l'ordre de 75 heures pour une population de 100 individus et 300 générations. Ce temps de calcul, relativement important, est la conséquence du grand nombre d'éléments à classer (5000 éléments).



**Figure 2.10.** Classification de l'ensemble des données  $S_0$



**Figure 2.11.** Classification de l'ensemble des données  $S_1$



**Figure 2.12.** Classification de l'ensemble des données  $S_2$

## 2.4 Application aux variables environnementales des systèmes électriques

### 2.4.1 Classification des missions pour des systèmes ferroviaires hybrides multi-segments

Dans ce paragraphe, nous étudions un exemple de classification des missions utile au dimensionnement des sources hybrides d'énergies intégrées dans des chaînes de traction ferroviaires multi-segments : le besoin (mission) spécifié ici est la puissance  $P(t)$  à fournir. L'architecture de ces sources hybrides d'énergie est basée sur un couplage entre une source d'énergie principale (par exemple un groupe Diesel ou la caténaire) avec une source secondaire représentée souvent par un organe de stockage (batteries, supercondensateurs,...).

Comme nous l'avons exposé au [chapitre 1](#), pour cette architecture particulière, une stratégie de gestion d'énergie possible consiste à produire la puissance moyenne de la mission à partir de la source d'énergie principale et le reste de la mission (partie fluctuante) par l'organe de stockage [EHS 99], [AKL 07]. Nous signalons encore que le dimensionnement d'un tel système d'alimentation dépend du triplet d'indicateurs de dimensionnement (défini au [chapitre 1](#)) : la puissance maximale  $P_{\max}$ , la puissance moyenne  $P_{\text{moy}}$  et l'énergie utile du stockage  $E_u$ .

Un « benchmark » de 105 missions ferroviaires est ici étudié afin d'illustrer l'intérêt de la classification dans un contexte de conception des systèmes hybrides d'alimentation multi segments [JAA 10a]. Il est composé de trois sous ensembles de missions relatives à trois systèmes ferroviaires différents : la locomotive BB63000, la locomotive BB460000 et les auxiliaires TGV POS (*Paris – Est de la France – Allemagne du Sud*). En ce qui concerne les deux premiers systèmes, il s'agit de deux locomotives Diesel de puissance nominale respectivement 610 kW et 1 MW. Leurs missions correspondent au besoin en puissance de chaque locomotive lors de ses trajets coutumiers. L'hybridation du système d'alimentation consiste à remplacer le groupe Diesel déjà installé par un autre de taille plus réduite et à fournir le manque de puissance par un élément de stockage. Le troisième système ferroviaire est différent des deux premiers. Il s'agit d'hybrider la source d'alimentation des auxiliaires « tronçon » du TGV. Plus précisément, ce sont les auxiliaires relatifs au confort des passagers tels que le chauffage, la climatisation, la ventilation, l'éclairage, la restauration, les installations frigorifiques du conditionnement alimentaire et les électroniques connectées au réseau de bord des voitures. Tous les auxiliaires, à l'exception du chauffage, sont alimentés par une ligne continue 500 V. Une ligne continue spéciale de 1800 V permet d'alimenter les résistances de chauffage de toutes les voitures. Ces deux lignes sont issues d'un « bloc auxiliaires » à base de deux ponts monophasés à commutation forcée connectés aux deux secondaires du transformateur principal et mis en parallèle pour définir la ligne 1800 V. Un hacheur abaisseur garantit la ligne 500 V. L'hybridation de ces auxiliaires consiste alors à embarquer des briques de stockage qui seront sollicitées en l'absence ou en présence de la caténaire. Dans ce dernier cas<sup>7</sup>, la stratégie de gestion d'énergie décrite précédemment reste valable. Dans le cas de perte de caténaire, le stockage pourra garantir un minimum de fonctionnement des auxiliaires TGV<sup>8</sup>. Les missions de ce système correspondent alors à la

---

<sup>7</sup> En présence de caténaire

<sup>8</sup> Par exemple, la climatisation est l'auxiliaire le plus prioritaire

consommation totale des auxiliaires TGV pour des trajets ordinaires. Pour plus de détails sur l'hybridation des auxiliaires TGV, nous invitons le lecteur à consulter l'[Annexe B](#).

Toutes les missions sont caractérisées selon le triplet d'indicateurs de dimensionnement  $\{P_{\max}, P_{\text{moy}}, E_u\}$  introduit au [chapitre 1](#). Le nombre de missions, par classe d'appartenance initiale, ainsi que les positions des barycentres correspondants sont résumés dans le [Tableau 2.3](#).

	Nombre de missions	Coordonnées du barycentre ( $P_{\max}$ (kW), $P_{\text{moy}}$ (kW), $E_u$ (kWh))
Auxiliaires TGV	63	(189, 80, 13)
BB 63000	15	(455, 91, 24)
BB 460000	27	(711, 35, 50)

**Tableau 2.3.** *Caractéristiques des 105 missions initiales des trois systèmes ferroviaires*

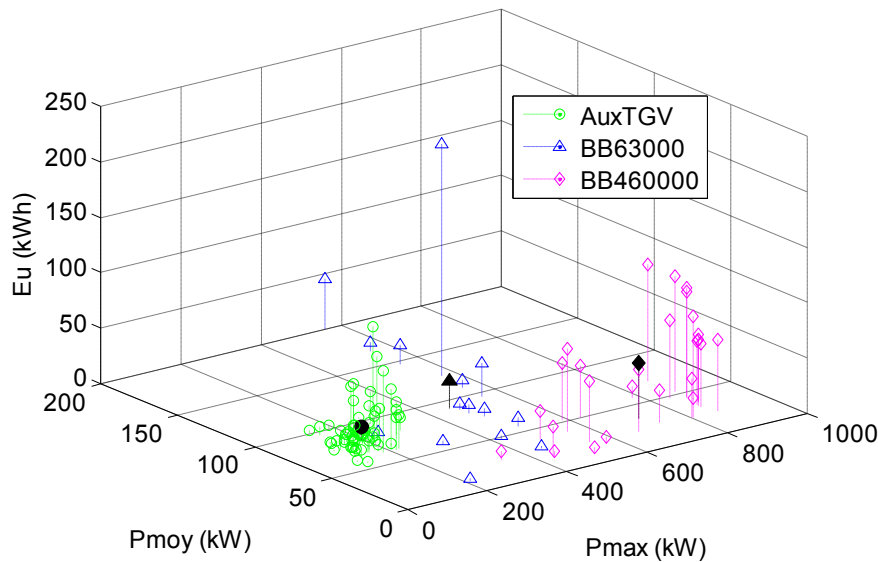
La représentation des sous ensembles des missions initiales selon le triplet  $\{P_{\max}, P_{\text{moy}}, E_u\}$  est donnée par la [Figure 2.13.a](#). Leurs barycentres sont aussi représentés par des motifs pleins. Le résultat de la classification obtenu en appliquant l'algorithme génétique développé dans la section précédente est donné par la [Figure 2.13.b](#). Les caractéristiques des classes trouvées sont également données par le [Tableau 2.4](#). Le nombre de générations est fixé à 500 et la taille de la population est égale à 100. Le temps de calcul correspondant est de l'ordre de 13 minutes sur un PC standard *Core Duo* 2 GHz. Nous comparons à la [Figure 2.13](#) la classification des données initiales avec le partitionnement issu de l'algorithme génétique.

	Nombre de missions	Coordonnées du barycentre ( $P_{\max}$ (kW), $P_{\text{moy}}$ (kW), $E_u$ (kWh))
Classe 1	76	(225, 75, 13)
Classe 2	6	(553, 141, 53)
Classe 3	23	(762, 37, 56)

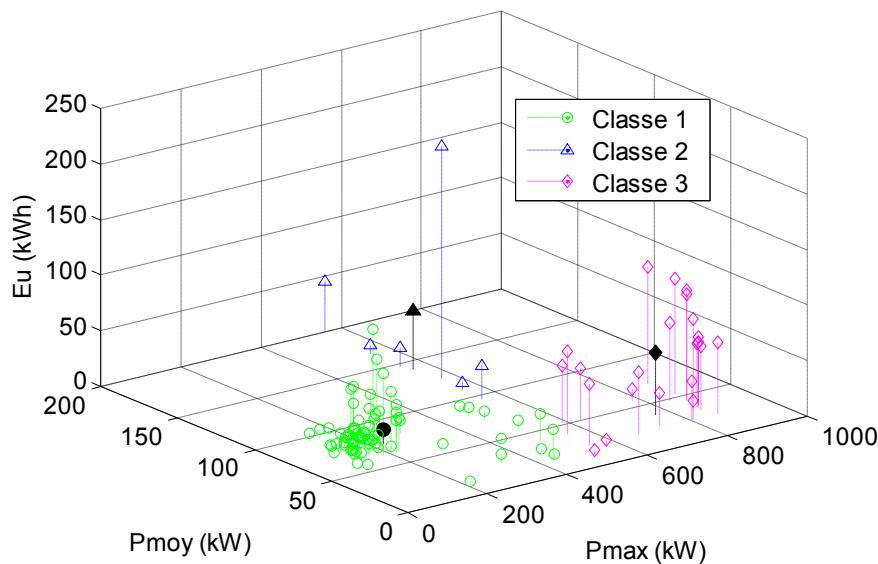
**Tableau 2.4.** *Caractéristiques des classes obtenues après la classification*

Nous constatons que l'on retrouve pratiquement la bonne classification en identifiant trois classes différentes. La différence entre la classification initiale et celle issue de l'algorithme génétique est seulement de 12% (13 missions ne sont pas bien affectées à leurs classes d'origine). Cette différence n'est pas due à une mauvaise convergence de l'algorithme. En

effet, bien que l'on dispose de trois systèmes d'alimentation différents, rien n'empêche que leurs missions présentent des indicateurs similaires sur le plan d'énergie/puissance. La [Figure 2.13.b](#) montre que toutes les missions sont globalement bien classées à l'exception de celles appartenant à la zone de recouvrement des sous ensembles initiaux. Dans cette région, de faible puissance et énergie, les missions peuvent être réalisées par les trois systèmes d'alimentation.



(a) Répartition initiale des missions



(b) Résultat de la classification

**Figure 2.13.** Exemple de classification des missions de trois systèmes ferroviaires

La plupart des missions de la zone de recouvrement sont attirées par la classe la plus dense regroupant les missions des auxiliaires TGV. Par conséquent, nous obtenons une déviation

importante du barycentre du sous ensemble des missions de la locomotive BB63000 au vu de sa faible densité. En ce qui concerne les deux autres classes de grandes tailles (Aux. TGV et BB460000), la déviation des barycentres reste négligeable.

A travers cet exemple, nous avons non seulement validé l'efficacité de la méthode de classification mais aussi montré l'utilité d'une telle démarche dans un processus d'analyse des variables environnementales (ici des profils de mission) en amont de la conception systémique. La classification peut alors faire partie d'un outil d'aide à la décision en l'absence d'*a priori* sur les profils de la variable d'environnement mise en jeu. Par exemple, supposons que nous n'ayons aucune information sur les classes initiales d'appartenance (BB63000, BB460000 et Aux. TGV) des 105 missions ferroviaires ainsi traitées. C'est alors par le biais de la classification que nous retrouvons trois segments différents. Dans une telle configuration, le concepteur se posera la question de la rentabilité de concevoir un seul système d'alimentation hybride satisfaisant toutes les missions ou, à l'opposé, un système par classe de missions. Nous traiterons, au [chapitre 4](#), un exemple concret de conception systémique exploitant la classification et la segmentation.

D'autre part, le fait de retrouver, par la voie de la classification, trois familles de missions correspondant bien aux classes initiales, prouve la pertinence du triplet d'indicateurs de caractérisation des missions  $\{P_{\max}, P_{\text{moy}}, E_u\}$ . Nous verrons dans le [chapitre 3](#) comment ces indicateurs serviront de base à la synthèse d'une mission représentative d'une classe ou famille de missions.

#### **2.4.2 Classification des profils de vitesse du vent**

Dans ce paragraphe, nous appliquons la même démarche de classification par algorithme génétique sur un autre type de variable environnementale d'un système. Il s'agit de classer des profils de la vitesse du vent pour deux sites géographiques différents : le site de « Sidi Daoued » en Tunisie et le site de « Petit Canal » en Guadeloupe (France). Pour chacun des deux sites, nous disposons de deux mois de mesures successives de la vitesse du vent. Dans l'objectif de valider la pertinence des indicateurs de caractérisation de la vitesse du vent proposés au [chapitre 1](#), nous choisissons de classer des profils journaliers de la vitesse du vent. Ceci consiste alors à déterminer les indicateurs de caractérisation pour chaque journée (24 heures) de vitesse du vent.

Nous rappelons au lecteur que les indicateurs de caractérisation de la vitesse du vent que nous avons proposés au [chapitre 1](#) sont la vitesse maximale ( $V_{\max}$ ), la vitesse cubique moyenne ( $\langle V^3 \rangle$ ), la distribution statistique ou la densité de probabilité de la vitesse du vent ( $pdf$ ) et l'indice de turbulence ( $I_{turb}$ ). Dans le cadre de ce « benchmark », nous ne considérons que les deux premiers indicateurs de dimensionnement de la chaîne éolienne  $V_{\max}$  et  $\langle V^3 \rangle$ . En ce qui concerne l'indice de turbulence, ce dernier est défini par la moyenne des écarts type des distributions gaussiennes de la vitesse du vent sur des intervalles de largeur 10 minutes. En raison de l'inadéquation des pas d'échantillonnage des mesures de vitesse du vent disponibles, cet indicateur n'a pas pu être exploité pour ce « benchmark ». Pour l'indicateur statistique, il n'est pas évident de traduire la densité de probabilité de la vitesse du vent par un seul indicateur pertinent que nous pouvons intégrer dans l'espace métrique<sup>9</sup> des données à classer. A ce propos et avec une approche purement statistique, une étude réalisée dans [\[CAL 08\]](#) a permis d'établir une classification des échantillons de la vitesse du vent de durée 10 minutes en considérant comme indicateurs les moments statistiques d'ordre  $n$ .

Le « benchmark » ainsi étudié consiste alors à reclasser 122 jours<sup>10</sup> de vitesse de vent caractérisés selon le doublet d'indicateurs de dimensionnement de la chaîne éolienne  $\{V_{\max}$  et  $\langle V^3 \rangle\}$ . Le nombre de jours par classe d'appartenance initiale ainsi que les positions des barycentres correspondants sont résumés dans le [Tableau 2.5](#).

	Nombre de jours	Coordonnées du barycentre ( $V_{\max}$ (m/s), $\langle V^3 \rangle$ ( $m^3/s^3$ ))
« Sidi Daoued »	61	(8.83, 306.9)
« Petit Canal »	61	(16, 1075)

**Tableau 2.5.** *Caractéristiques des classes initiales des journées du vent*

La représentation des journées du vent relatives aux deux sites géographiques selon les indicateurs de vitesse maximale ( $V_{\max}$ ) et de vitesse cubique moyenne ( $\langle V^3 \rangle$ ) est donnée par la [Figure 2.14.a](#). Les barycentres des classes initiales sont représentés par des motifs pleins. L'application de l'algorithme génétique avec une taille de population de 100 et un nombre d'individus de 200 conduit à la répartition de la [Figure 2.14.b](#). Le temps de calcul nécessaire à la résolution de ce problème est de l'ordre de 4 minutes sur un PC standard *Core Duo* 2 GHz.

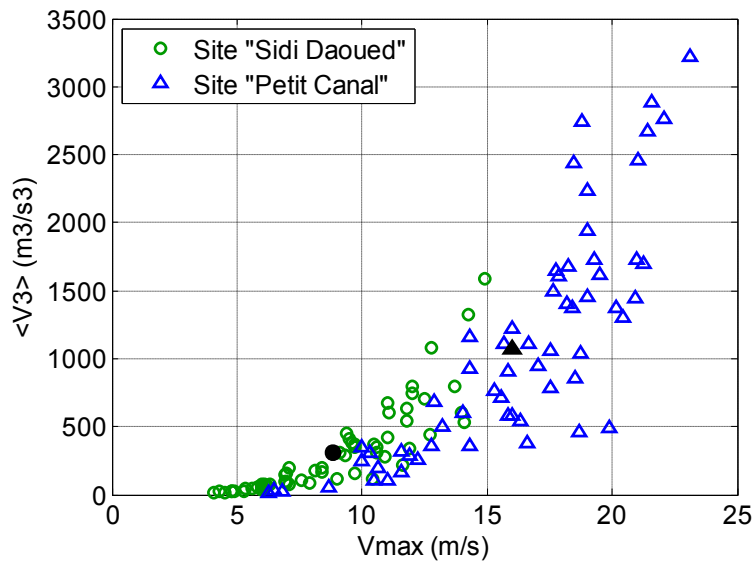
<sup>9</sup> En mathématiques, un espace métrique est un ensemble au sein duquel une notion de distance entre les éléments de l'ensemble est définie.

<sup>10</sup> 61 jours (2 mois) par site géographique

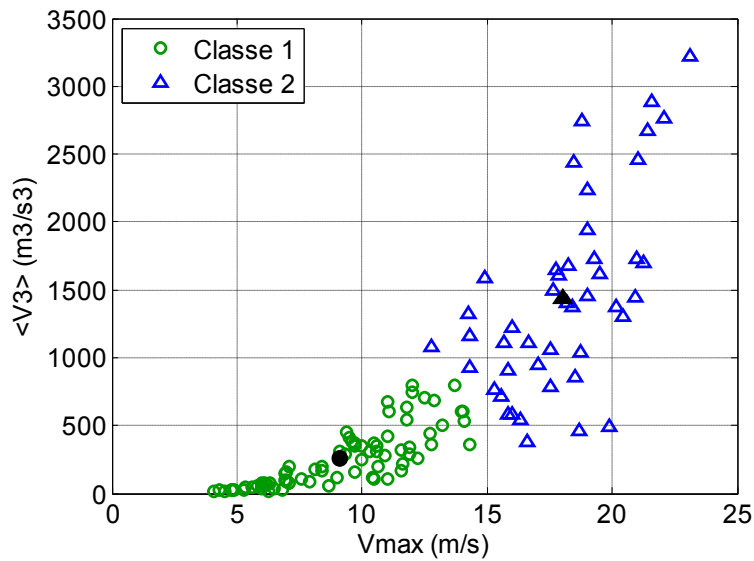
Le nombre de jours ainsi que les coordonnées des barycentres des classes obtenues sont résumés dans le [Tableau 2.6](#).

	Nombre de jours	Coordonnées du barycentre ( $V_{max}$ (m/s), $\langle V^3 \rangle$ (m <sup>3</sup> /s <sup>3</sup> ))
Classe 1	77	(9.11, 256.1)
Classe 2	45	(18.05, 1435)

**Tableau 2.6.** *Caractéristiques des classes obtenues après la classification*



(a) *Répartition initiale des journées de vitesse du vent*



(b) *Résultat de la classification*

**Figure 2.14.** *Classification des journées de vitesse du vent pour le site de « Sidi Daoued » en Tunisie et le site de « Petit Canal » en Guadeloupe*



La [Figure 2.14.a](#) montre que les journées du site de « Petit Canal » couvrent un domaine plus large en termes de vitesse maximale et de vitesse cubique moyenne que les journées relatives au site de « Sidi Daoued ». Ces dernières sont plutôt concentrées vers les faibles valeurs de vitesse maximale et de vitesse cubique moyenne. Les deux familles initiales de journées de vent présentent donc une zone de recouvrement qui correspond à une forte similarité entre certaines journées géographiquement différentes. Il en résulte alors une « classification » de 16 journées de vitesse du vent pour lesquelles on ne retrouve pas l'origine géographique des sites, ce qui correspond seulement à 13% de différence entre la classification géographique initiale et celle obtenue par l'algorithme génétique.

Dans la mesure où les données de vitesses du vent couvrent des horizons temporels assez larges (quelques années), ce travail pourra servir de référence pour mettre en place un outil d'aide à la décision pour la conception de chaînes éoliennes par classes de vent. A travers une modélisation systémique d'une chaîne éolienne, le concepteur pourra évaluer, sur une certaine base de critères, l'intérêt de concevoir une seule chaîne pour les deux sites géographiques ou une chaîne optimisée par site. Nous reviendrons au chapitre suivant sur l'intérêt de cette problématique en traitant de la synthèse d'un profil de vitesse de vent compact et représentatif d'une classe de profils. Le profil « fictif » issu de cette étape de synthèse « compacte » pourra être exploité avantageusement dans un processus de conception par optimisation en réduisant notablement les temps de conception.

## 2.5 Conclusion

Dans ce chapitre, nous avons exposé différentes variantes des techniques de classification. Nous avons mis en évidence la limite des méthodes « *exclusives non supervisées* » usuelles qui nécessitent de fixer *a priori* le nombre de classes. Nous avons contourné ce problème en utilisant une approche basée sur l'optimisation d'un critère de partitionnement (l'indice silhouette) associé à un algorithme génétique avec nichage. Cette approche permet de discerner une classification optimale avec un bon compromis entre la maximisation des distances inter-classes et la minimisation des distances intra-classes. Après avoir validé l'algorithme de classification développé sur trois « benchmarks » bidimensionnels, nous avons traité la classification de deux types de variables environnementales : des profils de missions ferroviaires dédiés à trois systèmes d'alimentation distincts et des profils de vitesses de vent issus de deux sites géographiques différents. Nous montrerons dans le [chapitre 4](#) l'intérêt d'un tel processus de classification dans une démarche de conception systémique.

## Chapitre 3

# Synthèse de profils de missions et de variables environnementales

### 3.1 Introduction

A l'issue de la méthodologie de classification des profils de missions et de variables environnementales présentée dans le chapitre précédent, la question qui se pose au niveau de ce chapitre est de savoir comment intégrer au mieux l'ensemble des profils d'une même classe dans une démarche de conception systémique. Autrement dit, nous cherchons à représenter le contenu informationnel de l'ensemble des profils par un seul *profil fictif* représentatif et de durée réduite afin de minimiser le coût de simulation dans un processus de conception systémique en particulier dans le contexte de la conception simultanée par optimisation. Dans ce chapitre, nous développons une méthode de synthèse d'un profil temporel permettant de construire un signal dont les indicateurs de conception correspondent aux indicateurs de référence, ces derniers étant déterminés à partir d'un ensemble de profils réels. Le profil synthétisé est construit à partir d'un modèle de signal combinant des motifs élémentaires (segments, sinus ou sinus cardinaux). Le paramétrage des motifs du signal est déterminé par résolution d'un problème inverse à l'aide d'un Algorithme Évolutionnaire (AE), en minimisant une fonction d'erreur permettant de garantir au profil fictif généré, les caractéristiques des indicateurs de référence imposés.

Dans une première partie de ce chapitre, nous décrivons plus en détail ce processus de synthèse exploitant trois modèles de signaux différents en fonction du type de motif utilisé. Dans une deuxième partie, nous critiquons le choix du type et du nombre de motifs élémentaires en fonction de la complexité du profil de la mission ou de la variable environnementale étudiée. Nous évaluons également la capacité des modèles de signaux proposés et les performances de trois algorithmes évolutionnaires différents, d'une part pour

reproduire un profil réel et d'autre part pour la régénération d'un profil à caractéristiques données. Nous nous appuyons dans les deux cas sur un exemple de mission ferroviaire. Enfin, nous appliquons notre processus de synthèse compacte de profils relatifs à deux types de variables environnementales. Le premier exemple consiste à synthétiser une mission fictive et compacte représentative d'une famille de missions de la locomotive BB63000 alors que la deuxième application a pour objectif de compacter deux mois de mesures de la vitesse du vent sur le site « Petit Canal » en Guadeloupe.

### 3.2 Processus de synthèse de profil de mission ou de variable environnementale

Comme nous l'avons signalé dans l'introduction de ce chapitre, nous procédons à la reconstruction d'un profil de mission ou d'une variable environnementale par concaténation ou agrégation de motifs élémentaires. A travers un paramétrage adéquat de ces motifs, le problème pourra être résolu par optimisation. Il s'agit en fait de trouver le bon jeu de paramètres des motifs de façon à ce que le signal obtenu satisfasse un ensemble donné d'indicateurs de conception de la variable environnementale en question. Trois types de motifs élémentaires sont mis en œuvre : le sinus, le sinus cardinal et le segment. Nous revenons, dans la section suivante, sur les détails de ces motifs et sur la manière avec laquelle ils sont combinés.

Le fait de générer un signal complexe avec une forte variabilité nécessite l'augmentation du nombre de motifs, ce qui accroît notablement la complexité du problème d'optimisation. De plus, en raison de la forme particulière oscillante des motifs proposés (spécialement dans le cas du sinus et du sinus cardinal), le problème inverse résultant est souvent multimodal. Comme dans le cas de la classification abordé au chapitre précédent, les algorithmes évolutionnaires sont ici particulièrement bien adaptés pour résoudre ce type de problèmes. Dans ce chapitre, nous comparerons l'efficacité de trois algorithmes évolutionnaires. Le premier algorithme est basé sur une Stratégie d'Evolution standard (ES) [SCH 95], alors que les deux autres algorithmes présentent deux techniques de *nichage* [SAR 98] : il s'agit plus précisément de la méthode d'Éclaircissement ou « Clearing » (CLR) [PET 96] et de la Sélection par Tournoi Restreint (RTS) [HAR 95]. Ces deux dernières méthodes sont associées à une procédure de recombinaison auto-adaptative [NGU 08].

Avant de décrire le processus de synthèse de profil, nous attirons l'attention du lecteur sur l'emploi de quelques termes fréquemment utilisés dans ce chapitre. Nous désignons par  $N_m$  le

nombre de motifs élémentaires (segments, sinus, sinus cardinaux) à sommer ou à concaténer pour obtenir un profil synthétisé complet. Chaque motif  $M_i$  est caractérisé par  $p$  paramètres  $\{x_{i1}, \dots, x_{ip}\}$ . Nous notons  $\mathbf{X}$ , le vecteur contenant successivement les paramètres des  $N_m$  motifs (soit  $p \times N_m$  paramètres).

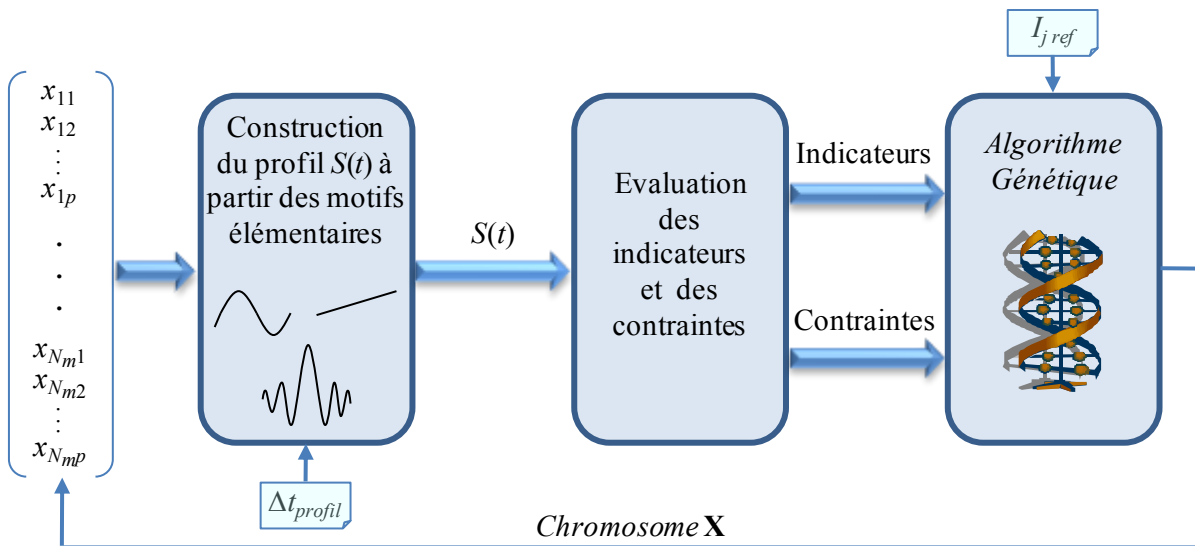
Le processus de synthèse de profil par optimisation est donné par la [Figure 3.1](#). La première étape consiste à construire le signal ou le profil  $S(t)$  de la variable environnementale à partir du chromosome généré par l'algorithme évolutionnaire. La durée  $\Delta t_{profil}$  du signal obtenu est considérée comme une entrée du problème. Bien que nous veillions à ce que le profil synthétisé soit aussi compact que possible, le choix de sa durée doit vérifier certaines contraintes. Nous montrerons un peu plus tard les couplages pouvant exister entre la durée du profil et ses indicateurs de conception. Au niveau de l'étape suivante, nous procédons au calcul des indicateurs du profil ainsi généré. Ces derniers sont ensuite comparés aux indicateurs de référence  $I_{jref}$  et permettent, dès lors, d'évaluer la *fonction d'erreur*  $\varepsilon$  (fonction à minimiser par l'algorithme évolutionnaire) exprimée par l'équation (3.1). Cette fonction est définie comme la somme des erreurs quadratiques normalisées de l'ensemble des indicateurs :

$$\varepsilon = \sum_j \left( \frac{I_j - I_{jref}}{I_{jref}} \right)^2 \quad (3.1)$$

Au-delà de la distance quadratique entre indicateurs et valeurs de référence, il peut être nécessaire d'imposer certaines contraintes au signal généré. Ces contraintes (dénotées  $C_k$ ) sont formulées en termes d'inégalités ( $C_k(\mathbf{X}) \leq 0$ ) et intégrées dans la fonction d'erreur sous forme de pénalisation :

$$\varepsilon = \sum_j \left( \frac{I_j - I_{jref}}{I_{jref}} \right)^2 + \sum_k \lambda_k \max(0, C_k(\mathbf{X}))^2 \quad (3.2)$$

où  $\lambda_k$  est un facteur de pénalité associé à la contrainte correspondante. La valeur de ce coefficient doit être judicieusement choisie de manière à garantir un « poids » équilibré entre les différentes contraintes lorsque celles-ci sont violées et une pénalisation suffisamment importante par rapport aux erreurs sur les différents indicateurs.



**Figure 3.1.** Processus de synthèse de profil d'une mission ou d'une variable environnementale

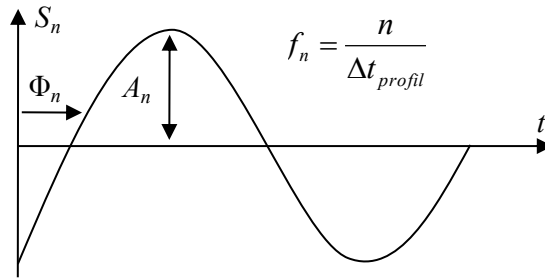
### 3.2.1 Motifs élémentaires pour la génération de profil

Dans cette section, nous revenons, d'une part, sur la description et le paramétrage des motifs élémentaires et d'autre part sur la manière de reconstruire le signal synthétisé  $S(t)$ . Nous cherchons à modéliser le profil temporel réel de la mission ou de la variable environnementale par des motifs élémentaires à faible nombre de paramètres. En fait, plus le nombre de paramètres d'un motif est important, plus la complexité du problème de synthèse de signal par optimisation est élevée. Trois types de motifs élémentaires sont proposés pour représenter le signal correspondant à la variable environnementale :

- *Modèle de type sinus*

Une première modélisation du signal ou de profil d'une mission ou d'une variable environnementale est basée sur la théorie de la décomposition en série de Fourier. Le profil  $S(t)$  est défini comme étant la somme d'une valeur moyenne et d'un certain nombre d'harmoniques caractérisés par leurs amplitudes, phases et fréquences. Un motif élémentaire n'est autre qu'un harmonique de rang  $n$  caractérisé par trois paramètres : une amplitude  $A_n$ , une phase  $\Phi_n$  et une fréquence  $f_n$  multiple de l'inverse de la durée désirée du profil<sup>11</sup>  $\Delta t_{profil}$  comme le montre la [Figure 3.2](#).

<sup>11</sup> Nous supposons que le profil de la variable d'environnement est périodique



**Figure 3.2.** Paramètres d'un motif de rang  $n$  :  $A_n$ ,  $f_n$  et  $\Phi_n$

La méthode des sinus nécessite de synthétiser la valeur moyenne du signal à générer. Dans certains problèmes de conception, cette dernière peut jouer le rôle d'un indicateur de conception de la variable environnementale traitée. Sa valeur de référence  $S_{moy\ ref}$  est alors bien déterminée et sera ajoutée directement à la somme des  $N_m$  harmoniques de moyenne nulle. A titre d'exemple, dans le cas de la conception systémique d'une chaîne de traction hybride couplant une source d'énergie principale et un organe de stockage d'énergie, la valeur moyenne du profil de mission est considérée comme un indicateur pertinent vis-à-vis de la conception. En effet, avec une stratégie de gestion d'énergie fréquentielle, la moyenne de la mission est assurée par la source d'énergie principale et le stockage fournit les harmoniques de la mission. Dans un tel cas de figure, le signal de mission  $S(t)$  s'écrit :

$$S(t) = S_{moyref} + \sum_{n=1}^{N_m} A_n \sin(2\pi f_n t + \Phi_n) \quad (3.3)$$

L'indicateur de la valeur moyenne  $S_{moy}$  est automatiquement vérifié et il ne sera pas intégré comme un critère d'optimisation dans la fonction d'erreur.

Au contraire, lorsque la valeur moyenne du profil de mission ou de la variable environnementale du système n'est pas significative au sens des indicateurs de conception, le choix de  $S_{moy\ ref}$  devient très délicat. Dans ce cas, nous considérons la valeur moyenne du signal à construire comme un paramètre d'optimisation. La bonne valeur de ce dernier est alors celle qui permettra au signal global de satisfaire au mieux l'ensemble des indicateurs de référence. Par exemple, dans le cas du traitement de la vitesse du vent pour la conception d'une chaîne éolienne, la valeur moyenne de cette variable environnementale n'est pas considérée explicitement comme un indicateur de conception. C'est plutôt la valeur moyenne du cube de la vitesse du vent dont nous devons tenir compte dans la synthèse du profil

représentatif car celle-ci est proportionnelle à la puissance délivrée par la masse d'air traversant la section des pales de l'éolienne.

Dans la [section 3.3](#), nous montrerons que, pour reproduire les principales caractéristiques de signaux compliqués à forte variabilité, le nombre d'harmoniques nécessaire devient important. Avec cette approche, la complexité du problème d'optimisation ( $p = 3$ , soit  $3 \times N_m$  paramètres) est alors assez élevée rendant la résolution délicate, tant du point de vue des temps de convergence que de la qualité des solutions trouvées. Ainsi, nous proposons de réduire les dimensions du problème à  $2 \times N_m$  paramètres en ne considérant que les  $N_m$  premiers harmoniques pris successivement. Autrement dit, nous exemptons l'algorithme d'optimisation de la recherche des fréquences ou des rangs des motifs. Ces dernières sont alors les multiples, jusqu'au rang  $N_m$ , de la fréquence du fondamental définie par l'inverse de la durée du profil à générer ( $1/\Delta t_{profil}$ ). En ce qui concerne les plages de variation des paramètres, nous faisons varier les phases  $\Phi_n$  des motifs entre 0 et  $2\pi$  et les amplitudes  $A_n$  entre 0 et la valeur maximale du signal à générer  $S_{max}$ .

Pour finir sur le motif élémentaire en sinus, nous signalons au lecteur que cette méthode doit être adaptée en fonction des caractéristiques des signaux à générer. Par exemple, dans le cas de la synthèse des profils de missions ferroviaires, le signal à générer (ici la puissance à fournir) doit être forcément supérieur ou égal à 0 ( $S(t) \geq 0$ ). Une première alternative consiste à intégrer cette caractéristique comme contrainte dans la fonction d'erreur, comme indiqué par l'équation (3.2). Ceci étant, dans le cas d'application aux missions ferroviaires où nous imposons *a priori* la valeur moyenne du signal  $S_{moy\ ref}$  (cf. équation (3.3)), nous avons constaté que la plupart des solutions proposées par l'algorithme d'optimisation (individus d'une génération donnée), sont rejetées car ne respectant pas cette contrainte. En effet, une telle violation de la contrainte ( $S(t) \geq 0$ ) se produit lorsque la valeur minimale<sup>12</sup> de la somme des  $N_m$  harmoniques dépasse, en valeur absolue,  $S_{moy\ ref}$ . La deuxième alternative consiste à décaler la somme des harmoniques de leur valeur minimale afin que le signal généré soit positif (cf. équation (3.4)).

$$S(t) = S_0 + S'(t) \tag{3.4}$$

$$\text{avec } S'(t) = \sum_{n=1}^{N_m} A_n \sin(2\pi f_n t + \Phi_n) \quad \text{et} \quad S_0 = -\min_t(S'(t))$$

---

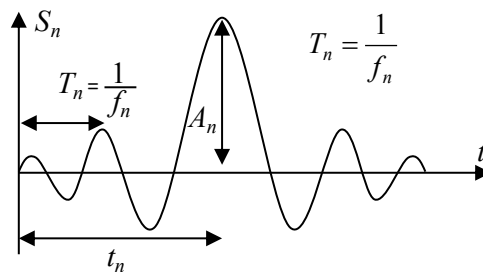
<sup>12</sup> Valeur négative

Suite à ce décalage, la valeur moyenne du signal synthétisé ne correspond plus forcément à la valeur de référence  $S_{moy\ ref}$  que l'on souhaite imposer. Il est donc nécessaire de considérer la valeur moyenne du signal parmi les indicateurs de référence à satisfaire et de l'intégrer à la fonction d'erreur conformément à l'équation (3.1). Cette dernière alternative sera retenue pour le traitement des missions ferroviaires dans la suite de ce travail.

- *Modèle de type sinus cardinal*

Un deuxième type de motif élémentaire est basé sur l'utilisation du sinus cardinal. Ce modèle est inspiré de la théorie de la transformée en ondelettes [MAL 00] qui permet de décomposer un signal en plusieurs types de signaux élémentaires, paramétrés ou non, nommés ondelettes. La forme « sinus cardinal » constitue souvent une ondelette mère du signal à traiter.

Le principe de génération d'un profil ou d'un signal est identique à celui de la méthode de sinus. Il s'agit de sommer  $N_m$  sinus cardinaux caractérisés chacun par trois paramètres : une amplitude  $A_n$  ( $0 \leq A_n \leq S_{max}$ ), un coefficient de fréquence  $f_n$  ( $f_{min} \leq f_n \leq f_{max}$ ) et un coefficient de localisation temporelle  $t_n$  ( $0 \leq t_n \leq \Delta t_{profil}$ ). La Figure 3.3 est une illustration des différents paramètres d'un motif élémentaire. Son expression est également donnée par l'équation (3.5).



**Figure 3.3.** Paramètres d'un sinus cardinal de rang  $n$  :  $A_n$ ,  $f_n$  et  $t_n$

$$S_n(t) = \begin{cases} A_n \times \frac{\sin(2\pi f_n(t-t_n))}{f_n(t-t_n)} & \text{si } t \neq t_n \\ A_n & \text{si } t = t_n \end{cases} \quad (3.5)$$

L'expression du signal généré  $S(t)$ , obtenu à partir de la somme de  $N_m$  sinus cardinaux est :

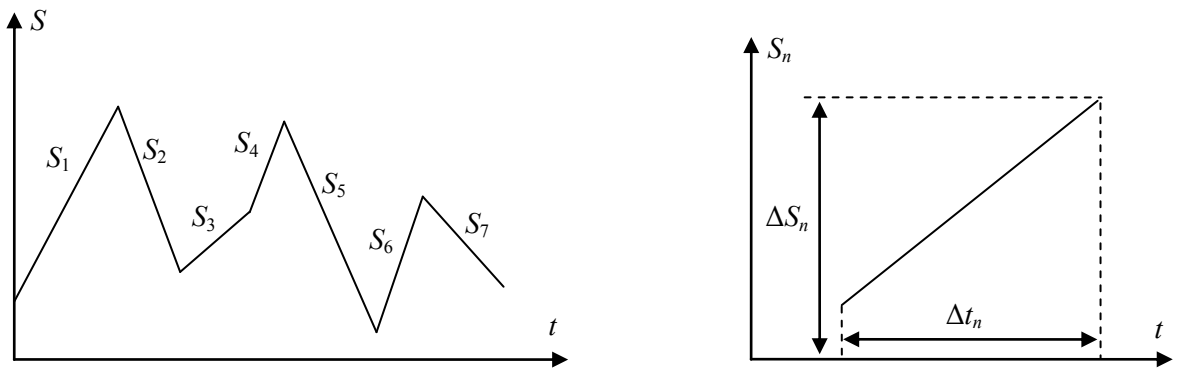
$$S(t) = \sum_{n=1}^{N_m} S_n(t) \quad (3.6)$$



Dans le cas où le signal de mission ou de variable environnementale à synthétiser est défini positif (cas des missions ferroviaires par exemple), nous procédons au décalage du signal  $S(t)$  de sa valeur minimale, si cette dernière est inférieure à zéro, comme nous l'avons décrit pour la méthode de sinus (cf. équation (3.4)).

- *Modèle de type segment*

Le modèle de segment consiste à représenter le profil d'une variable environnementale par une concaténation ou une agrégation de  $N_m$  segments (cf. Figure 3.4.a). Au niveau du processus de synthèse de la Figure 3.1, un motif élémentaire est alors un segment caractérisé par son amplitude  $\Delta S_n$  et sa durée  $\Delta t_n$  comme le montre la Figure 3.4.b. La dimension du problème d'optimisation vaut donc  $2 \times N_m$ . Nous soulignons que ce paramétrage n'était pas évident et a résulté de différents tests exposés à l'Annexe E<sup>13</sup>. Le modèle constitue donc une représentation échantillonnée du profil (avec un échantillonnage à pas variable) associée à une interpolation d'ordre 1. On notera que cette approche pourrait être généralisée à d'autres fonctions interpolation (quadratique cubique, ou de type B-spline par exemple) mais avec un coût de complexité nettement supérieur.



(a) Décomposition d'un profil en segments

(b) Paramètres d'un motif:  $\Delta S_n$  et  $\Delta t_n$

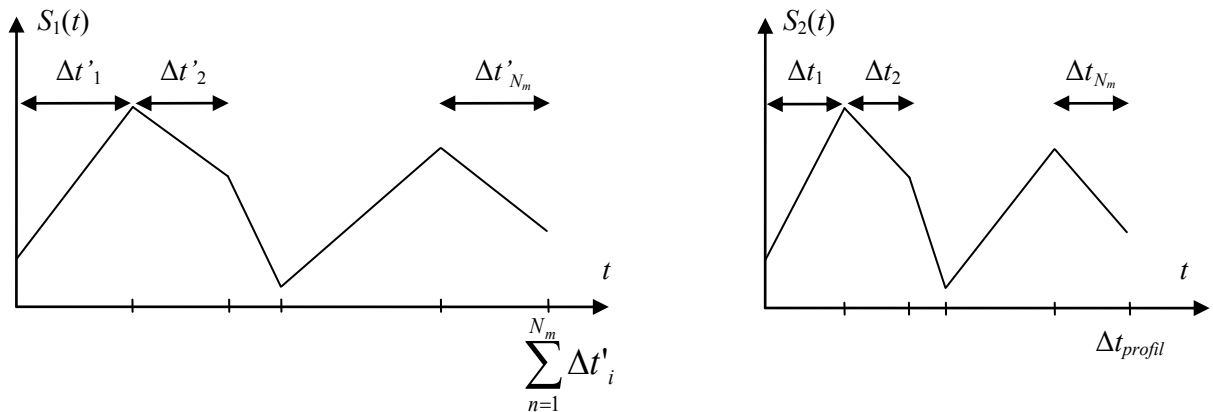
**Figure 3.4.** Modélisation du profil en segments élémentaires

L'amplitude  $\Delta S_n$  d'un segment de rang  $n$  varie entre les deux valeurs extrêmes  $S_{\min ref}$  et  $S_{\max ref}$  à définir en fonction des caractéristiques de l'ensemble des signaux à représenter (missions ou variables environnementales). Les paramètres relatifs aux durées  $\Delta t_n$  de chaque motif élémentaire, varient dans l'intervalle  $]0, \Delta t_{profil}]$ . Une première formulation du problème d'optimisation consiste à introduire une contrainte d'égalité sur les paramètres  $\Delta t_n$  afin que

<sup>13</sup> Dans cette annexe, nous étudions particulièrement l'influence du paramétrage de l'amplitude  $\Delta S_n$  d'un segment vis-à-vis de la résolution du problème inverse.

leur somme soit égale à la durée désirée du signal ( $\Delta t_{profil}$ ). Toutefois, cette contrainte est très difficile à respecter en raison du grand nombre de segments nécessaires à la synthèse du signal. Nous proposons donc, comme solution, une procédure de « réparation » de ces paramètres qui consiste à réaliser une mise à l'échelle selon l'axe temporel pour garantir la durée  $\Delta t_{profil}$ . Soit  $\Delta t'_i$  et  $\Delta t_i$  ( $i = \{1, \dots, N_m\}$ ) les durées de  $N_m$  motifs élémentaires respectivement avant et après la mise à l'échelle (cf. Figure 3.5). Le passage du signal  $S_1(t)$  de la Figure 3.5.a avant la mise à l'échelle au signal  $S_2(t)$  de la Figure 3.5.b après la mise à l'échelle est assuré par la transformation suivante :

$$\Delta t_i = \frac{\Delta t'_i}{\sum_{n=1}^{N_m} \Delta t'_i} \times \Delta t_{profil} \tag{3.7}$$



(a) Signal avant la mise à l'échelle

(b) Signal après la mise à l'échelle

**Figure 3.5.** Mise à l'échelle du signal selon l'axe temporel

### 3.2.2 Codage des chromosomes

Le choix du nombre de motifs  $N_m$  à considérer dans le processus de synthèse n'est pas évident. Prendre un nombre de motifs trop petit ne permettra pas d'introduire suffisamment de variabilité dans le signal généré et donc de satisfaire les indicateurs de référence liés aux caractéristiques du profil recherché. A l'inverse, un nombre de motifs trop important augmentera considérablement la complexité du problème inverse, rendant sa résolution difficile. Fixer le nombre de motifs *a priori*, n'est donc pas évident. Cette première approche a toutefois été envisagée en premier lieu en raison de sa simplicité. Puis, nous avons développé une deuxième approche permettant d'introduire le nombre de motifs comme paramètre dans le processus de synthèse de la Figure 3.1. Ces deux approches se distinguent

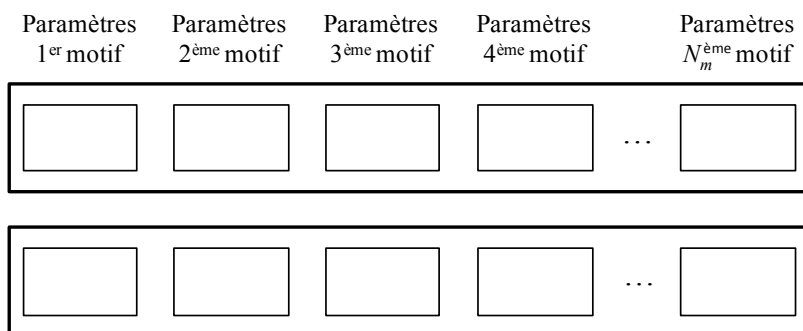
tout particulièrement au niveau de la représentation des solutions et du codage correspondant des chromosomes dans l’algorithme évolutionnaire.

- *Codage du chromosome avec nombre de motifs fixe*

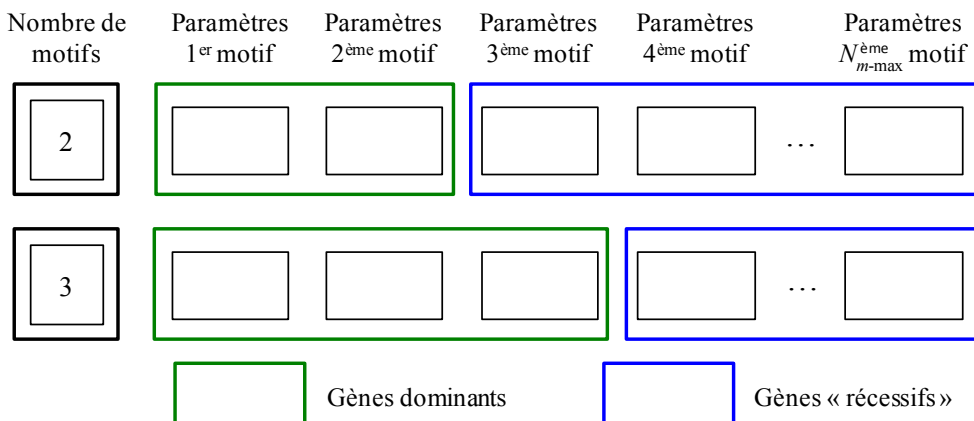
Il s’agit de l’approche la plus classique dans laquelle le nombre de motifs  $N_m$  est fixé *a priori*, de façon identique pour tous les individus de la population. En conséquence, le chromosome des individus est constitué par les paramètres relatifs aux  $N_m$  motifs. (cf. [Figure 3.6.a](#)). La taille du chromosome dépend directement du nombre de motifs  $N_m$  et du nombre de paramètres par motifs.

- *Codage du chromosome avec nombre de motifs variable*

Cette deuxième approche est identique à celle utilisée dans le [chapitre 2](#) pour la classification par algorithme évolutionnaire avec un nombre de classes variable. Elle consiste à ajouter au chromosome de chaque individu un *gène* supplémentaire représentatif du nombre de motifs. Ce nombre peut varier entre 1 et un nombre maximal de motifs  $N_{m-max}$  fixé au préalable.



(a) *Codage du chromosome avec un nombre  $N_m$  de motifs fixe*



(b) *Codage du chromosome avec un nombre de motifs variable ( $N_m \in [1, N_{m-max}]$ )*

**Figure 3.6.** *Stratégies de codage du chromosome*

La taille du chromosome est identique pour tous les individus de la population mais elle dépend cette fois-ci de ce nombre maximal de motifs et du nombre de paramètres par motifs (cf. [Figure 3.6.b](#)). Néanmoins, seule la partie du chromosome dont la taille est déterminée en fonction du *gène* supplémentaire représentatif du nombre de motifs est considérée dans le décodage des individus. Les gènes restants ne sont pas exprimés et sont considérés comme « récessifs ». Ils ne sont pas utilisés dans la procédure de décodage des individus mais ils participent aux opérations de croisement et de mutation.

### 3.3 Génération d'un profil en fonction du type de motif

Dans ce paragraphe, nous mettons en évidence la complexité du problème de synthèse de profils et nous analysons la cohérence entre choix du motif élémentaire et la forme du profil réel. Nous signalons au lecteur, que la synthèse du signal représentatif (cf. [Figure 3.1](#)) est relative non pas à l'ensemble du signal réel mais plutôt aux indicateurs de conception considérés. Dans une étape intermédiaire, il nous semble intéressant de vérifier la capacité des motifs proposés et des algorithmes évolutionnaires à retrouver un signal réel. L'application du processus de synthèse décrit au début de ce chapitre aura lieu un peu plus tard. Notre propos, dans cette section, consiste alors à dire que « si nos motifs élémentaires sont capables de décrire un profil réel d'une mission ou d'une variable environnementale donnée, alors ils pourront reproduire l'intégralité des indicateurs de conception considérés ».

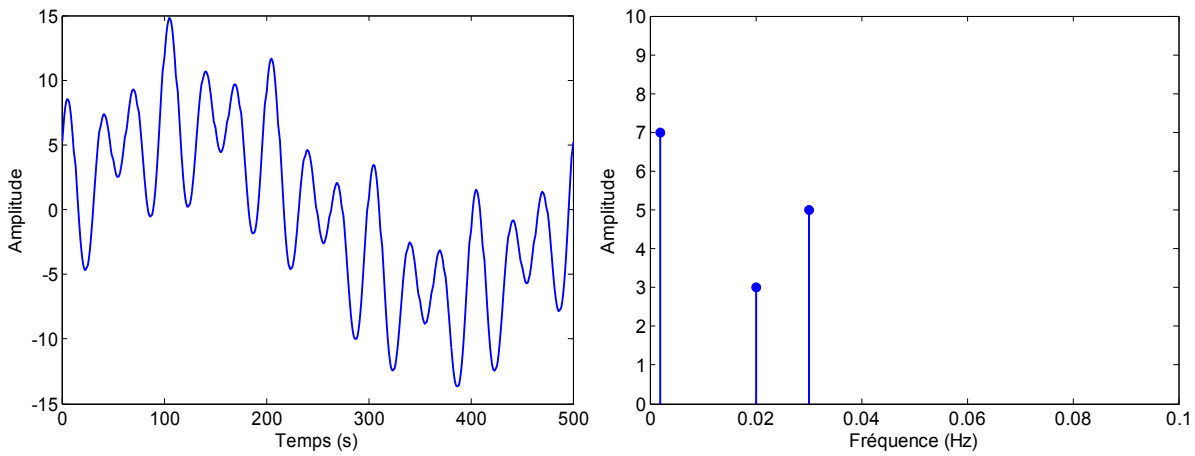
#### 3.3.1 Cas de profils « académiques »

Nous nous basons dans ce paragraphe sur des exemples de signaux « académiques » simples afin de mettre en évidence le choix adéquat des motifs élémentaires vis-à-vis de la forme et les caractéristiques du profil réel. Nous commençons par une phase d'analyse de trois signaux simples puis nous vérifions, par l'optimisation, la capacité des motifs élémentaires proposés précédemment à reproduire ces signaux.

- *Analyse*

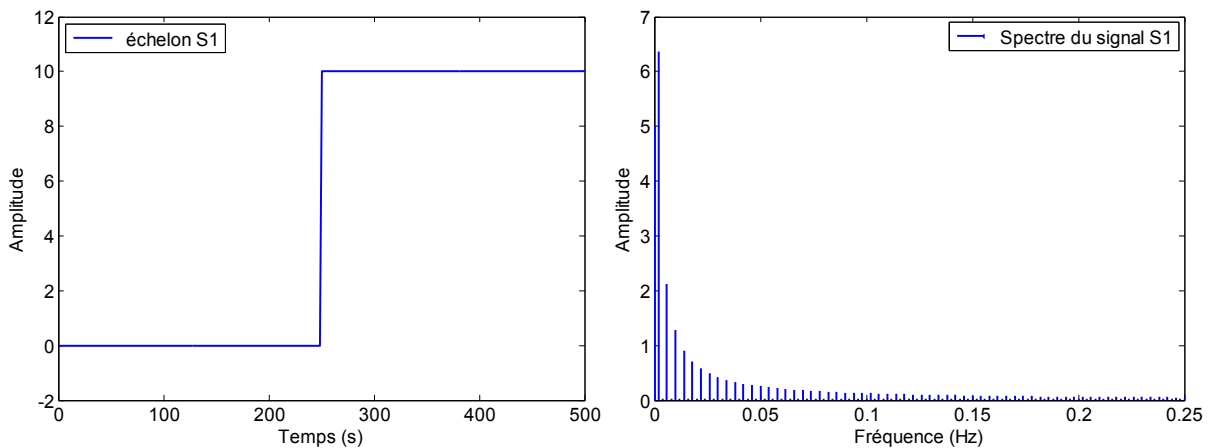
Le profil « réel » de la [Figure 3.7](#) de durée 500 s est issu de la somme de trois signaux sinusoïdaux de fréquences multiples du fondamental  $f_1$  ( $f_1 = 1/500$ ,  $f_{10} = 10 \times f_1$  et  $f_{15} = 15 \times f_1$ ). Il est évident que, dans un tel cas de figure, le motif élémentaire en sinus est le plus adéquat

pour la génération de ce même signal. Les motifs de type segment ou sinus cardinal restent relativement plus compliqués pour la reconstitution de ce profil réel particulier.

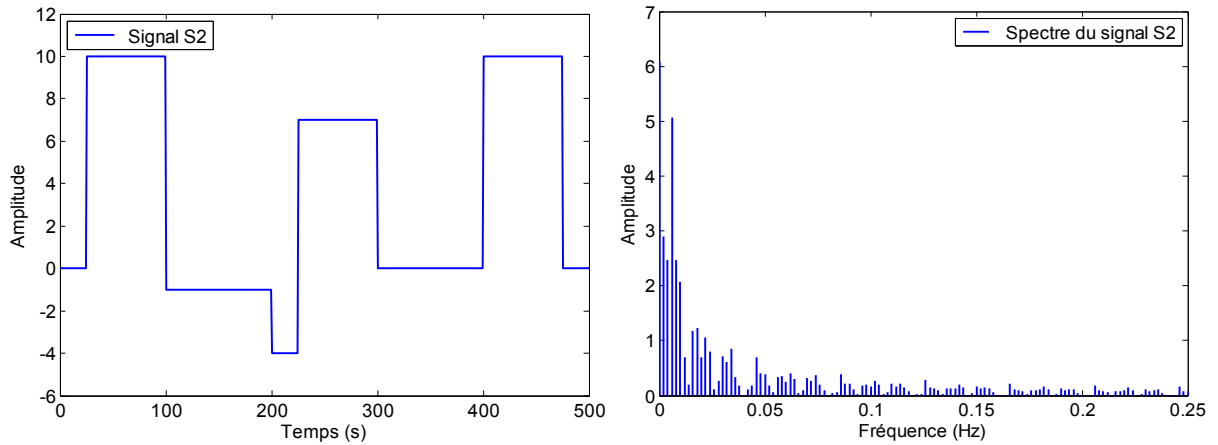


**Figure 3.7.** Exemple d'un profil issu de l'interférence de trois signaux sinusoïdaux

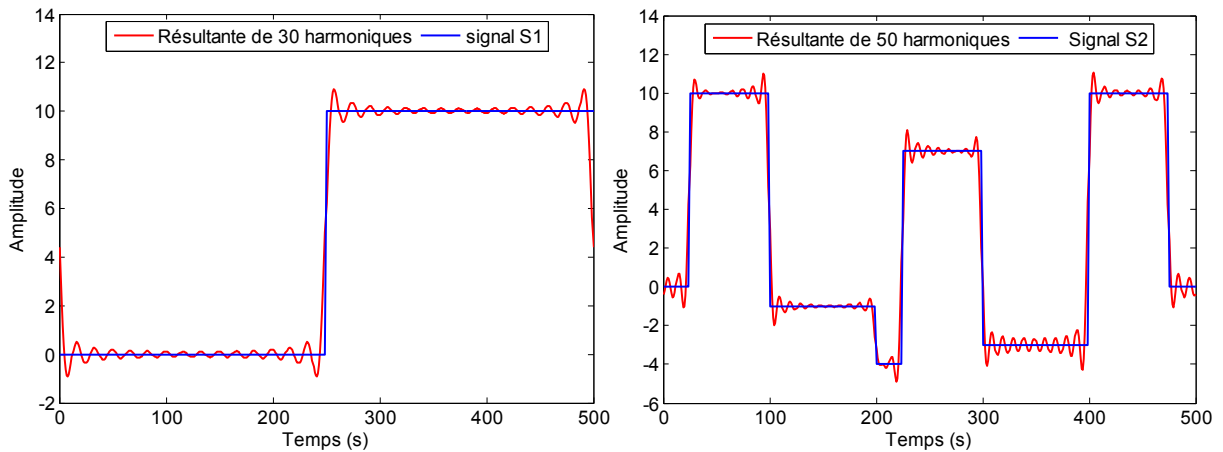
Par contre, le modèle en sinus trouve rapidement ses limites avec les profils  $S_1$  et  $S_2$  de la Figure 3.8 et de la Figure 3.9. Il s'agit respectivement d'un échelon et d'une somme de signaux rectangulaires. Les spectres fréquentiels de ces deux signaux montrent que le nombre d'harmoniques nécessaire à leur régénération est très important. En effet, la Figure 3.10.a montre qu'avec 30 harmoniques successifs, nous n'arrivons pas à reproduire parfaitement l'échelon  $S_1$ . De même, la Figure 3.10.b montre l'insuffisance de 50 harmoniques pour dupliquer la forme exacte du signal  $S_2$ . Dans ce cas, l'utilisation du sinus comme motif élémentaire semble moins appropriée et ne fait qu'amplifier la dimension du problème d'optimisation et par la suite sa complexité. C'est alors le modèle de segment qui semble le plus adéquat pour la représentation de ces deux signaux avec un faible coût d'optimisation (nombre réduit de paramètres).



**Figure 3.8.** Exemple d'un signal carré



**Figure 3.9.** Exemple d'un profil à base de signaux rectangulaires



(a) Régénération de  $S_1$  à partir de 30 harmoniques

(b) Régénération de  $S_2$  à partir de 50 harmoniques

**Figure 3.10.** Régénération des signaux  $S_1$  et  $S_2$  à partir de 30 ou 50 harmoniques

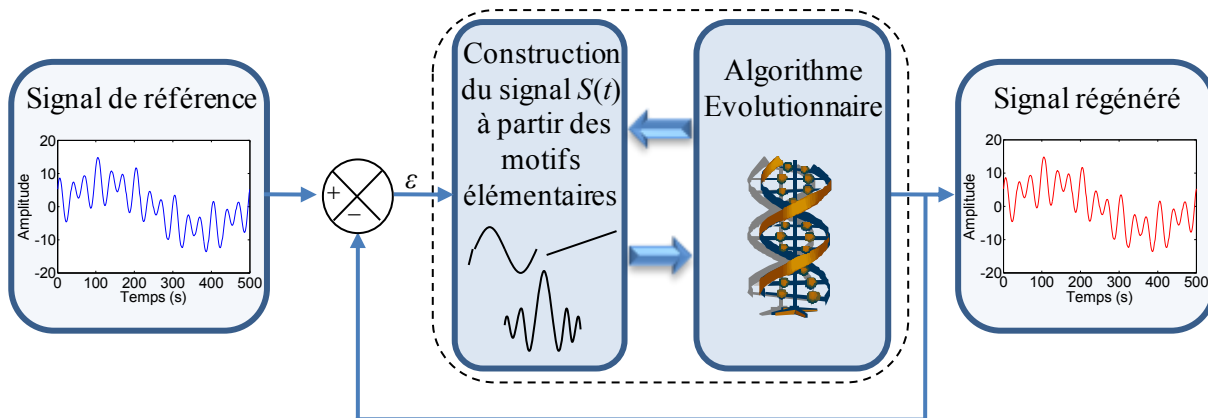
- *Vérification par l'optimisation*

Dans cette partie, nous procédons à la régénération des signaux académiques précédents à partir des motifs élémentaires (sinus, sinus cardinal et segment). Nous soulignons que dans ce cas, la synthèse du signal ne se fait pas à partir des indicateurs de conception mais plutôt en minimisant l'erreur quadratique moyenne ( $\varepsilon$ ) entre le signal académique de référence ( $S_{ref}$ ) et le signal généré ( $S_{gén}$ ) définie par l'expression suivante :

$$\varepsilon = \frac{1}{\sqrt{N}} \sqrt{\sum_{i=1}^N (S_{ref}(t) - S_{gén}(t))^2} \quad (3.8)$$

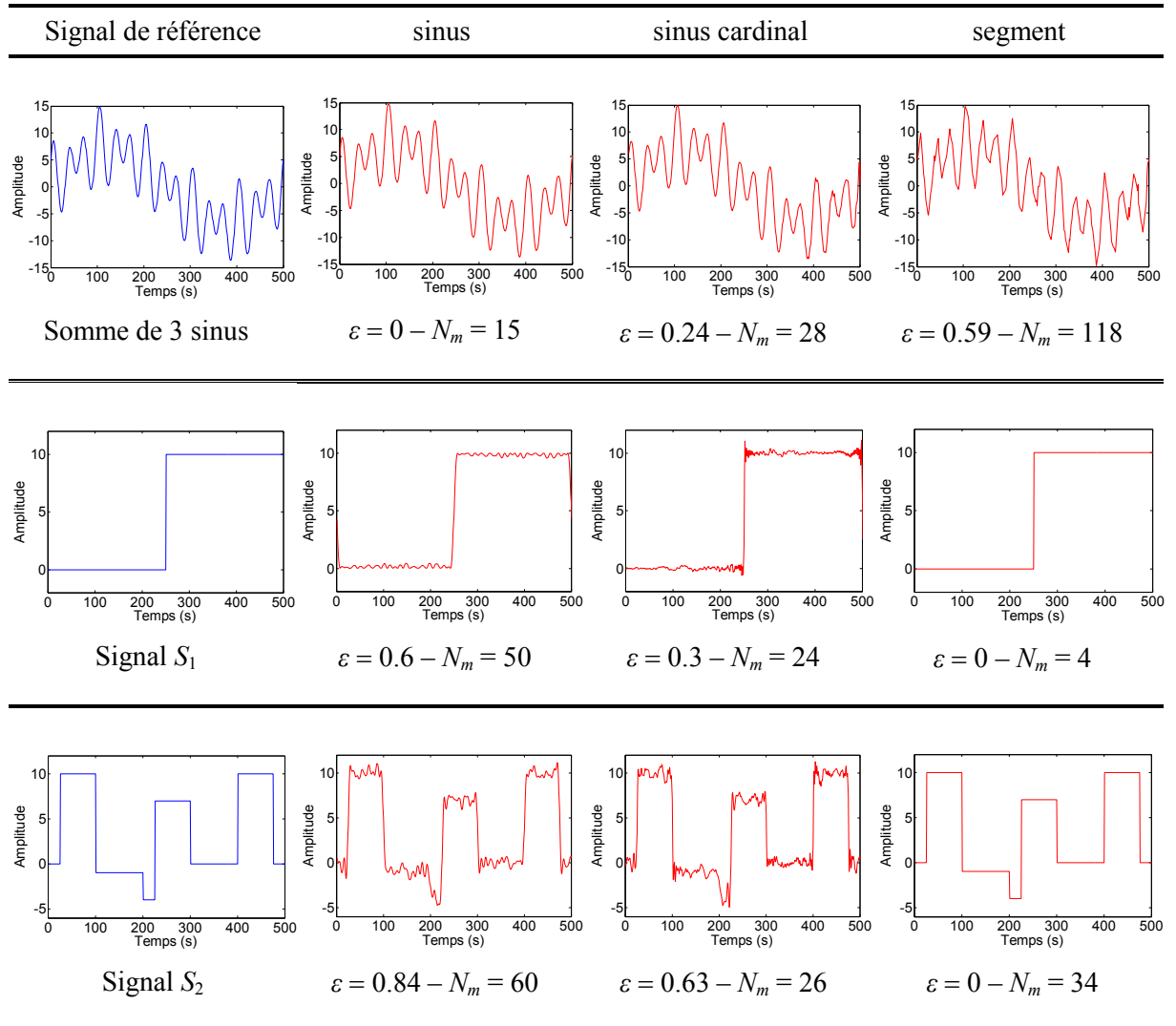
où  $N$  désigne le nombre de points échantillonnant le signal  $S_{ref}$ . Pour les trois signaux étudiés, le pas d'échantillonnage vaut 1 s et  $N$  vaut 501 points, ce qui correspond à une durée de 500 s.

Le processus de régénération d'un signal  $S_{ref}$  est donné par la [Figure 3.11](#). Nous choisissons d'utiliser un algorithme évolutionnaire avec la méthode d'éclaircissement comme technique de nichage (CLR) et avec un nombre de motifs variable. Le nombre de motifs maximal pour chaque configuration testée est choisi de sorte que nous disposons d'une même taille de problèmes d'optimisation (301 paramètres). Il en résulte alors, pour chacun des trois signaux à régénérer, 150 motifs respectivement avec le sinus et le segment ( $2 \times 150 + 1$ ) et 100 motifs avec le sinus cardinal ( $3 \times 100 + 1$ ). La taille de la population est fixée à 100 individus et le nombre de générations à 10 000. Les résultats de synthèse de chaque signal avec les trois types de motifs élémentaires sont exposés au [Tableau 3.1](#). Nous donnons également l'erreur quadratique moyenne ( $\varepsilon$ ) et le nombre de motifs ( $N_m$ ) utilisés.



**Figure 3.11.** Processus de régénération du signal réel de référence  $S_{ref}$

Les résultats du [Tableau 3.1](#) montrent d'une part la bonne convergence de l'algorithme d'optimisation et d'autre part l'importance du choix adéquat du motif élémentaire dans la régénération d'un signal réel. En effet, en considérant le sinus comme motif élémentaire, nous arrivons à reproduire parfaitement le signal sinusoïdal de la [Figure 3.7](#) (somme de 3 sinus) avec seulement 15 harmoniques successifs alors qu'il nous faut 28 motifs avec le sinus cardinal et 118 motifs avec le segment pour se rapprocher au mieux de ce signal. Néanmoins, le motif de type sinus, n'est pas le plus approprié pour les signaux  $S_1$  et  $S_2$ . Ces deux signaux sont facilement générés en considérant le motif de type segment. Effectivement, les signaux  $S_1$  et  $S_2$  sont parfaitement régénérés en concaténant respectivement 4 et 34 segments seulement. Le motif de type sinus cardinal est limité pour une reproduction parfaite des signaux  $S_1$  et  $S_2$  mais a donné lieu à des résultats meilleurs par rapport au motif de type sinus.



**Tableau 3.1.** Différents résultats de synthèse de signaux académiques

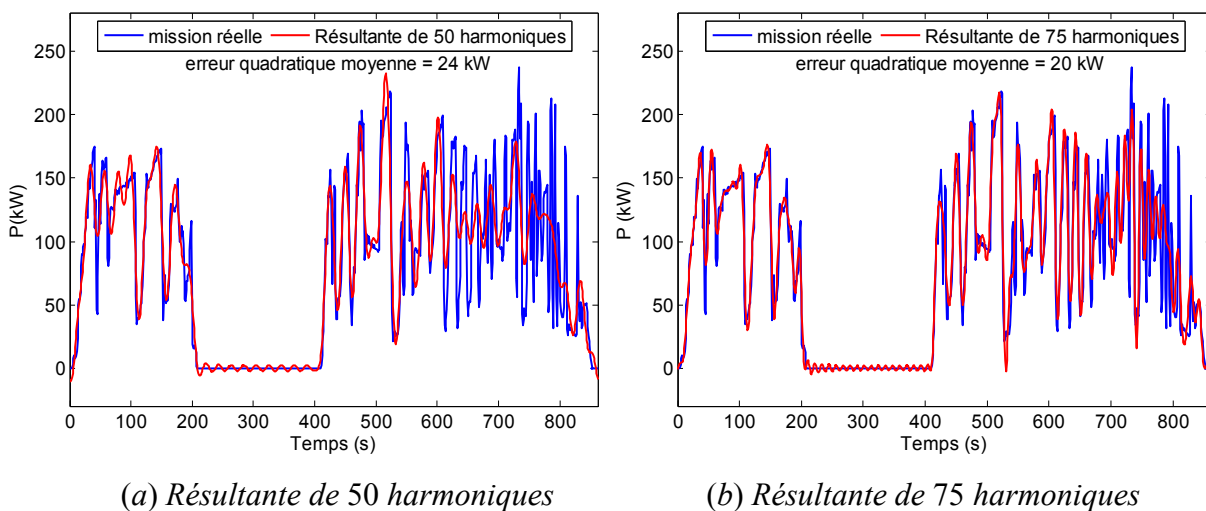
Pour les trois signaux et les trois motifs ainsi testés, nous avons constaté que notre algorithme d'optimisation converge vers un nombre de motifs adéquat vis-à-vis de la forme du signal considéré. En effet, dans le cas du premier signal sinusoïdal de la Figure 3.7 et en considérant le sinus comme motif élémentaire, malgré un nombre de motifs maximal de 150, l'algorithme d'optimisation converge vers un nombre de motifs qui correspond au nombre d'harmoniques minimal nécessaire à la reproduction du signal (15 harmoniques). Dans le cas du signal de type échelon ( $S_1$ ) et le signal rectangulaire ( $S_2$ ) et en considérant le segment comme motif élémentaire, l'algorithme d'optimisation converge vers un nombre légèrement supérieur au nombre de segments minimal nécessaire pour reproduire le signal initial. Cette légère différence se justifie par la forte multimodalité des problèmes inverses considérés. En effet, un palier (niveau constant sur un temps donné) peut être représenté par un ou plusieurs segments de même niveau d'amplitude ( $\Delta P$ ) et de durées différentes.



### 3.3.2 Cas du profil d'une mission ferroviaire

- *Analyse*

En complément de l'analyse précédente, ce paragraphe donne une idée sur le contenu harmonique d'une mission de circulation ferroviaire réelle de la locomotive BB63000 de durée 862 s avec un pas d'échantillonnage de 1 s. Il s'agit d'un exemple de profil très riche en harmoniques, ce qui rend délicate sa reconstruction à partir des motifs sinusoïdaux. En effet, la [Figure 3.12](#) montre que sommer 50 harmoniques, voire 75 harmoniques, reste insuffisant pour une reproduction correcte de la mission réelle. Dans le premier cas ([Figure 3.12.a](#)), nous obtenons une erreur quadratique moyenne de l'ordre de 24 kW entre la mission réelle de référence et la mission reconstruite à partir de la somme des 50 harmoniques. Le cas avec 75 harmoniques ([Figure 3.12.b](#)), présente une erreur plus faible de l'ordre de 20 kW. Nous revenons dans la section suivante sur la régénération de ce profil réel à partir des différents motifs élémentaires proposés.



**Figure 3.12.** Reconstruction d'un signal par la somme des premiers harmoniques de la mission réelle

À ce niveau, nous profitons de cette analyse fréquentielle pour avoir une idée sur le nombre d'harmoniques à considérer afin que le signal obtenu<sup>14</sup> respecte les indicateurs de conception de cette mission ferroviaire définis dans le [chapitre 1](#) ( $P_{\max}$ ,  $P_{\text{moy}}$ ,  $E_u$ ,  $N_{\text{cyc}}$  et  $I_{\text{stat}}$ ). Nous avons constaté qu'il faut au moins 100 harmoniques successifs pour respecter l'ensemble des indicateurs de référence avec une erreur inférieure ou égale à 10%. Autrement dit, si nous désirons obtenir ce même résultat par le processus de synthèse de profil de la

<sup>14</sup> Le signal est obtenu par la somme des harmoniques et ne pas par le processus de synthèse.

Figure 3.1 en considérant le sinus comme motif élémentaire, nous sommes alors face à un problème d'optimisation d'au moins 200 paramètres ( $2 \times 100$ ). Il s'agit d'un problème à coût de complexité très élevé, non seulement par son nombre important de paramètres mais aussi par sa multimodalité. Nous verrons plus tard si les autres motifs élémentaires offrent des solutions plus efficaces pour la synthèse de profil respectant les indicateurs de référence.

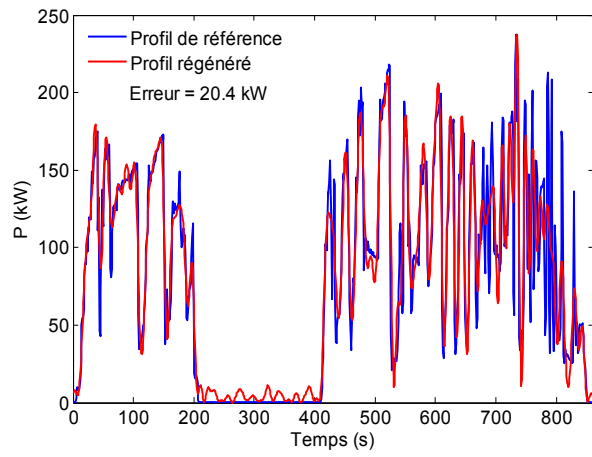
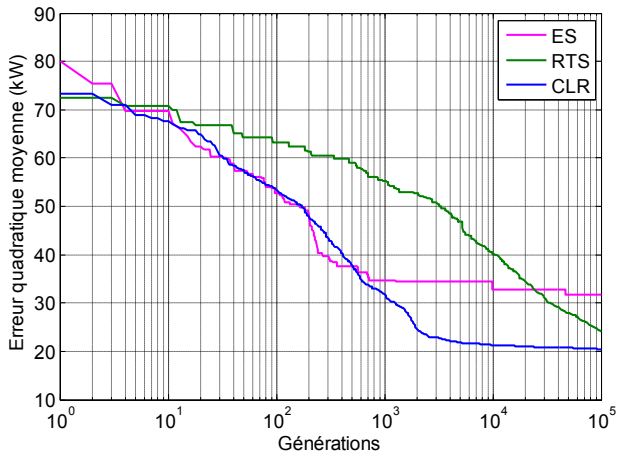
- *Vérification de la capacité des motifs élémentaires à la régénération d'une mission réelle*

Dans cette partie, nous évaluons la capacité des motifs élémentaires proposés (sinus, sinus cardinal et segment) à régénérer la mission de circulation réelle de la locomotive BB63000 analysée dans le paragraphe précédent. Il s'agit en fait de mettre en évidence la complexité de la résolution d'un problème inverse qui consiste à trouver les paramètres adéquats des motifs élémentaires afin de minimiser l'erreur quadratique moyenne ( $\varepsilon$ ) entre la mission réelle de référence  $P_{ref}$  et la mission fictive générée  $P_{gén}$  conformément à la Figure 3.11 et à l'équation (3.8). Nous examinons également les performances de trois algorithmes évolutionnaires sans et avec technique de nichage : une stratégie d'évolution (ES), la sélection par tournoi restreint (RTS) et la méthode d'éclaircissement (CLR).

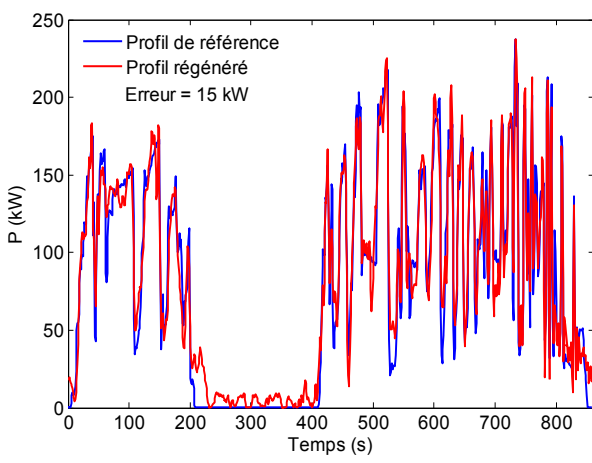
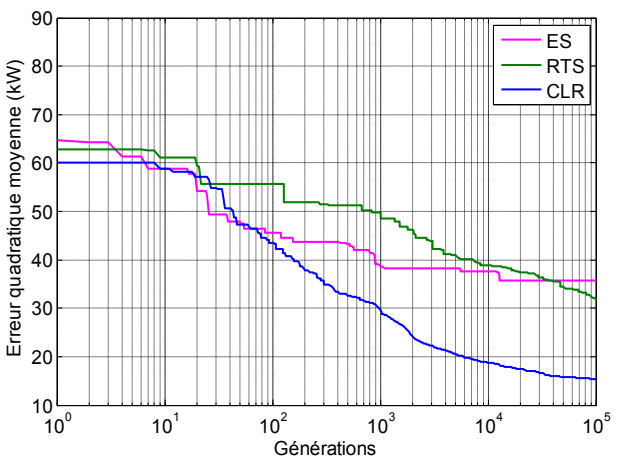
(a) *Résultats avec nombre de motifs fixe :*

Afin de comparer l'efficacité des algorithmes évolutionnaires, nous choisissons le même nombre de paramètres d'optimisation pour le processus de synthèse avec les trois motifs élémentaires. Ce nombre est fixé à 150 paramètres. Il en résulte 75 motifs<sup>15</sup> respectivement pour les motifs de type segment et de type sinus et 50 motifs pour le sinus cardinal. La taille de la population de chaque algorithme évolutionnaire est fixée à 100 et le nombre de générations à 100 000. Pour chaque combinaison (algorithme évolutionnaire et motif élémentaire), cinq tests indépendants sont réalisés [JAA 10b]. Les meilleurs résultats sont représentés par la Figure 3.13. Les évolutions de l'erreur quadratique moyenne ( $\varepsilon$ ) en fonction du nombre de générations pour chaque modèle du signal et avec les trois algorithmes évolutionnaires sont données par la Figure 3.13.a. Nous constatons que le CLR est le meilleur algorithme en termes de convergence. A noter que, même avec le meilleur algorithme, l'erreur quadratique moyenne ( $\varepsilon$ ) reste relativement importante (supérieure à 10 kW, soit ~5% de la puissance maximale du profil) à l'issue de l'optimisation. Une comparaison entre le profil de référence  $P_{ref}$  et le profil généré  $P_{gén}$  par chaque modèle de signal avec le meilleur algorithme évolutionnaire (CLR) est donnée par la Figure 3.13.b.

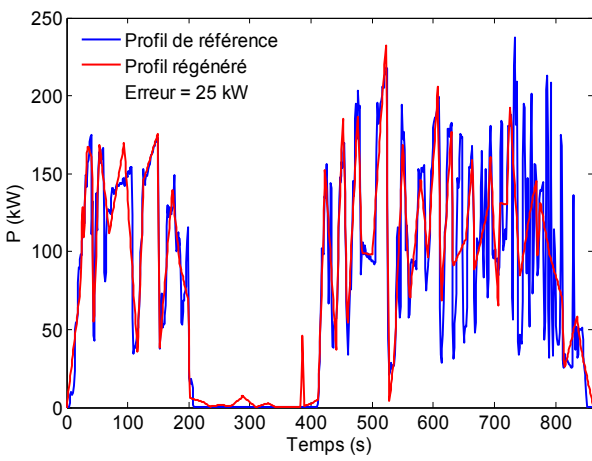
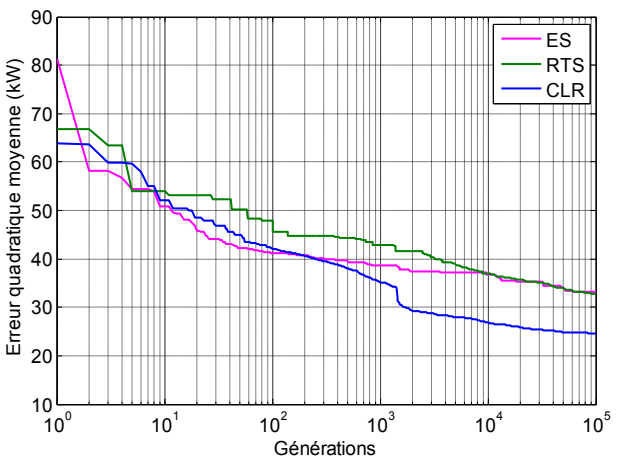
<sup>15</sup> Nous rappelons que le sinus ou le segment sont paramétrés chacun par 2 paramètres.



*Motif de type sinus « Fourier »*



*Motif de type sinus cardinal*



*Motif de type segment*

(a)

(b)

**Figure 3.13.** (a) Evolution de l'erreur moyenne  $\varepsilon$  en fonction du nombre de générations  
 (b) Comparaison entre le profil de référence et le profil généré avec le CLR

Nous constatons que le motif élémentaire de type sinus cardinal est, dans ce cas d'application, le mieux adapté à la régénération du profil de référence. En effet, ce dernier permet d'obtenir une erreur de 15 kW, plus faible que l'erreur quadratique moyenne (20 kW) obtenue par la somme de 75 harmoniques déterminés à partir de l'analyse de Fourier comme le montre la [Figure 3.12.b](#) de la sous-section précédente. Moins satisfaisant par rapport au sinus cardinal, le motif élémentaire en sinus présente une erreur de l'ordre de 20 kW, comparable au résultat de la [Figure 3.12.b](#). En ce qui concerne le modèle de segment, il semble que le résultat obtenu par cette méthode ( $\varepsilon = 25$  kW) pourra être nettement meilleur avec un nombre de motifs plus important.

Nous donnons au niveau du [Tableau 3.2](#) une idée sur le temps de simulation nécessaire pour effectuer 100 000 générations pour chaque modèle de synthèse avec les trois algorithmes évolutionnaires. Ces temps de calcul sont évalués sur un PC standard (*Core Duo* 2 GHz). Le meilleur résultat obtenu par le modèle de sinus cardinal couplé au CLR nécessite un temps de calcul qui s'élève jusqu'à 2 jours.

	sinus $N_m = 75$	sinus cardinal $N_m = 50$	segment $N_m = 75$
ES	2.3 jours	1.5 jours	2 heures
RTS	2.5 jours	1.7 jours	5 heures
CLR	3 jours	2 jours	0.5 jours

**Tableau 3.2.** Temps de simulation avec un PC standard (*Core Duo* 2 GHz)

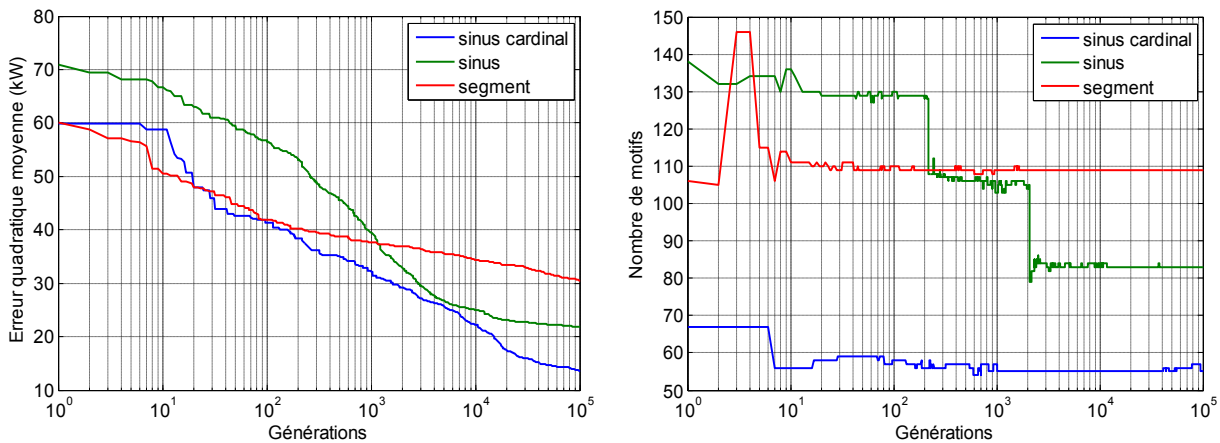
(b) Résultats avec nombre de motifs variable :

Au niveau de cette partie, nous examinons la capacité des motifs élémentaires à reproduire un profil réel avec un nombre de motifs variable. Nous reprenons le même profil de mission ferroviaire étudié ci-dessus avec le CLR comme algorithme évolutionnaire. Les résultats de simulations pour les trois motifs élémentaires sont donnés par la [Figure 3.14](#). Nous représentons respectivement l'évolution de l'erreur quadratique moyenne et la variation du nombre de motifs au cours de 100 000 générations.

Les résultats ainsi obtenus montrent que l'algorithme évolutionnaire a tendance à converger vers un nombre de motifs plus important que précédemment. En effet, avec le modèle de sinus, le nombre de motifs passe de 75 à 83 sauf que nous signalons une légère dégradation de l'erreur  $\varepsilon$  qui passe de 20 kW à 22 kW. Il en est de même pour la méthode des

segments dont le nombre de motifs croît de 75 à 109, alors que l'erreur  $\varepsilon$  passe de 25 kW à 30.5 kW. En ce qui concerne le modèle de sinus cardinal, nous remarquons une diminution de l'erreur de 15 kW à 14 kW avec une faible croissance du nombre de motifs de 50 à 53. Ces résultats s'expliquent par la complexité du problème d'optimisation à grand nombre de paramètres et dont la multimodalité est d'autant plus accentuée avec un nombre de motifs variable. En conséquence, la convergence de l'algorithme évolutionnaire devient beaucoup plus lente avec de nombreux motifs et il est absolument nécessaire d'augmenter le nombre de générations. Nous signalons qu'avec le modèle de sinus cardinal, les 100 000 générations requièrent environ 5 jours de calcul [JAA 10b].

De plus, il est intéressant de constater sur la [Figure 3.14.b](#), la stabilisation du nombre de motifs élémentaires qui devient quasi constant au-delà de quelques milliers de générations.



(a) Evolution de l'erreur

(b) Evolution du nombre de motifs

**Figure 3.14.** Régénération de la mission de référence avec nombre de motifs variable

Dans l'objectif de valider les constatations précédentes, nous avons testé le modèle de sinus cardinal couplé au CLR sur un million de générations. Ce dernier cas correspond à 50 jours de calcul environ. Nous avons alors constaté une amélioration de l'erreur quadratique moyenne  $\varepsilon$  qui diminue de 14 à 11 kW avec 61 motifs.

A travers l'exemple de mission ferroviaire de la locomotive BB63000, nous avons constaté que le choix des motifs élémentaires n'est pas évident *a priori*. En outre, compte tenu des formes compliquées de certains signaux réels, ces motifs demeurent limités dans la reproduction de ces signaux avec une haute précision. En revanche, les limites évoquées ne présentent pas un inconvénient majeur pour la simplification de profils (compactage d'un profil ou synthèse d'un profil représentatif d'une classe), approche pour laquelle la

vérification des indicateurs de conception est beaucoup plus importante que la forme du profil réel. Nous verrons par la suite que les modèles proposés sont efficaces dans un tel contexte.

### 3.4 Capacité des motifs élémentaires à synthétiser un profil fictif à caractéristiques données

Dans cette section, nous examinons à travers un exemple de mission de circulation réelle de la locomotive BB63000 la capacité des motifs élémentaires proposés à synthétiser un profil à caractéristiques données. Il s'agit de générer un profil fictif dont les indicateurs de conception correspondent bien aux indicateurs de référence. Dans un premier temps, afin d'évaluer les performances des trois algorithmes évolutionnaires (ES, RTS et CLR) face à ce genre de problème, nous ne considérons que des configurations à même nombre de paramètres quel que soit le type de motif utilisé. Autrement dit, nous fixons le nombre de motifs pour chaque modèle élémentaire de manière à avoir la même dimension pour le problème inverse. Dans un deuxième temps et après avoir choisi le meilleur algorithme d'optimisation, nous reprenons le même problème de synthèse avec un nombre de motifs variable. Nous faisons l'hypothèse que le profil fictif et représentatif de la mission en question possède la même durée que la mission initiale réelle : la mission n'est donc pas « compactée » ici. Nous verrons à la fin de ce chapitre que cette approche permet aussi de rendre les profils de missions et de variables environnementales plus compacts.

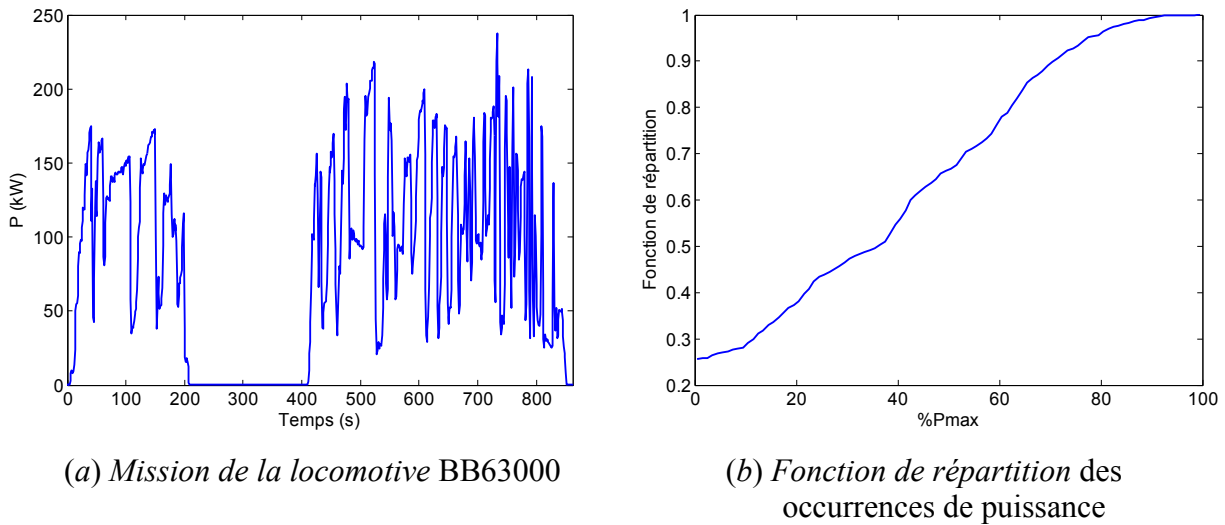
La mission réelle de la locomotive BB63000, de durée  $\Delta t_{profil} = 862$  s, dont nous voulons reproduire certaines caractéristiques est donnée par la [Figure 3.15.a](#). Sa fonction de répartition traduisant l'indicateur statistique de référence, est représentée par la [Figure 3.15.b](#). Les valeurs des autres indicateurs à considérer dans le contexte de la conception d'une locomotive hybride sont résumées dans le [Tableau 3.3](#). Ces valeurs seront prises par la suite comme indicateurs de référence dans le processus de synthèse décrit au début de ce chapitre par la [Figure 3.1](#).

Le critère d'erreur à minimiser dans cet exemple de synthèse de mission est le suivant :

$$\mathcal{E}_{obj} = \left( \frac{P_{\max} - P_{\max ref}}{P_{\max ref}} \right)^2 + \left( \frac{P_{\text{moy}} - P_{\text{moyref}}}{P_{\text{moyref}}} \right)^2 + \left( \frac{E_u - E_{u ref}}{E_{u ref}} \right)^2 + \left( \frac{N_{cyc} - N_{cyc ref}}{N_{cyc ref}} \right)^2 + \mathcal{E}_{stat} \quad (3.9)$$

où  $\mathcal{E}_{stat}$  désigne l'erreur quadratique moyenne entre la fonction de répartition de référence et celle de la mission fictive générée évaluée sur 100 intervalles égaux entre 0 et  $P_{\max ref}$ .

$$\mathcal{E}_{stat} = \frac{1}{100} \times \sum_{k=1}^{100} \left( \frac{I_{stat}(k) - I_{stat.ref}(k)}{I_{stat.ref}(k)} \right)^2 \quad (3.10)$$



**Figure 3.15.** Profil de mission de la locomotive BB63000 et sa fonction de répartition

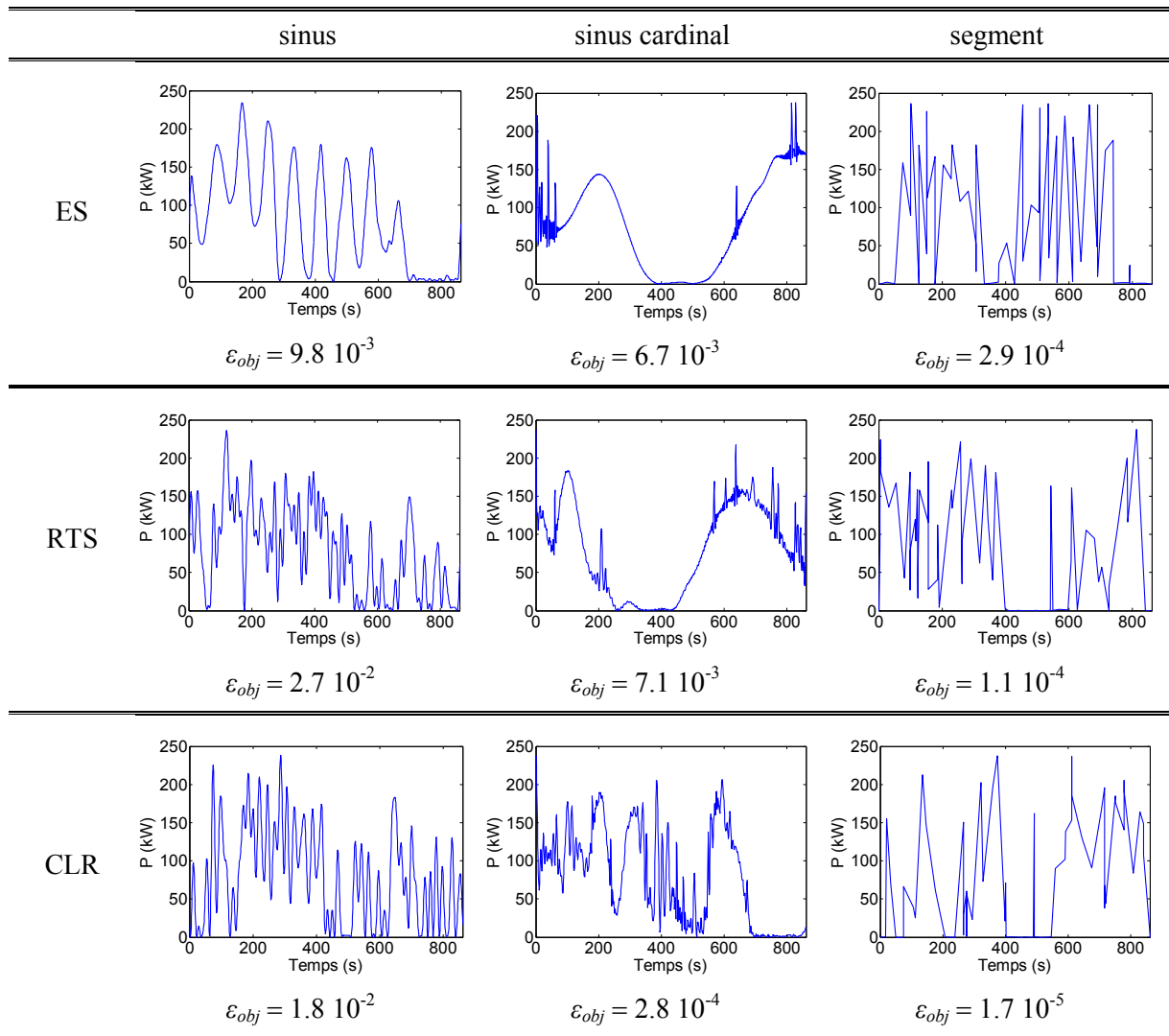
Indicateur	Valeur de référence
Puissance maximale : $P_{max.ref}$ (kW)	237
Puissance moyenne : $P_{moy.ref}$ (kW)	81.3
Energie utile : $E_{u.ref}$ (kWh)	3.97
Nombre de cycles : $N_{cyc.ref}$	0.68

**Tableau 3.3.** Valeurs des indicateurs de référence

### 3.4.1 Résultats avec nombre de motifs fixe et choix de l'algorithme évolutionnaire

Comme nous l'avons signalé précédemment, le nombre de paramètres est fixé à 120 pour toutes les configurations testées permettant de croiser les motifs élémentaires d'une part et les algorithmes évolutionnaires d'autre part. Il en résulte 60 motifs pour les modèles de type sinus et de segment et seulement 40 motifs pour le modèle de sinus cardinal. Tous les algorithmes évolutionnaires (ES, RTS et CLR) sont configurés avec une population de 100 individus. Plusieurs tests de simulation ont été réalisés sur 100 000 générations; les meilleurs résultats des profils générés sont représentés dans le [Tableau 3.4](#). Nous donnons, au [Tableau 3.5](#), le taux d'erreur des indicateurs de chaque profil représentatif obtenu par rapport aux valeurs de référence. Enfin, nous donnons au [Tableau 3.6](#) un ordre de grandeur sur le temps de simulation par configuration testée.

À l'issue de cette étude, nous constatons que les trois méthodes de synthèse sont capables de générer un profil représentatif de la mission réelle de la locomotive BB63000 de même durée et avec un taux d'erreur inférieur à 10% pour chaque indicateur de conception. Toutefois, le [Tableau 3.5](#) montre que les meilleurs résultats sont obtenus avec les motifs de type segment, en particulier lorsque ceux-ci sont couplés au CLR comme algorithme évolutionnaire. Le [Tableau 3.4](#), donnant les meilleurs profils obtenus pour chaque motif et chaque algorithme évolutionnaire, prouve la forte multimodalité du problème inverse. En effet, avec des taux d'erreur pratiquement comparables, nous arrivons à résoudre le problème avec 9 profils dont les formes diffèrent sensiblement. L'utilisation de méthodes de nichage, permettant notamment d'éviter une convergence prématurée pour de tel problème, se justifie ici pleinement.



**Tableau 3.4.** Différents résultats des profils représentatifs de la mission réelle



	sinus			sinus cardinal			Segment		
	ES	RTS	CLR	ES	RTS	CLR	ES	RTS	CLR
$\varepsilon_{P_{\max}}$ (%)	1.5	0.5	0.3	0	0	0	0.5	0	0
$\varepsilon_{P_{\text{moy}}}$ (%)	0.4	4	2.3	2.2	2	1.4	0	0	0.1
$\varepsilon_{E_u}$ (%)	5.2	6.5	5	0.5	1.4	2.4	0.1	0.2	0
$\varepsilon_{N_{\text{cyc}}}$ (%)	3.5	0.4	2.4	1.3	1	0.5	0	0.1	0.1
$\varepsilon_{I_{\text{stat}}}$ (%)	4	8.5	7	4.2	4.3	3	1.2	0.7	0.4

**Tableau 3.5.** Taux d'erreur en % sur les indicateurs pour les diverses configurations

Pour ce qui concerne les temps d'exécution, celui du CLR est légèrement plus important que pour les deux autres algorithmes (RTS et ES). En effet, avec les motifs de type segment, qui donnent la convergence la plus rapide par rapport aux deux autres motifs, ce temps s'élève jusqu'à 0.5 jour pour 100 000 générations effectuées. Il peut atteindre 1.8 jours avec le motif sinus cardinal et 2.3 jours avec le sinus (cf. [Tableau 3.6](#)).

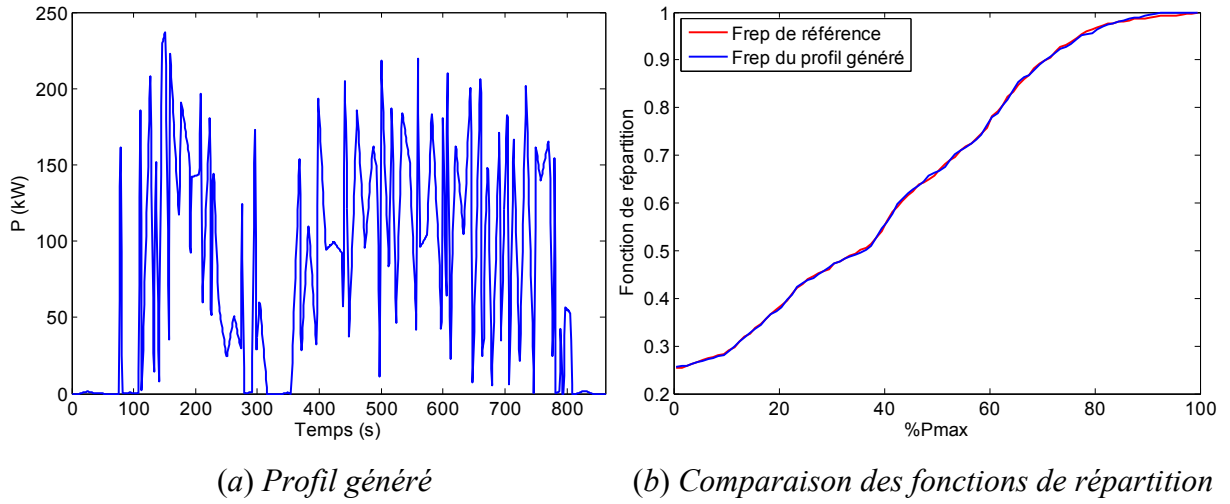
	sinus $N_m = 60$	sinus cardinal $N_m = 40$	segment $N_m = 60$
ES	2 jours	1.5 jours	6 heures
RTS	2.1 jours	1.6 jours	9 heures
CLR	2.3 jours	1.8 jours	0.5 jour

**Tableau 3.6.** Temps de simulation avec un PC standard (Core Duo 2 GHz)

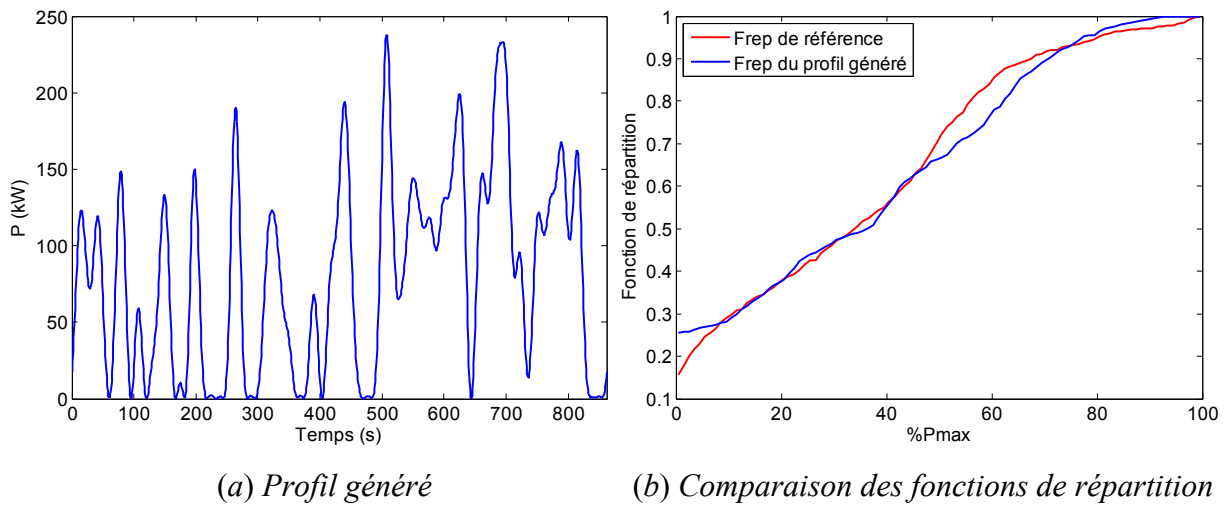
Dans la suite de ce travail, nous ne considérerons que le CLR comme algorithme d'optimisation dans le processus de synthèse de profil représentatif. Avant de confirmer notre choix de motif élémentaire dans le cas où la variable d'environnement est une mission ferroviaire, nous évaluons la performance de la synthèse à partir des motifs élémentaires en considérant un nombre de motifs variable. Ceci fait l'objet du paragraphe suivant.

### 3.4.2 Résultats avec nombre de motifs variable

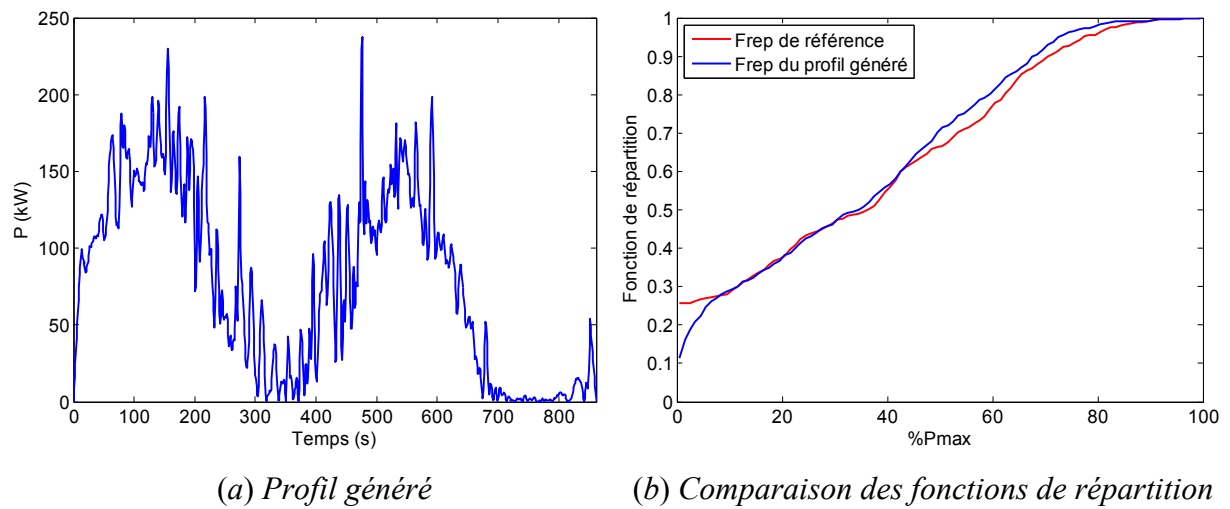
Contrairement à l'étude précédente, nous traitons ici la problématique de synthèse de mission représentative avec un nombre de motifs variable. Nous rappelons le lecteur que ce nombre de motifs  $N_m$  est considéré dans ce cas comme un paramètre d'optimisation. Les trois méthodes de synthèse sont ainsi testées en considérant le CLR comme algorithme d'optimisation. Les résultats de simulation sont donnés par les [Figure 3.16](#), [Figure 3.17](#) et [Figure 3.18](#).



**Figure 3.16.** Résultat du processus de synthèse avec le motif segment



**Figure 3.17.** Résultat du processus de synthèse avec le motif sinus



**Figure 3.18.** Résultat du processus de synthèse avec le motif sinus cardinal

La taille de la population de l'algorithme évolutionnaire CLR est de 100 individus. Nous fixons le nombre de générations à 100 000 afin qu'il soit largement suffisant pour la convergence de l'algorithme. Les caractéristiques des profils obtenus sont résumées dans le [Tableau 3.7](#). Ce tableau montre que le meilleur résultat est donné à partir du motif segment. Bien qu'il s'agisse d'un problème d'optimisation difficile, les indicateurs du profil représentatif obtenu coïncident parfaitement avec les indicateurs de référence. Pour l'indicateur statistique, la [Figure 3.16.b](#) montre la superposition de la fonction de répartition du signal généré avec celle de référence. En ce qui concerne les deux autres motifs élémentaires, les taux d'erreur sur les indicateurs de conception ainsi trouvés sont moins précis que ceux donnés par les segments mais restent aussi acceptables.

	sinus	sinus cardinal	segment	
Nombre de motifs	51	102	123	
Nombre de paramètres	103	307	227	
Erreur (%)	$P_{\max}$	0.3	0	<b>0</b>
	$P_{\text{moy}}$	1.6	1.9	<b>0</b>
	$E_u$	5	0.6	<b>0</b>
	$N_{\text{cyc}}$	2.9	1.8	<b>0</b>
	$I_{\text{stat}}$	4.3	3.8	<b>0.4</b>

**Tableau 3.7.** Caractéristiques des profils obtenus pour les trois configurations testées

Il est intéressant (et même atypique) de constater que, tandis qu'à la [section 3.3](#), le segment s'est révélé moins efficace que les autres motifs pour reproduire la « forme réelle » de la mission traitée, les résultats ci-dessus montrent que ce dernier motif est *a contrario* le plus efficace pour la synthèse d'un profil de mission ferroviaire à « caractéristiques données ». Ainsi, on peut conclure que synthétiser un profil au sens d'une « forme de signal » ou au contraire au sens d'« indicateurs » relatifs à la conception ne correspond donc pas du tout à la même problématique. Le processus de synthèse avec le modèle de segment est d'autant plus précis que l'on utilise le CLR comme algorithme évolutionnaire avec un choix automatique du nombre de motifs.

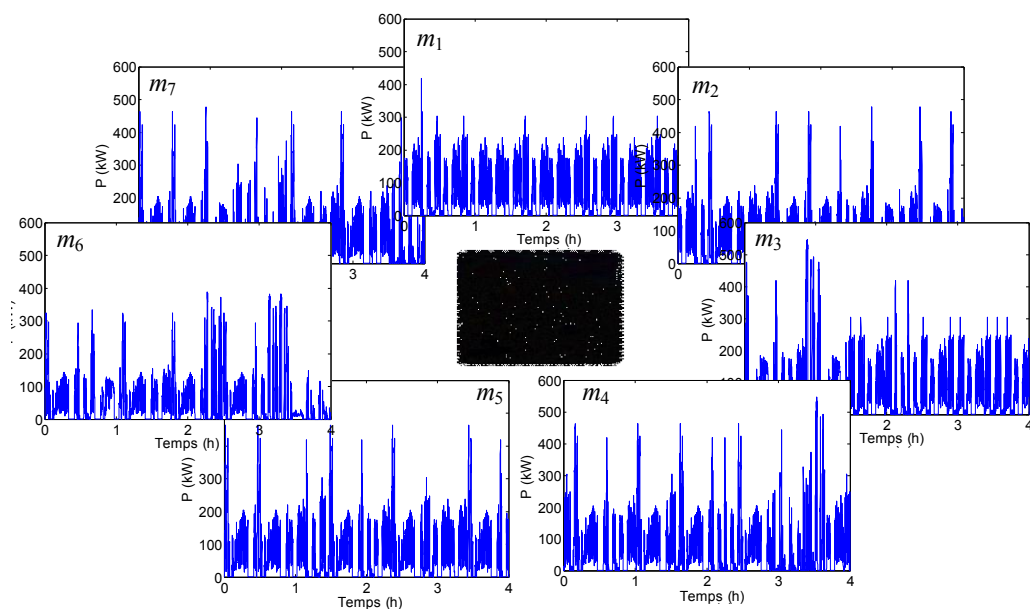
### 3.5 Exemples d'application

Dans cette dernière partie, nous appliquons le processus de synthèse d'un profil fictif représentatif sur deux exemples de missions et variables environnementales. Le premier s'inscrit dans le cadre d'une démarche de conception simultanée par optimisation d'une

locomotive hybride que nous aborderons plus en détails au [chapitre 4](#). Il s'agit de synthétiser une mission représentative d'une famille ou d'une classe de missions ferroviaires réalisées par la locomotive diesel BB63000. Nous montrons comment le processus de synthèse permet de générer une seule mission de durée réduite et spécifique du contenu informationnel (indicateurs) de l'ensemble des missions réelles. Le deuxième exemple traite du compactage de la vitesse réelle du vent pour deux mois de mesures relevées sur le site de « Petit Canal » de la Guadeloupe. Nous comparons notre approche de synthèse de profil représentatif et compact avec une approche équivalente purement statistique.

### 3.5.1 Application à une famille de missions ferroviaires

Nous traitons en premier lieu de la synthèse d'une mission représentative d'une famille de 7 missions ferroviaires confiées à la locomotive hybride PLATHEE (diesel, stockage sur une base BB63000). Dans le projet PLATHEE, il s'agissait de remplacer le groupe diesel de 610 kW de la locomotive BB63000 par un autre de taille plus réduite et de compléter le manque de puissance par des batteries d'accumulateurs de type Ni–Cd. Les 7 missions de cette famille sont extraites à partir d'une base de missions de circulations réelles relativement à des segments de type triage et desserte locale (cf. [Figure 3.19](#)).



**Figure 3.19.** Ensemble de 7 missions ferroviaires de la locomotive hybride PLATHEE

En absence des données suffisantes à l'usage de la locomotive BB63000, notamment à l'occurrence de ses missions de durées très hétérogènes, nous proposons d'étudier un scénario

de 7 missions iso-durées (4 heures par mission) et de même occurrence obtenues par un enchaînement de plusieurs missions réelles (cf. [Annexe A](#)). Nous faisons l'hypothèse que l'état de charge du stockage, batteries Ni–Cd est ramené à 100% après chaque séquence de missions de durée 4 heures. Nous utilisons alors les indicateurs de conception de missions présentés au premier chapitre ( $P_{\max}$ ,  $P_{\text{moy}}$ ,  $E_u$ ,  $N_{\text{cyc}}$ ,  $I_{\text{stat}}$ ). Contrairement à la problématique de synthèse d'une mission à caractéristiques données traitée précédemment, les indicateurs de référence dans cette étude ne sont pas issus d'une seule mission mais d'un ensemble de missions témoignant de la « vie réelle » de la locomotive hybride.

- *Choix des indicateurs de référence*

Généralement, au sein d'une famille ou une classe de missions, il n'existe pas de mission dominante vis-à-vis de tous les indicateurs de conception. Le plus souvent, ce sont les missions énergétiquement difficiles qui imposent les indicateurs de dimensionnement de référence tandis que d'autres indicateurs pourtant potentiellement importants sont négligés. Dans notre approche, toutes les missions de la classe considérée participent à l'évaluation des indicateurs de performance et de durée de vie à prendre comme référence. Ces indicateurs de référence permettant de caractériser la mission représentative sont choisis de la manière suivante :

- les indicateurs de dimensionnement sont fixés aux valeurs les plus contraignantes sur l'ensemble des missions d'une même classe. Ceci nous permet de garantir le fait que le système à concevoir satisfera toutes les missions de la classe étudiée.
- les indicateurs de performance et de durée de vie sont considérés en moyenne.

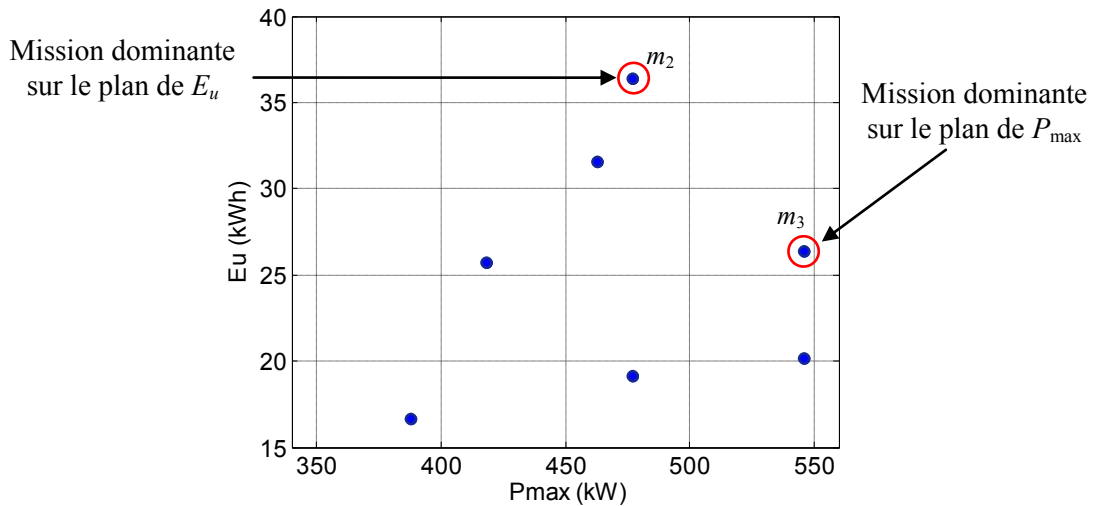
Une étude menée dans l'[Annexe F](#) a prouvé l'intérêt de choisir la puissance moyenne de référence  $P_{\text{moy ref}}$  à la valeur moyenne de la concaténation de toutes les missions. Dans le présent exemple, cette valeur est égale à 70 kW. Nous rappelons que cette valeur de puissance correspond au dimensionnement du groupe Diesel, dispositif dont la consommation et la « pollution » sont optimales lorsque ce dernier fonctionne à sa puissance nominale.

Une fois la puissance moyenne de référence est fixée, nous calculons la valeur de l'énergie utile du stockage (batteries) nécessaire à la satisfaction de chacune des missions selon l'expression suivante :

$$E_{ui} = \max E_{\text{stocki}}(t) - \min E_{\text{stocki}}(t) \quad (3.11)$$

$$\text{avec } E_{stocki}(t) = -\int_0^t (P_{missioni}(\tau) - P_{moyref}) d\tau \quad \text{où } t \in [0, \Delta t_{missioni}] \text{ et } i = \{1, \dots, 7\} \quad (3.12)$$

La répartition des missions selon les deux critères de dimensionnement  $P_{max}$  et  $E_u$  est donnée par la [Figure 3.20](#). Cette dernière montre qu'il n'y a pas une seule mission dominante sur le plan des deux critères  $P_{max}$  et  $E_u$ . En effet, la mission  $m_2$  est dominante sur le plan de l'énergie utile alors que la mission  $m_3$  présente la valeur la plus élevée de la puissance maximale. Ainsi, comme nous l'avons signalé plus haut, nous positionnons les indicateurs de dimensionnement de référence  $P_{max ref}$  et  $E_u ref$  aux valeurs les plus contraignantes ( $P_{max ref} = 546 \text{ kW}$  et  $E_u ref = 36.4 \text{ kWh}$ ).



**Figure 3.20.** Répartition des missions selon les indicateurs de dimensionnement  $P_{max}$  et  $E_u$

Le choix de l'énergie utile de référence  $E_u ref$  nous permet de déterminer la taille du stockage et par conséquent nous pouvons estimer le nombre de cycles consommés pour chacune des missions. Le nombre de cycles de référence  $N_{cyc ref}$  est déterminé alors par la valeur moyenne des nombres de cycles par heure de toutes les missions. Les différentes valeurs des indicateurs de conception des 7 missions étudiées sont données par le [Tableau 3.8](#). Signalons que l'indicateur de l'énergie utile pour chaque mission est calculé après avoir fixé la puissance moyenne de référence  $P_{moy ref}$ . Il en est de même pour l'indicateur du nombre de cycles évalué après le choix de  $E_u ref$ .

Nous récapitulons dans le [Tableau 3.9](#) les valeurs des indicateurs de référence. Pour l'indicateur statistique de référence  $I_{stat ref}$ , nous considérons la fonction de répartition de la mission totale issue de la concaténation des 7 missions.

	$P_{\max}$ (kW)	$P_{\text{moy}}$ (kW)	$E_u$ (kWh)	$Nb_{\text{cyc}}$ (/h)
$m_1$	418	75	25.7	0.078
$m_2$	477	77	36.4	0.176
$m_3$	546	71	26.3	0.085
$m_4$	546	70	20.2	0.090
$m_5$	463	76	31.6	0.134
$m_6$	388	54	16.6	0.030
$m_7$	477	65	19.1	0.070

**Tableau 3.8.** Indicateurs de conception des 7 missions ferroviaires

	$P_{\max \text{ ref}}$ (kW)	$P_{\text{moy ref}}$ (kW)	$E_{u \text{ ref}}$ (kWh)	$Nb_{\text{cyc ref}}$ (/h)
<i>mission représentative</i>	546	70	36.4	0.095

**Tableau 3.9.** Indicateurs de référence à imposer à la mission représentative

- *Choix de la durée de la mission représentative*

Le choix de la durée de la mission représentative est soumis à deux contraintes liées à la satisfaction de l'énergie utile de référence (faisant le couplage entre la puissance de la mission et le domaine temporel) ainsi qu'au nombre de cycles pour le stockage. La première contrainte traduit directement le temps minimal pour satisfaire au moins le triplet des indicateurs de dimensionnement de référence ( $P_{\max \text{ ref}}$ ,  $P_{\text{moy ref}}$ ,  $E_{u \text{ ref}}$ ). En effet, en faisant abstraction à la fois de l'indicateur statistique et du nombre de cycles, nous pouvons générer un profil représentatif et simple de la manière suivante :

Pendant un temps  $t_1$ , nous maintenons la puissance de la mission à la valeur maximale de référence  $P_{\max \text{ ref}}$  jusqu'à ce que l'énergie utile  $E_{u \text{ ref}}$  soit vérifiée. Nous rappelons que  $E_u$  est calculée par intégration de la mission de puissance (cf. [Figure 3.21](#)). L'expression du temps  $t_1$  est donnée par l'équation ci-dessous.

$$t_1 = \frac{E_{u \text{ ref}}}{P_{\max \text{ ref}} - P_{\text{moy ref}}} \quad (3.13)$$

Ensuite, nous complétons la mission par une phase de puissance nulle et de durée  $t_2$  tant que la moyenne du signal est supérieure à  $P_{\text{moy ref}}$ . Ce temps  $t_2$  doit vérifier la relation suivante :

$$P_{\text{moyref}} = \frac{P_{\text{maxref}} \times t_1}{t_1 + t_2} \quad (3.14)$$

Compte tenu de l'équation (3.13), il en résulte :

$$t_2 = \frac{E_{u\text{ref}}}{P_{\text{moyref}}} \quad (3.15)$$

Dans ce cas extrême, la durée minimale à respecter  $\Delta t_{\text{miss min 1}}$  est la somme des deux durées  $t_1$  et  $t_2$ . Son expression est alors la suivante :

$$\Delta t_{\text{miss min 1}} = t_1 + t_2 = E_{u\text{ref}} \times \frac{P_{\text{maxref}}}{P_{\text{moyref}}(P_{\text{maxref}} - P_{\text{moyref}})} \quad (3.16)$$

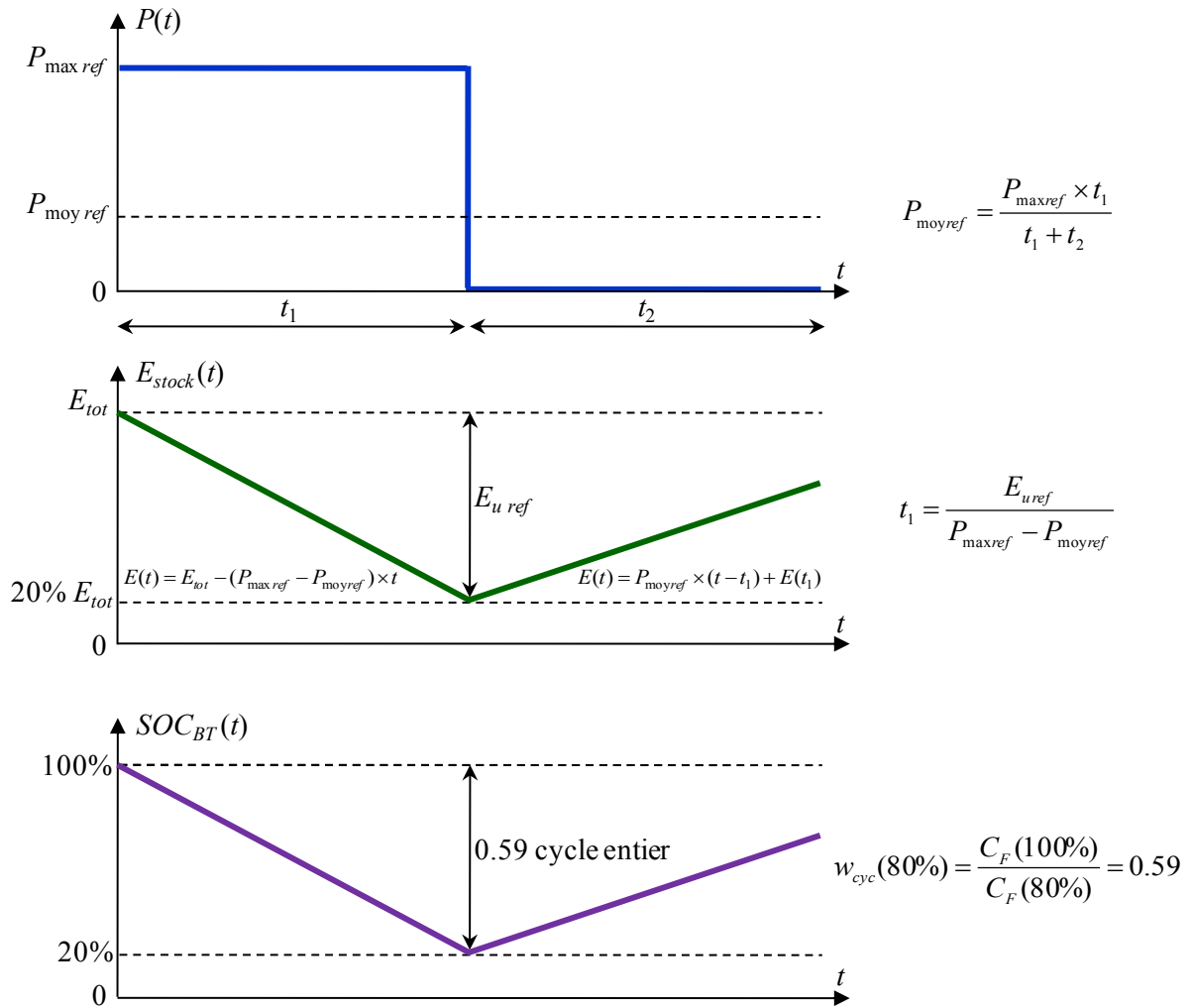


Figure 3.21. Illustration des contraintes de durée minimale à respecter



La deuxième contrainte est imposée par le nombre de cycles de référence par heure. En fait, pour les batteries Ni–Cd nous autorisons une exploitation énergétique de 80% de leur capacité de stockage totale  $E_{BT\ tot}$ . Il en découle l'apparition d'un grand cycle de profondeur 80% et qui correspond à la satisfaction de l'énergie utile de référence  $E_{u\ ref}$  (cf. [Figure 3.21](#)). Autrement dit, si notre mission représentative vérifie l'énergie utile de référence, alors les batteries seront sollicitées au moins sur un cycle de profondeur 80%. Comme le nombre de cycles est déterminé à partir de la courbe de l'état de charge des batteries  $SOC_{BT}(t)$  [[BAE 08](#)], il en résulte alors un nombre de cycles minimal à respecter. D'après la courbe constructeur, donnant le nombre de cycles admissibles en fonction de l'état de charge des batteries, un cycle de profondeur 80% représente 0.59 cycle entier ( $w_{cyc}(80\%)$ ).

$$w_{cyc}(80\%) = \frac{C_F(100\%)}{C_F(80\%)} = 0.8^{2.37} = 0.59 \quad (3.17)$$

Ce phénomène impose une deuxième durée minimale  $\Delta t_{miss\ min\ 2}$  à respecter de sorte que le produit de cette dernière par le nombre de cycles de référence par heure de mission vaille 0.59.

$$\Delta t_{miss\ min\ 2} = \frac{0.59}{N_{cyc\ ref\ par\ heure}} \quad (3.18)$$

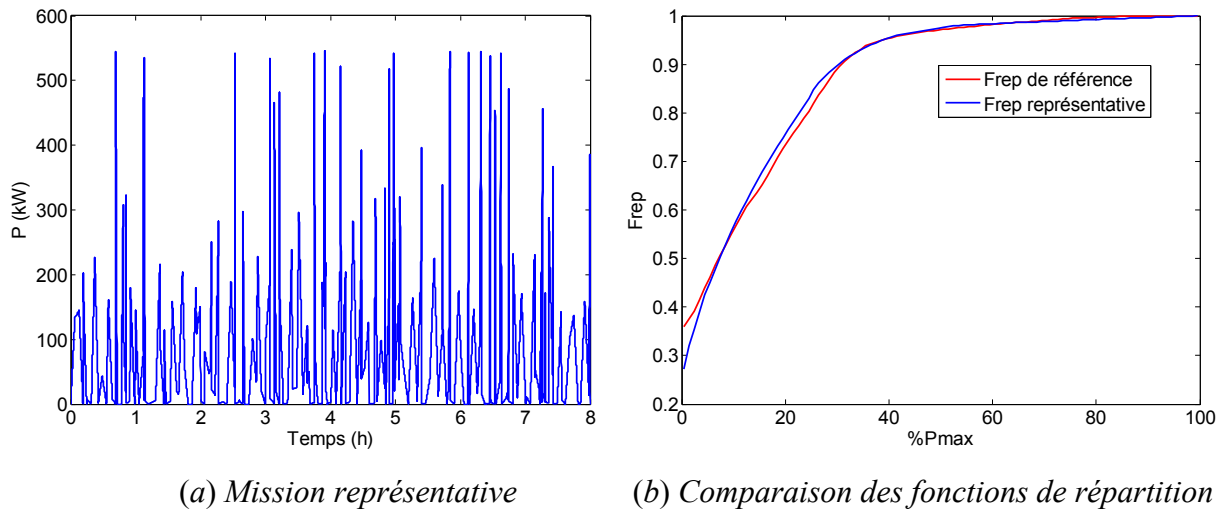
La durée de la mission représentative doit être alors bien supérieure à la plus grande durée entre  $\Delta t_{miss\ min\ 1}$  et  $\Delta t_{miss\ min\ 2}$  si l'on souhaite par ailleurs satisfaire les autres indicateurs de référence (nombre de cycle et distribution statistique).

Revenons à notre exemple, la durée minimale  $\Delta t_{miss\ min\ 1}$  vaut 36 minutes alors que la durée minimale relative au nombre de cycles  $\Delta t_{miss\ min\ 2}$  est égale à 6 heures et 13 minutes. Nous choisissons de fixer la durée de la mission représentative  $\Delta t_{miss}$  à 8 heures, soit presque 30% des  $7 \times 4 = 28$  heures pour les 7 missions concaténées.

- *Résultat du processus de synthèse*

Le résultat du processus de synthèse de mission est obtenu à partir du motif de type segment, associé au CLR comme algorithme évolutionnaire, avec la stratégie permettant un nombre de motifs variable. La taille de population est fixée à 100 individus et le nombre de générations vaut 20 000. La [Figure 3.22.a](#) représente la mission fictive ainsi générée de durée 8 heures et représentative des 7 missions de la locomotive BB63000. La mission synthétisée à

partir de la concaténation de 263 motifs (527 paramètres), respecte parfaitement les indicateurs de référence imposés (cf. [Tableau 3.10](#) et [Figure 3.22.b](#)).



**Figure 3.22.** Résultat du processus de synthèse avec le motif de type segment

	Indicateurs de référence	Indicateurs de la mission représentative	Erreur (%)
$P_{\max}$ (kW)	546	546	0.0
$P_{\text{moy}}$ (kW)	70	69.7	0.4
$E_u$ (kWh)	36.4	36.3	0.3
$Nb_{\text{cyc}}$ (h)	0.095	0.095	0.0
$I_{\text{stat}}$	—	—	1.4

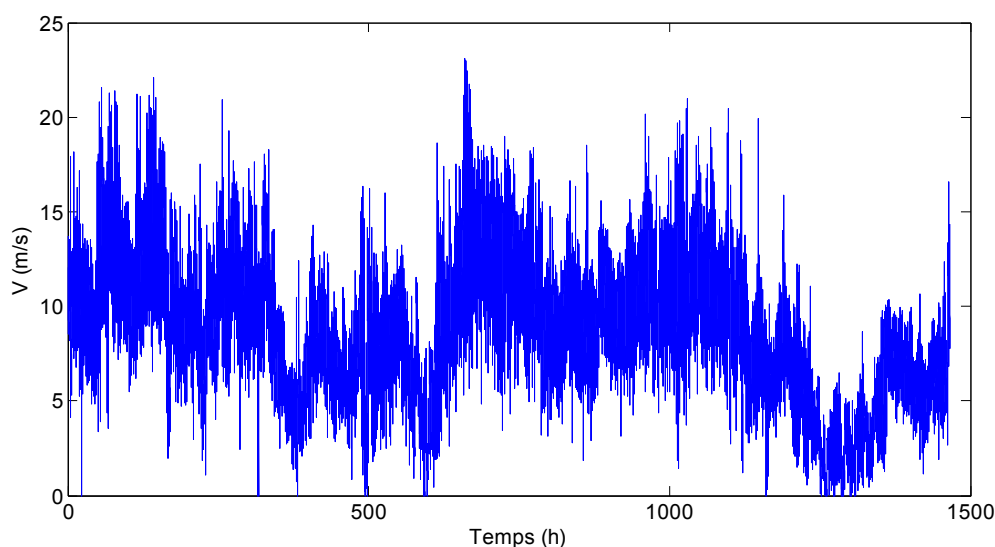
**Tableau 3.10.** Indicateurs de conception de la mission représentative

Enfin, à travers cet exemple issu de missions ferroviaires à coût de complexité très élevé, nous avons prouvé l'efficacité du processus de synthèse avec le motif de type segment couplé à l'algorithme d'optimisation CLR. Nous avons pu réduire la durée totale de toutes les missions du cahier des charges de 28 heures ( $7 \times 4$  h) à 8 heures de mission seulement, soit un gain de 3.5 sur la durée globale des missions. Nous montrerons au [chapitre 4](#) comment ce rapport peut se traduire par un gain notable en termes de temps de traitement du processus de conception simultanée par optimisation.

### 3.5.2 Application à un profil de vitesse du vent

Le deuxième exemple s'inscrit dans le cadre d'une démarche de conception systémique d'une chaîne éolienne dont la variable environnementale dimensionnante est la vitesse du

vent. En raison de son caractère intermittent voire capricieux, il est inévitable de tenir compte des caractéristiques statistiques de ce vecteur d'énergie primaire sur des horizons de temps suffisamment longs<sup>16</sup>. Afin d'optimiser le coût de traitement lors des phases de conception par optimisation, il est alors plus judicieux de considérer un profil de vitesse du vent de durée minimale tout en restant pertinent vis-à-vis des critères et contraintes de conception. Nous signalons qu'au niveau de ce manuscrit, nous portons notre attention seulement sur la synthèse d'un profil de vitesse du vent représentatif et compact. Pour plus de détails sur la conception d'une chaîne éolienne par optimisation, nous invitons le lecteur à consulter [ABD 07] et [TRA10]. Deux approches différentes de synthèse de profil représentatif ont été appliquées pour deux mois de mesures de la vitesse du vent sur le site de « Petit Canal » en Guadeloupe avec un pas d'échantillonnage de 1 seconde (cf. Figure 3.23). La première approche utilise le processus de synthèse de signal représentatif par optimisation décrit au début de ce chapitre. La deuxième reprend une approche purement statistique établie dans [ABD 07] et basée sur des générateurs de nombres aléatoires dont les densités de probabilité sont issues de la distribution statistique des données de vitesse du vent étudiées. Enfin, nous établissons une étude comparative des deux approches pour le même coût de simulation.



**Figure 3.23.** Relevé de deux mois de mesures de la vitesse du vent sur le site de « Petit Canal » en Guadeloupe

- *Approche de synthèse par algorithme évolutionnaire*

La première approche de synthèse d'un profil de vitesse du vent représentatif et compact consiste à utiliser le processus de synthèse par optimisation d'une variable environnementale

<sup>16</sup> De l'ordre de quelques années

présenté au début de ce chapitre. Les indicateurs de conception de la vitesse du vent sont : la vitesse maximale  $V_{\max}$ , la vitesse cubique moyenne  $\langle V^3 \rangle$ , la distribution statistique *pdf* et l'indicateur de turbulence  $I_{turb}$ . Nous rappelons au lecteur que ces indicateurs sont définis dans le [chapitre 1](#). En plus de ces indicateurs de conception, nous imposons une contrainte relative à la dynamique de la vitesse du vent. Il s'agit d'imposer le gradient maximal de vitesse du vent (accélération de la masse d'air). Pour les données de vitesse du vent du site « Petit Canal » de la Guadeloupe, le gradient maximal de la vitesse du vent est de  $12 \text{ m/s}^2$ . Dans le processus de synthèse de profil représentatif par optimisation, cette contrainte est exprimée comme suit :

$$g = \frac{dv(t)}{dt} - 12 \leq 0 \quad (3.19)$$

Contrairement à la synthèse d'une mission représentative d'une famille de missions ferroviaires, le problème de choix des indicateurs de référence ne se pose pas pour cet exemple de synthèse d'un profil de vitesse du vent. En effet, nous disposons d'un seul profil initial (deux mois de mesures de la vitesse du vent) dont nous cherchons à reproduire les caractéristiques avec une durée plus compacte. Tous les indicateurs de référence sont alors issus directement de ce profil (cf. [Tableau 3.11](#)). Néanmoins, le choix de la durée du profil représentatif n'est pas évident. Bien que les indicateurs de conception n'imposent aucune durée minimale, cette dernière doit contenir suffisamment d'échantillons afin de respecter la distribution statistique de référence (*pdf*).

La fonction d'erreur à optimiser est la suivante :

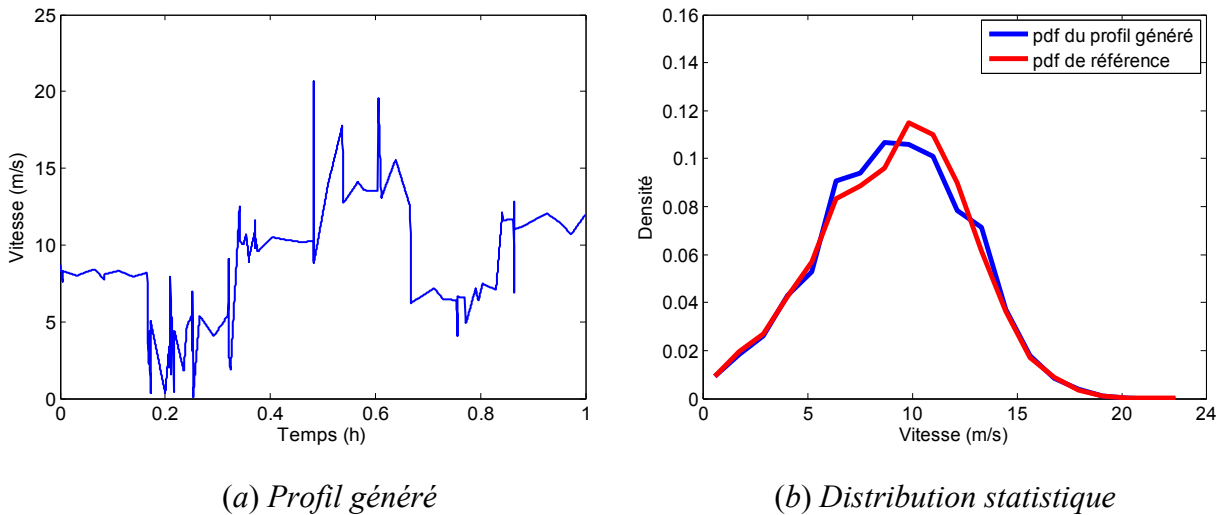
$$\mathcal{E} = \left( \frac{V_{\max} - V_{\max \text{ ref}}}{V_{\max \text{ ref}}} \right)^2 + \left( \frac{\langle V^3 \rangle - \langle V^3 \rangle_{\text{ref}}}{\langle V^3 \rangle_{\text{ref}}} \right)^2 + \left( \frac{I_{turb} - I_{turb \text{ ref}}}{I_{turb \text{ ref}}} \right)^2 + \mathcal{E}_{\text{stat}} \quad (3.20)$$

$\mathcal{E}_{\text{stat}}$  désignant l'erreur quadratique moyenne entre la densité de probabilité de référence et celle du profil fictif généré évaluée sur 20 intervalles égaux entre 0 et  $V_{\max \text{ ref}}$ .

$$\mathcal{E}_{\text{stat}} = \frac{1}{20} \times \sum_{k=1}^{20} \left( \frac{pdf(k) - pdf_{\text{ref}}(k)}{pdf_{\text{ref}}(k)} \right)^2 \quad (3.21)$$

Nous signalons que la densité de probabilité de référence  $pdf_{\text{ref}}$  est déterminée à partir de la vitesse réelle du vent mesurée avec un pas d'échantillonnage de 1 seconde.

Plusieurs tests ont été effectués avec les trois motifs élémentaires (sinus, sinus cardinal, segment). Là encore, comme pour l'exemple précédent, les meilleurs résultats sont obtenus avec les segments couplés au CLR comme algorithme évolutionnaire. La [Figure 3.24](#) donne un résultat de synthèse de profil représentatif de durée 1 heure (à comparer avec la base de données initiale qui dure 2 mois, soit presque 1500 heures). La taille de la population est fixée à 100 individus et le nombre de générations est égal à 75 000. Le nombre de motifs élémentaires (segments) obtenu est de 103.



**Figure 3.24.** Résultat du processus de synthèse pour une durée de 1 heure

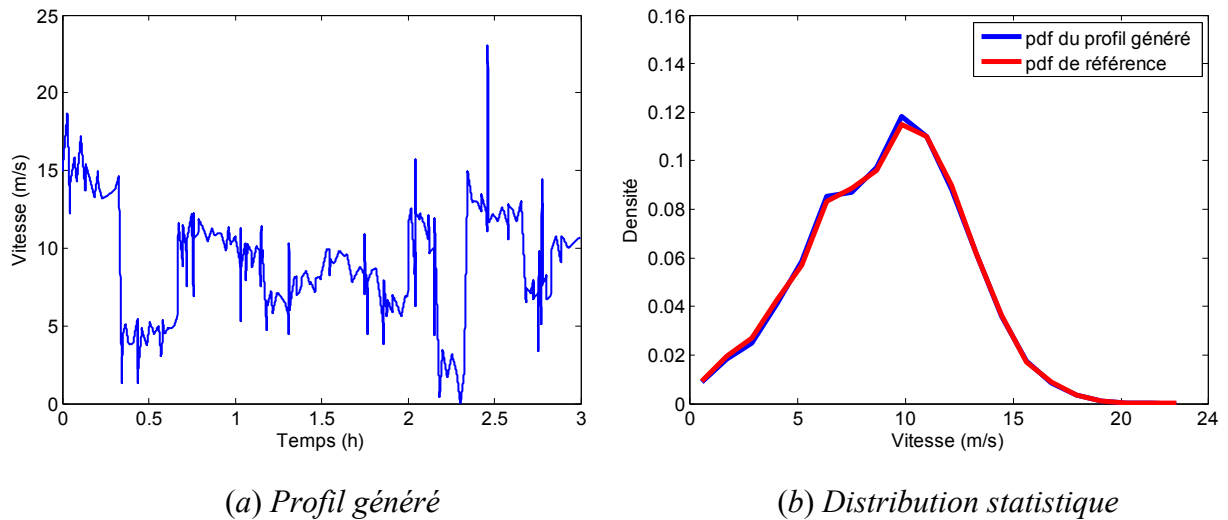
	Indicateurs de référence	Indicateurs du profil généré	Erreur (%)
$V_{\max}$ (m/s)	23.1	20.7	10.4
$\langle V^3 \rangle$ ( $m^3/s^3$ )	1076	1074	0.2
$I_{turb}$	0.72	0.78	8.3

**Tableau 3.11.** Indicateurs de conception du profil représentatif

Les caractéristiques du profil de vitesse du vent généré sont résumées dans le [Tableau 3.11](#). Nous constatons à partir de ce tableau que les indicateurs de référence ne sont pas exactement « retrouvés » : le taux d'erreur est à peine supérieur à 10% pour l'indicateur de vitesse maximale et à 8% pour l'indicateur de turbulence. Cet écart n'est pas la conséquence d'un faible nombre de générations ou d'une convergence prématurée mais est plutôt lié à la façon avec laquelle nous avons défini l'indicateur de turbulence. En effet, pour un profil de durée 1 heure, nous disposons seulement de 6 valeurs<sup>17</sup> d'écart-types, ce qui présente un effectif relativement faible pour satisfaire l'indice de turbulence défini par la moyenne de ces valeurs.

<sup>17</sup>  $60 \text{ min}/T_e$  avec  $T_e = 10 \text{ min}$

Une première solution possible pour remédier à ce problème consiste à augmenter tout simplement la durée du profil généré afin d'accroître l'effectif des écarts-types utiles au calcul de l'indicateur de turbulence. Par exemple avec un profil de durée 3 heures, donc 18 valeurs d'écarts-type, les résultats sont nettement meilleurs par rapport au premier cas (cf. [Figure 3.25](#) et [Tableau 3.12](#)). Le profil généré est obtenu avec 217 motifs au bout de 100 000 générations avec une population de 100 individus.



**Figure 3.25.** Résultat du processus de synthèse pour une durée de 3 heures

	Indicateurs de référence	Indicateurs du profil généré	Erreur (%)
$V_{\max}$ (m/s)	23.1	23.1	0.0
$\langle V^3 \rangle$ ( $\text{m}^3/\text{s}^3$ )	1076	1076	0.0
$I_{\text{turb}}$	0.72	0.74	2.7

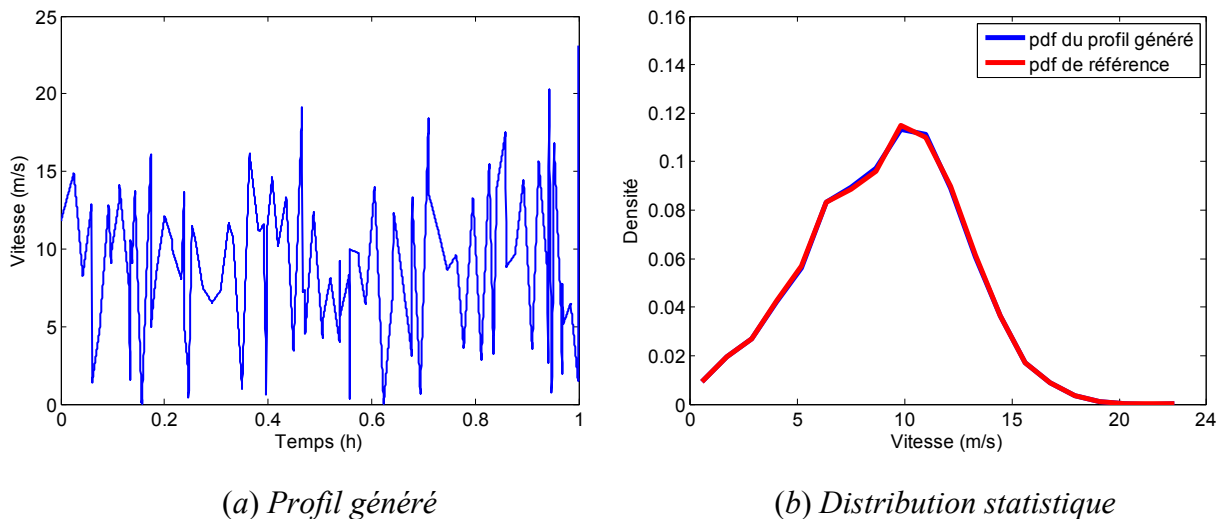
**Tableau 3.12.** Indicateurs de conception du profil représentatif

À noter qu'on obtient ici des résultats « quasi idéaux » au sens des indicateurs de conception tandis que la base de données de 2 mois de mesures (1500 heures) est *compactée dans un rapport 500* ( $\sim 1500/3$ ). Par rapport à l'exemple de missions ferroviaires, ce taux de réduction de l'horizon est supérieur en raison de la nature des indicateurs : en effet, comme nous l'avons souligné au paragraphe précédent, les indicateurs d'énergie utile et de cyclage imposent des temps minimums substantiels dans le cas de la locomotive alors que les indicateurs du système éolien sont « rapidement synthétisables ».

Une deuxième solution pour améliorer la précision consiste à intégrer le phénomène de turbulence comme contrainte dans le processus de synthèse par optimisation. Au lieu d'imposer un écart-type moyen (indicateur de turbulence), nous imposons une plage de

variation de l'écart-type délimitée par un écart-type minimal ( $\sigma_{\min}$ ) et un écart-type maximal ( $\sigma_{\max}$ ). Pour les 2 mois de mesures de la vitesse du vent du site « Petit Canal », l'écart-type des échantillons du vent de 10 minutes varie entre  $\sigma_{\min} = 0.05$  m/s et  $\sigma_{\max} = 4.9$  m/s.

Une illustration de cette solution pour un profil de durée 1 heure est donnée par la [Figure 3.29.a](#). Ce profil de vitesse du vent est obtenu par concaténation de 105 segments au bout de 15 000 générations seulement ce qui correspond à 5 heures de simulation sur un PC standard (*Core Duo* 2 GHz). Nous signalons que, malgré la complexité du problème qui s'élève à 211 paramètres, le CLR converge rapidement vers une solution très convenable en respectant à la fois la contrainte relative au phénomène de turbulence et la contrainte liée à la dynamique de la vitesse du vent. En effet, la [Figure 3.29.b](#) montre que la distribution statistique du profil généré coïncide parfaitement avec la distribution de référence. Le [Tableau 3.13](#) montre aussi le faible écart entre les indicateurs de référence et les indicateurs de conception du profil généré.



**Figure 3.26.** Résultat du processus de synthèse pour une durée de 1 heure

	Indicateurs de référence	Indicateurs du profil généré	Erreur (%)
$V_{\max}$ (m/s)	23.1	23.1	0.0
$\langle V^3 \rangle$ (m <sup>3</sup> /s <sup>3</sup> )	1076	1078	0.2

**Tableau 3.13.** Indicateurs de caractérisation du profil représentatif

En conclusion, l'approche de synthèse par algorithme évolutionnaire a montré son efficacité dans la précision et dans le compactage de l'information contenue dans un profil réel de vitesse du vent. En effet, en considérant l'indicateur de turbulence comme critère d'optimisation, nous avons pu compacter deux mois de mesures de la vitesse du vent en

3 heures seulement. Dans le cas où nous intégrons le phénomène de turbulence comme contrainte dans le processus de synthèse, la durée du profil représentatif peut être réduite jusqu'à 1 heure avec une très bonne précision sur les trois autres critères. Ce compactage de l'information apportera un gain notable en termes de temps de calcul lors d'un processus de conception simultanée par optimisation. Dans le paragraphe suivant, nous étudions à titre de comparaison, une autre approche de synthèse de profil représentatif.

- *Approche de synthèse purement statistique*

Cette deuxième méthode de synthèse de la vitesse du vent est basée sur une approche purement statistique. Elle consiste à générer un signal temporel continu à partir de la distribution statistique des données réelles de la vitesse du vent. Parmi les lois de probabilité qui peuvent caractériser la distribution statistique de la vitesse moyenne du vent, on note la distribution *log-normale*, la distribution *gaussienne bivariate* et la distribution de *Weibull*. Cette dernière est souvent la plus appropriée à la description des propriétés statistiques du vent réel moyen [KEL 92]. L'expression de la densité de probabilité de la loi de *Weibull* en fonction de la vitesse moyenne du vent ( $\bar{v}$ ) est donnée par l'expression suivante :

$$f(\bar{v}) = \frac{k}{c} \left( \frac{\bar{v}}{c} \right)^{k-1} \exp \left( - \left( \frac{\bar{v}}{c} \right)^k \right) \quad (3.22)$$

avec  $k$  est un facteur de forme sans dimension et  $c$  est le paramètre d'échelle qui traduit la vitesse moyenne des données du vent traitées.

La [Figure 3.27](#) illustre la distribution de la vitesse du vent pour les deux mois de mesures sur le site de « Petit Canal » avec un pas d'échantillonnage de 10 minutes. Cette distribution est identifiée à une loi de *Weibull* de paramètres  $k = 3.03$  et  $c = 10.62$ .

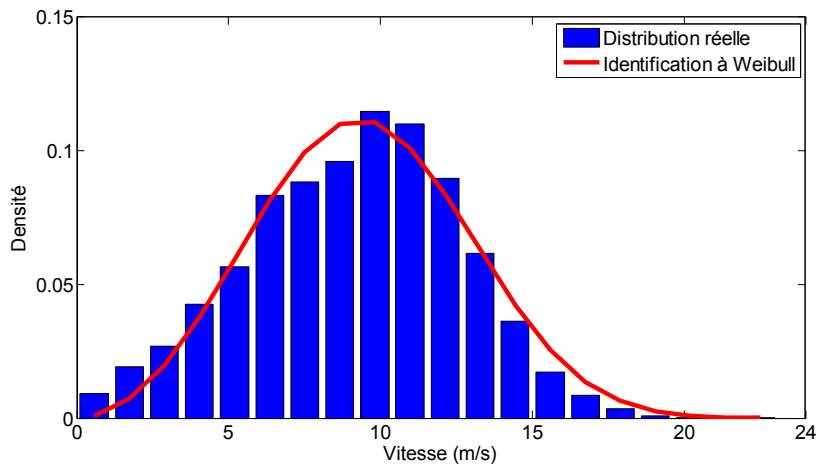
Une synthèse sur la base de statistiques de la vitesse instantanée du vent réel développée dans [ABD 07] consiste à décomposer cette dernière selon deux dynamiques différentes : une dynamique lente, qui caractérise les variations de la vitesse moyenne selon une distribution statistique de *Weibull*, et une dynamique rapide caractéristique du phénomène de turbulence et qui suit une distribution *gaussienne* [STR 90] (cf. [Figure 3.28](#)). Selon la première dynamique, qualifiée de lente ou moyenne, l'approche purement statistique consiste à générer un signal de vitesse du vent dont la distribution statistique suit la même loi de probabilité déterminée à partir d'une phase d'analyse des données réelles de la vitesse moyenne du vent. Il s'agit de



générer un certain nombre d'échantillons  $N_e$  selon un générateur de nombres aléatoires ayant comme densité la loi de probabilité de référence ainsi déterminée. Dans le cas d'une distribution de *Weibull*, le générateur de nombres aléatoires  $W(c, k)$  est défini à partir de la transformation inverse de la densité de probabilité donnée par l'équation (3.22). En se référant à un générateur de nombres aléatoires à densité uniforme dans l'intervalle  $[0, 1]$  ( $U(0,1)$ ) et en connaissant les paramètres de la distribution de *Weibull* ( $c$  et  $k$ ), l'expression du générateur de nombres aléatoires  $W(c, k)$  est la suivante :

$$W(c, k) = c(-\ln U(0, 1))^{\frac{1}{k}} \quad (3.23)$$

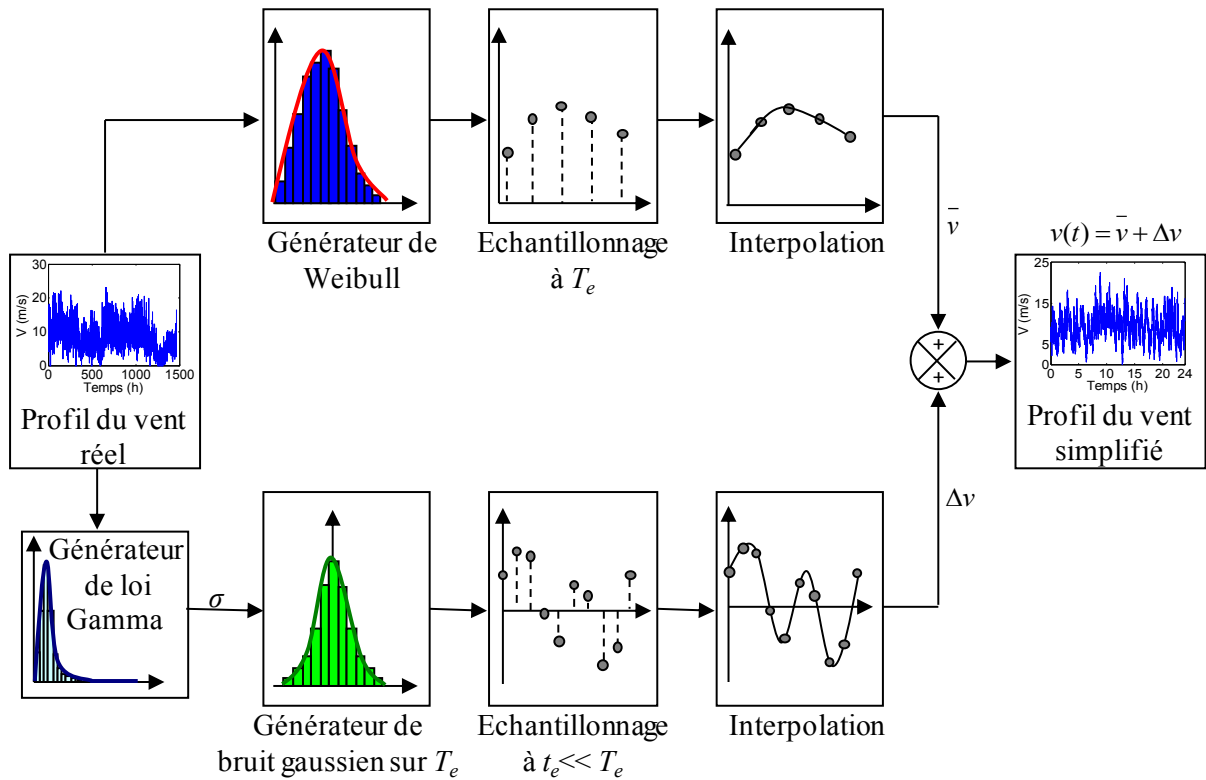
Le signal continu est obtenu par interpolation des  $N_e$  échantillons ainsi générés.



**Figure 3.27.** Identification de la distribution statistique de la vitesse du vent du site « Petit Canal » à une loi de Weibull

La deuxième dynamique, qualifiée de rapide, permet d'intégrer le phénomène de turbulence<sup>18</sup> de la vitesse du vent dans le signal temporel représentatif. Sur chaque période d'échantillonnage de la dynamique lente  $T_e$ , la turbulence de la vitesse du vent est modélisée par un bruit gaussien de moyenne nulle ( $\mu = 0$ ) et d'écart-type  $\sigma$ . Le signal représentatif de cette turbulence est généré à partir d'un générateur de nombres aléatoires à densité *gaussienne*. Il s'agit du même principe que le générateur à densité de *Weibull* (cf. Figure 3.28) sauf que la période d'échantillonnage  $t_e$  doit être très faible par rapport à  $T_e$ . Enfin, le signal de vitesse du vent global est obtenu par la somme de la composante lente ( $\bar{v}$ ) générée à partir de la densité de *Weibull* et de la composante rapide ( $\Delta v$ ).

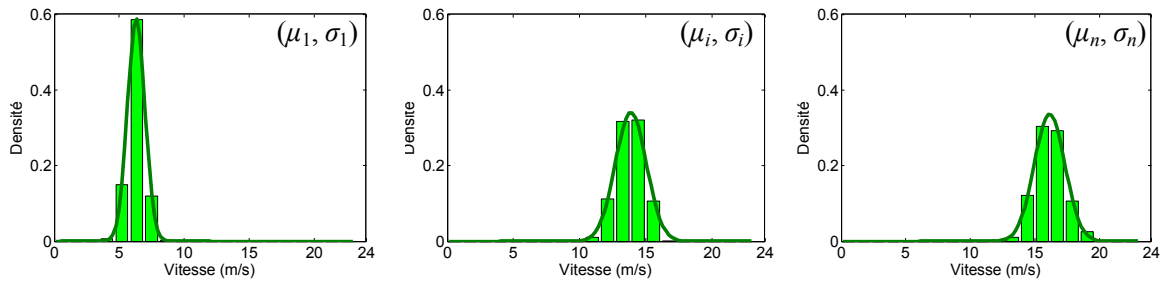
<sup>18</sup> Nous signalons que la turbulence du vent est un phénomène important car couplé à la fatigue mécanique de l'éolienne.



**Figure 3.28.** *Processus de génération statistique d'un signal du vent représentatif*

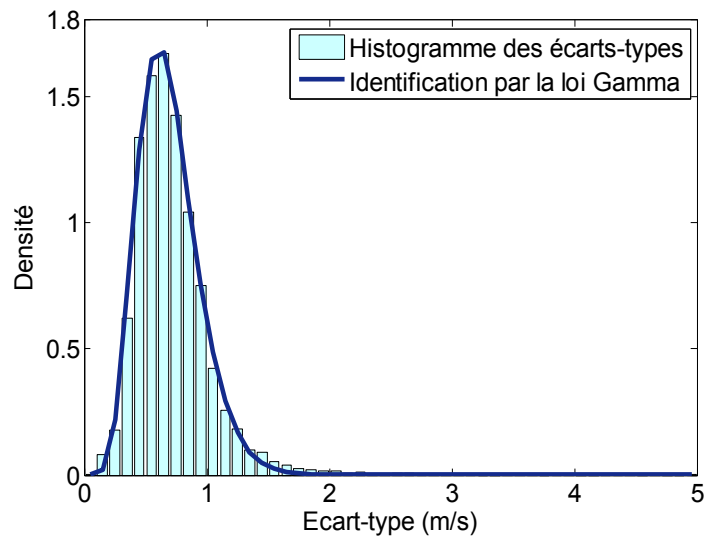
Dans [ABD 07], l'écart-type  $\sigma$  de la loi *gaussienne* a été considéré constant tout le long du signal du vent généré. Dans ce travail, nous proposons de rendre plus réaliste l'approche statistique en produisant la turbulence (bruit gaussien) avec un écart-type variable. Par le biais d'une phase d'analyse des données réelles de la vitesse du vent, nous identifions la distribution de la composante rapide, sur chaque intervalle  $i$  de durée  $T_e$ , à une loi *gaussienne* d'écart-type  $\sigma_i$ . Ensuite, nous déterminons la distribution statistique  $D_\sigma$  des écarts-types ( $\sigma_i$ ). Au niveau du profil du vent généré, sur chaque période  $T_e$ , l'écart-type du bruit gaussien est déterminé à partir d'un générateur de nombres aléatoires de même densité que la distribution statistique des écarts-types ( $D_\sigma$ ). La Figure 3.29 donne la distribution statistique de la vitesse réelle du vent du site « Petit Canal » sur quelques fenêtres de largeur  $T_e = 10$  minutes. Nous vérifions bien que ces distributions suivent, dans la plupart des cas, une loi *gaussienne* de paramètres  $(\mu, \sigma)$ , centrée sur la valeur de la composante lente de la vitesse du vent dans la fenêtre considérée de largeur  $T_e$  et d'expression :

$$f(v) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(v-\mu)^2}{2\sigma^2}\right) \quad (3.24)$$



**Figure 3.29.** Distribution statistique de la vitesse réelle du vent du site « Petit Canal » sur quelques fenêtres de largeur  $T_e = 10$  minutes

La distribution statistique  $D_\sigma$  des écarts-types ( $\sigma_i$ ) sur les deux mois de mesures de la vitesse du vent du site « Petit Canal » est donnée par l'histogramme de la Figure 3.30. Cette distribution est identifiée à une loi de densité *Gamma* de paramètre  $\alpha = 7.9$  et  $\beta = 11.3$ .



**Figure 3.30.** Distribution statistique des écarts-types

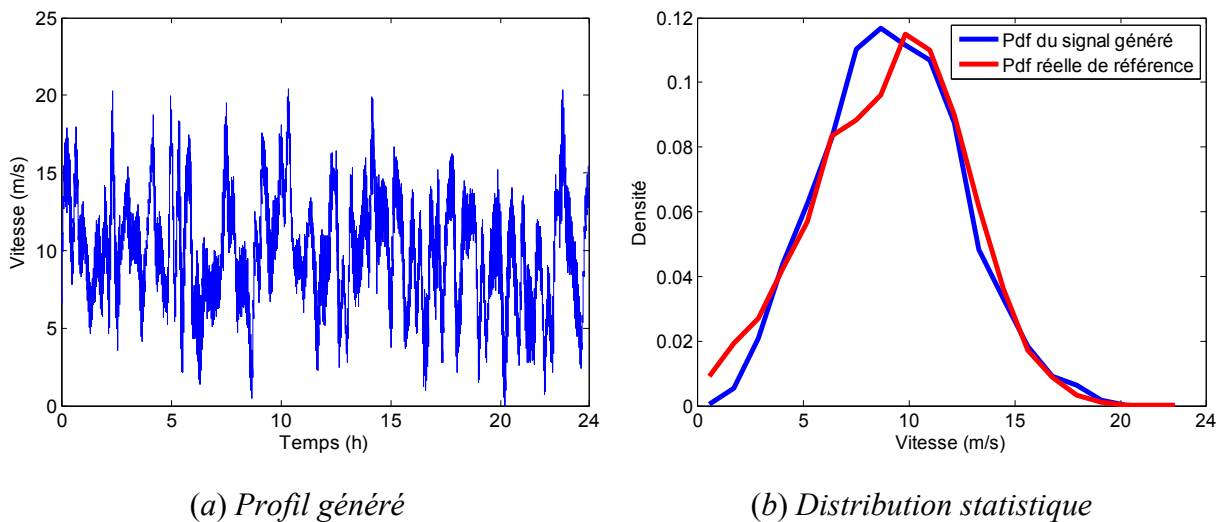
L'expression d'une loi de probabilité de densité *Gamma*  $f(\sigma)$  et de paramètres  $(\alpha, \beta)$  est donnée par la relation suivante :

$$f(\sigma) = \frac{\beta^\alpha}{\Gamma(\alpha)} \sigma^{\alpha-1} \exp(-\beta\sigma) \quad (3.25)$$

où  $\Gamma$  est la fonction mathématique définie pour tout réel  $x$  positif non nul par :

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$$

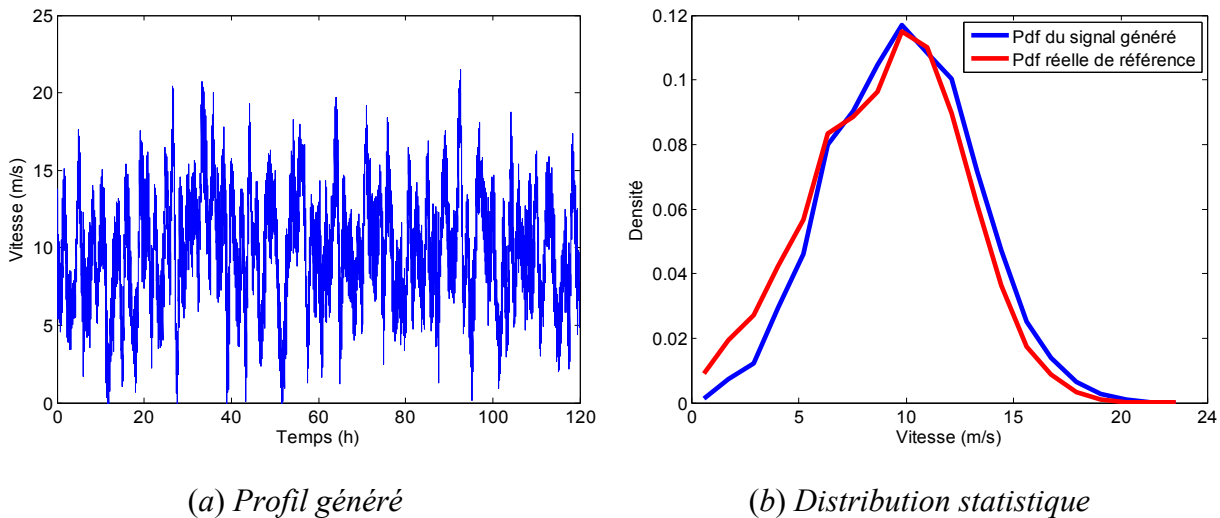
Un résultat de simulation de l'approche statistique pour un profil de vitesse du vent de durée 24 heures et représentatif des deux mois de mesures sur le site « Petit Canal » est donné par la [Figure 3.31.a](#). La période d'échantillonnage de la composante lente du vent  $T_e$  est fixée à 10 minutes et celle de la dynamique rapide (turbulence)  $t_e$  à 1 seconde. Il en résulte 144 échantillons pour le générateur de *Weibull* de paramètres  $k = 3.03$  et  $c = 10.65$  et 600 échantillons pour le générateur du bruit gaussien par intervalle de durée  $T_e$ . La [Figure 3.31.b](#) donne une comparaison entre les distributions statistiques (*pdf*) de la vitesse du vent du signal généré d'une part et du signal initial (2 mois) d'autre part. Nous signalons que ces distributions statistiques ne décrivent pas une dynamique particulière mais plutôt la dynamique totale du signal traité. En effet, elles sont déterminées par un échantillonnage fin avec un pas de 1 s.



**Figure 3.31.** Résultat de l'approche statistique pour un horizon temporel de 24 heures

A travers une analyse de la dynamique rapide du profil ainsi obtenue, nous avons constaté que la distribution statistique sur quelques fenêtres de 10 minutes ne suit pas la loi *gaussienne* de référence. Cette différence est due au fort gradient de vitesse entre deux échantillons successifs issus du générateur de *Weibull* et distants d'une durée  $T_e$  ce qui modifie la distribution statistique de la composante rapide à densité *gaussienne*. Pour remédier à ce genre de problème, une solution consiste à augmenter la période d'échantillonnage du générateur de *Weibull* (par exemple  $5T_e$ ) afin de garantir au mieux le découplage entre dynamiques lente et rapide. Cette solution trouve rapidement ses limites en termes de durée de profil représentatif. En effet, choisir un pas d'échantillonnage multiple de  $T_e$  ne fait qu'augmenter la durée du signal généré, ce qui pénalise le coût de calcul dans un processus de conception par optimisation. Nous présentons à la [Figure 3.32](#) une simulation de l'approche

statistique avec un pas d'échantillonnage de  $5T_e$  pour le générateur de *Weibull* (les  $N_e$  échantillons issus du générateur de *Weibull* sont distants de  $5T_e = 50$  minutes). Il s'agit d'un profil représentatif de durée 120 heures (5 jours) dont la distribution statistique (*pdf*) suit pratiquement la distribution de référence (cf. Figure 3.32.b). En outre, nous vérifions bien que la distribution de la dynamique rapide du signal généré suit, dans la plupart des cas, une distribution *gaussienne* sur chaque intervalle de 10 minutes.



**Figure 3.32.** Résultat de l'approche statistique pour un horizon temporel de 120 heures

Pour conclure, cette approche statistique permet de générer des signaux de vent relativement réalistes à condition que la durée du signal synthétisé soit suffisamment importante. Néanmoins, elle ne garantit pas en l'état le contrôle de l'accélération de la masse d'air représentée par le signal généré. Autrement dit, aucune contrainte n'est imposée sur le gradient de la vitesse du vent entre deux échantillons successifs issus des générateurs de nombres aléatoires.

- *Comparaison des deux approches de synthèse*

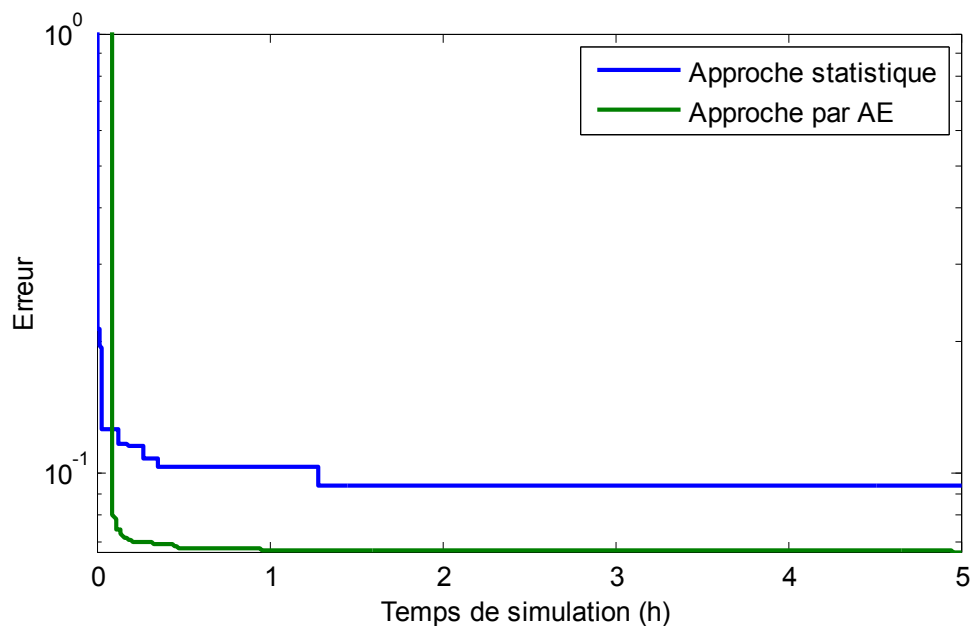
Dans cette partie, nous procédons à la comparaison des deux approches de synthèse de profil représentatif développées précédemment pour *un temps CPU* de 5 heures. Ce temps de simulation correspond au temps de convergence nécessaire à l'approche de synthèse par algorithme évolutionnaire pour converger correctement sur le problème « Petit Canal » présenté en début de section. Pour l'approche de synthèse purement statistique, nous procédons à la génération d'un signal de vitesse de vent par des itérations successives jusqu'à obtenir un temps de traitement identique à celui nécessaire à la convergence de l'algorithme évolutionnaire (5 heures). Le signal issu de chaque itération est retenu s'il améliore le même

critère d'erreur  $\varepsilon$  utilisé dans l'approche de synthèse par algorithme évolutionnaire. Tel qu'il est formulé (cf. équation (3.26)), ce critère d'erreur vise à optimiser simultanément la vitesse maximale, la vitesse cubique moyenne et la distribution statistique de la vitesse du vent, indicateurs considérés comme pertinents à la conception d'une éolienne.

$$\varepsilon = \left( \frac{V_{\max} - V_{\max ref}}{V_{\max ref}} \right)^2 + \left( \frac{\langle V^3 \rangle - \langle V^3 \rangle_{ref}}{\langle V^3 \rangle_{ref}} \right)^2 + \frac{1}{20} \times \sum_{k=1}^{20} \left( \frac{pdf(k) - pdf_{ref}(k)}{pdf_{ref}(k)} \right)^2 \quad (3.26)$$

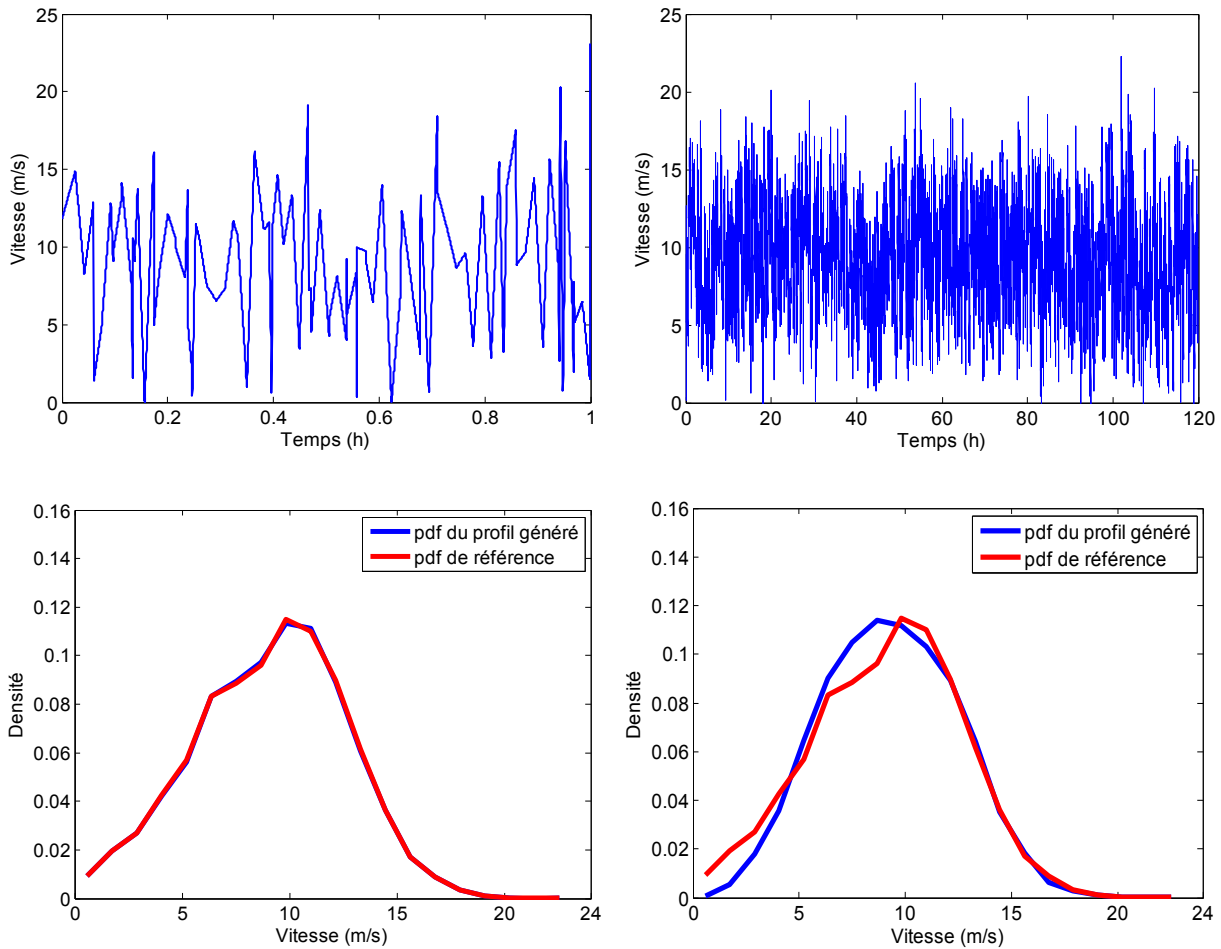
Que ce soit avec l'approche statistique ou avec l'approche par algorithme évolutionnaire, nous considérons la distribution statistique du vent réel du site « Petit Canal » comme référence avec un pas d'échantillonnage de 1 seconde.

La durée du signal représentatif est fixée à 1 heure pour l'approche de synthèse par algorithme évolutionnaire et 5 jours (120 heures) pour l'approche purement statistique. Pour les premières minutes, nous notons une violation des contraintes du problème d'optimisation avec l'approche de synthèse par algorithme évolutionnaire ce qui conduit à une pénalisation du critère d'erreur (cf. Figure 3.33).



**Figure 3.33.** Évolution du critère d'erreur pour les deux approches de synthèse

Les profils représentatifs obtenus, après 5 heures de calcul sur PC standard (*Core Duo* 2 GHz), ainsi que leurs caractéristiques sont donnés respectivement par la Figure 3.34 et le Tableau 3.14.



*Approche par algorithme évolutionnaire*

*Approche purement statistique*

**Figure 3.34.** Comparaison des résultats de génération de profil représentatif par les deux approches de synthèse

		$V_{\max}$ (m/s)	$\langle V^3 \rangle$ ( $\text{m}^3/\text{s}^3$ )
Indicateurs de référence		23.1	1076
Indicateurs du profil fictif	AE	23.1	1078
	Statistique	22.2	1077
Erreur (%)	AE	0	0.2
	Statistique	3.9	0.1

**Tableau 3.14.** Comparaison des caractéristiques des profils générés par les deux approches de synthèse

Les résultats précédents montrent que l'approche de synthèse par algorithme évolutionnaire permet d'avoir une meilleure précision sur l'ensemble des indicateurs de conception. En particulier, la distribution statistique du signal obtenu avec cette approche suit parfaitement la distribution de référence, alors qu'elle présente des écarts importants dans le

cas de l'approche purement statistique (cf. [Figure 3.34](#)). Compte tenu des indicateurs de conception pertinents vis-à-vis des critères et contraintes de conception d'une éolienne, l'approche de synthèse de signal par algorithme évolutionnaire nous offre aussi l'avantage d'une durée plus compacte. Ceci est bien en faveur d'une démarche de conception par optimisation, démarche pour laquelle un horizon temporel réduit de la vitesse du vent apporte un gain notable en termes de temps de calcul. En effet, deux mois de mesures de la vitesse du vent sont représentés seulement par un signal de durée de 1 heure.

### 3.6 Conclusion

Dans ce chapitre, nous avons mis en œuvre un processus de synthèse de profil représentatif d'un ensemble de profils d'une mission ou variable environnementale. Ce processus est basé sur l'agrégation ou la concaténation de motifs élémentaires dont le nombre et les paramètres sont déterminés par optimisation à l'aide d'un algorithme évolutionnaire. L'objectif est de synthétiser un profil fictif et de durée minimale qui respecte un jeu d'indicateurs de conception pertinents vis-à-vis des critères et contraintes de conception. Trois motifs élémentaires étaient proposés : le sinus, le sinus cardinal et le segment. Nous avons montré que le choix du motif élémentaire n'est pas évident et dépend de la complexité et de la variabilité du profil réel de la variable environnementale en question. Ensuite, nous avons évalué la capacité des motifs élémentaires à reproduire, d'une part, un profil réel et d'une autre part à synthétiser un profil à caractéristiques données au sens des indicateurs de conception et ce en utilisant trois algorithmes évolutionnaires différents avec et sans technique de nichage. Nous avons constaté que, même si certains motifs sont incapables de reproduire la forme du profil réel, ceci ne présente pas un inconvénient majeur pour l'approche de synthèse de profil représentatif de quelques caractéristiques (indicateurs) du signal. A travers deux exemples de profils réels de missions ferroviaires et de variable environnementale d'un système éolien (i.e. la vitesse du vent), nous avons montré la capacité du processus de synthèse à générer un profil représentatif des indicateurs de conception et de durée réduite.

Le taux de réduction de l'horizon temporel à partir du profil de base dépend de l'application, et plus précisément des critères de conception qui dictent les indicateurs auxquels se réfère le processus de synthèse : ainsi, l'application à la synthèse de vitesse de vent conduit à des rapports de « compactage » nettement supérieurs au cas de la locomotive



hybride, application pour laquelle des temps minima incompressibles sont imposés par certains indicateurs tels que l'énergie utile ou le cyclage de l'organe de stockage dans la chaîne de traction.

Pour les deux exemples traités, les meilleurs résultats sont obtenus avec le motif élémentaire de type segment, couplé au Clearing (CLR) comme algorithme évolutionnaire. Enfin, bien que les résultats de synthèse obtenus dans ce chapitre soient très satisfaisants, le coût de complexité des problèmes traités ne garantit pas que des résultats pareils puissent être atteints si nous accroissons la complexité du problème en ajoutant quelques contraintes ou quelques indicateurs supplémentaires. La recherche d'un ou de plusieurs motifs élémentaires du profil d'une variable environnementale, pertinents et à faible coût de paramètres, s'inscrit dans les perspectives de ce travail.

## Chapitre 4

# Application à la conception simultanée par optimisation d'une chaîne de traction hybride

### 4.1 Introduction

Après avoir décrit et testé le processus de synthèse de profil représentatif et compact d'une variable environnementale, nous montrons dans ce chapitre comment ce dernier accroît notablement l'efficacité de l'optimisation en minimisant le coût de calcul facilitant dès lors une approche de conception simultanée par optimisation. Nous montrons également que la démarche de classification présentée au deuxième chapitre permet d'évaluer l'intérêt d'un dispositif spécifique, dédié et optimisé par classe de missions par rapport à un seul dispositif capable de satisfaire l'ensemble des classes. Ces aspects sont illustrés à travers la conception optimale d'une locomotive hybride économe en énergie et respectueuse de l'environnement. Il s'agit plus précisément de concevoir une locomotive Diesel-électrique à partir d'une même locomotive tout Diesel en remplaçant son groupe thermique par un autre de puissance nominale plus réduite et en complétant le manque de puissance par des batteries d'accumulateurs. La conception simultanée de la locomotive vise à optimiser le « coût de possession » annuel intégrant les coûts d'investissement, de maintenance et d'exploitation.

Dans la première section de ce chapitre, nous introduisons les modèles de conception de la locomotive hybride liés aux aspects énergétique, climatique, financier, géométrique et de durée de vie,... Ensuite, nous posons le problème de dimensionnement de la locomotive dans une approche de conception simultanée par optimisation. Nous définissons ainsi l'ensemble des paramètres, des contraintes et des critères pertinents vis-à-vis de la conception. Après avoir validé le choix de la puissance nominale du groupe Diesel, nous procédons à la conception d'une locomotive hybride optimale capable de satisfaire l'ensemble des 7 missions ferroviaires de la locomotive BB63000 à partir de la mission représentative de cette famille de

missions. Le résultat est ensuite comparé à celui obtenu en considérant la concaténation des 7 missions du cahier des charges comme mission dimensionnante. A travers un deuxième exemple de missions ferroviaires, nous montrons qu'une démarche systémique globale allant de la classification des missions jusqu'à la conception simultanée par optimisation en utilisant le processus de synthèse de profil représentatif, peut être considérée comme un outil d'aide à la décision dans une phase amont de la conception. Enfin, nous montrons que la mission représentative est intimement liée à la pertinence des indicateurs de conception et qu'il est absolument nécessaire d'adapter ces derniers en fonction de la complexité du problème traité.

## 4.2 Modélisation de la locomotive hybride

Nous reprenons ici et adaptons les modèles développés dans la thèse de Rockys AKLI [AKL 07].

### 4.2.1 Architecture d'hybridation et gestion énergétique

Nous rappelons au lecteur que l'hybridation de la locomotive Diesel consiste à remplacer son groupe Diesel initial par un autre, de taille plus réduite, et de lui associer des batteries d'accumulateurs (cf. Figure 4.1). Nous considérerons des batteries de type Ni-Cd (Hoppecke, FNC 1502 HR) de capacité nominale  $C = 135$  Ah [HOP]. La gestion d'énergie fait en sorte que le groupe Diesel fournisse la puissance moyenne de la mission et que les batteries fournissent le complément de puissance (cf. Figure 4.2) [EHS 99]. Un bus continu régulé en tension permet de connecter le groupe Diesel et les batteries. Ces dernières sont organisées en modules parallèles d'éléments mis en série et reliés au bus continu à travers des hacheurs réversibles en courant et élévateurs vers le bus [AKL 07], [THI 08].

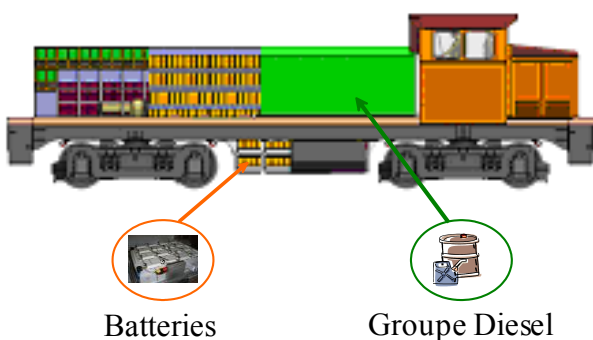


Figure 4.1. La locomotive PLATHEE

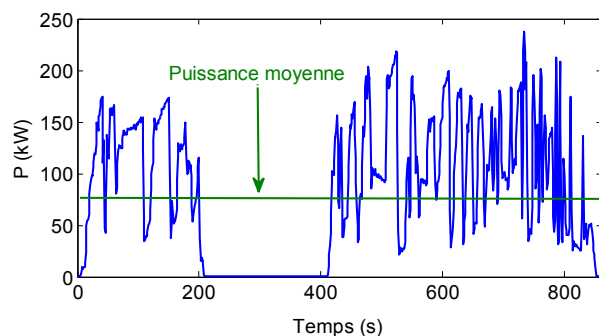


Figure 4.2. Principe de la gestion d'énergie

### 4.2.2 Modèle en flux de puissance

Le modèle en flux de puissance du groupe Diesel est donné à la Figure 4.3. Il permet de déterminer à partir de la puissance de référence du groupe Diesel  $P_{GDref}$  et de sa consigne de marche/arrêt ( $MA_{GD} = 0$  ou 1) la puissance fournie par ce dernier  $P_{GD}$ , la quantité de gasoil consommée  $Q_{fuel}$  et la quantité de  $CO_2$  émise ( $Q_{CO_2}(kg) = 2.66 \times Q_{fuel}(L)$ ) [GRA 06]. La quantité de gasoil consommée est obtenue à partir de la cartographie de consommation du groupe Diesel  $C_s = f(P_{GD-r})$  traduisant la puissance réellement fournie par le groupe Diesel ( $P_{GD-r}$ ) en débit de fuel [FPT].

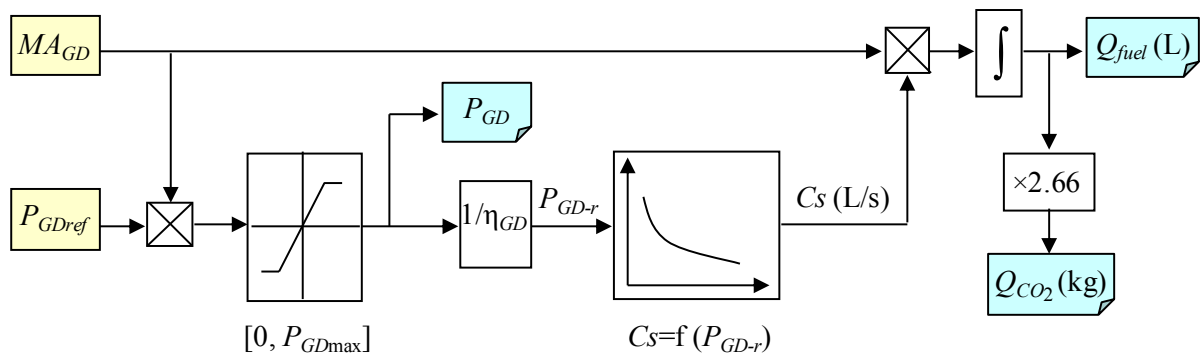


Figure 4.3. Modèle en flux de puissance du groupe Diesel

Le modèle en flux de puissance des batteries (cf. Figure 4.4) admet comme paramètres la mission de référence des batteries ( $P_{BTref} = P_{mission} - P_{GDN}$  où  $P_{GDN}$  désigne la puissance nominale du groupe Diesel), les puissances maximales de décharge  $P_{BTdchmax}$  et de charge  $P_{BTchmax}$  du pack de batteries, le rendement  $\eta_{BT}$  et la quantité d'énergie initiale  $E_{BT0}$ .

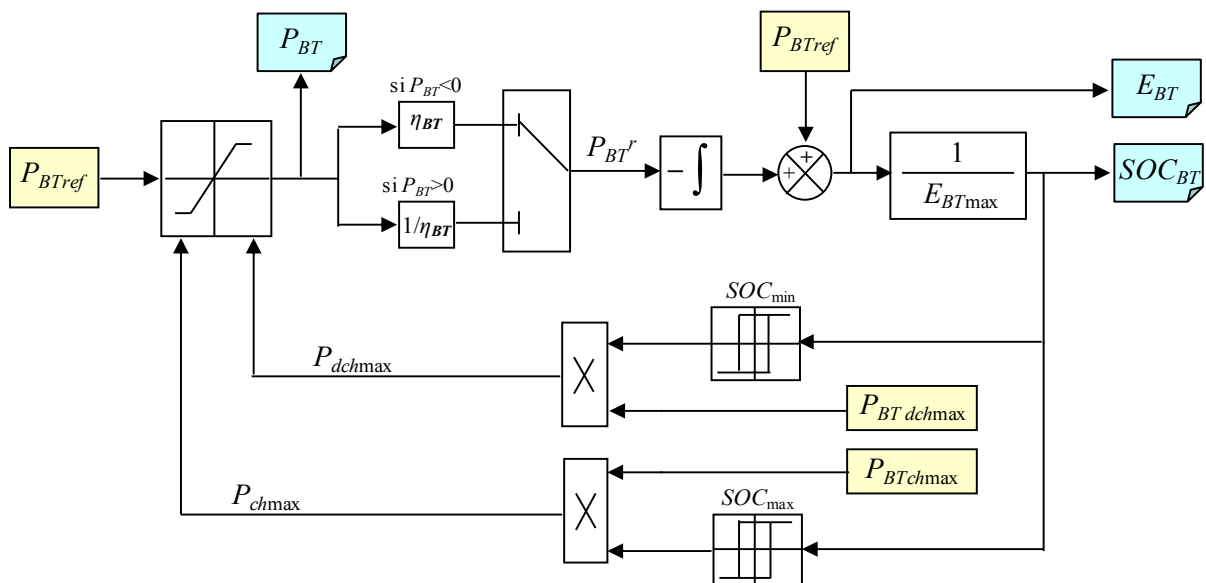


Figure 4.4. Modèle en flux de puissance des batteries

En fixant un seuil minimal d'état de charge des batteries  $SOC_{min}$ , respectivement maximal  $SOC_{max}$ , nous déterminons les puissances maximales de décharge  $P_{dchmax}$ , respectivement de charge  $P_{chmax}$ , réellement admissibles par le pack de batteries en tenant compte, d'une manière instantanée, de son état de charge. Par exemple, lorsque l'état de charge des batteries atteint le seuil  $SOC_{min}$ , la puissance maximale de décharge admissible  $P_{dchmax}$  passe de  $P_{BTdchmax}$  à 0. Dit autrement, nous inhibons complètement la décharge des batteries lorsque le seuil  $SOC_{min}$  est atteint. En sortie, ce modèle en flux de puissance permet de déterminer la puissance des batteries  $P_{BT}$ , leur énergie  $E_{BT}$  ainsi que leur état de charge  $SOC_{BT}$ .

### 4.2.3 Modèle électrique statique

Le modèle électrique statique permet de traduire les grandeurs issues du modèle en flux de puissance (puissance, énergie) en grandeurs électriques (courant, tension). Afin de ne pas compliquer le problème de conception, le modèle électrique du groupe Diesel n'est pas pris en compte. La Figure 4.5 donne le modèle électrique statique des batteries d'accumulateur. Il s'agit d'un modèle simplifié de type  $R, E$ . Le courant  $i_{BT}$  et la tension  $v_{BT}$  sont obtenus à partir de la puissance ramenée à l'échelle d'une seule cellule  $p_{BT}$  ( $p_{BT} = P_{BT}/(NP_{BT} \times NS_{BT})$ ) où  $NP_{BT}$  et  $NS_{BT}$  sont respectivement le nombre de cellules de batteries en parallèle et en série.

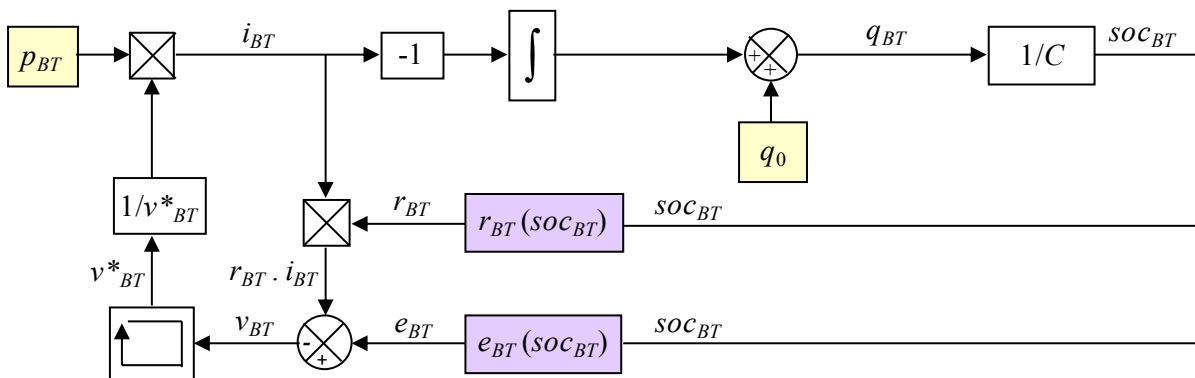


Figure 4.5. Modèle électrique statique des batteries

L'interpolation des points de mesures fournis par le fabricant a permis d'établir l'évolution de la résistance de la batterie  $r_{BT}$  et de sa tension  $e_{BT}$  en fonction de l'état de charge  $SOC_{BT}$  de la cellule selon l'équation (4.1) [AKL 08]. L'état de charge d'une cellule de batterie est obtenu par le rapport de la quantité de charge  $q_{BT}$  de la cellule par sa capacité nominale  $C$  ( $C = 135$  Ah). La quantité de charge  $q_{BT}$  est définie par la somme de l'intégrale du courant  $i_{BT}$

et de la quantité de charge initiale de la cellule ( $q_0$ ). Nous signalons que selon la convention de signe, le courant de décharge des batteries est considéré positif.

$$\begin{cases} e_{BT} = 2.83 - 12.88 soc_{BT} + 24.88 (soc_{BT})^2 - 20.83 (soc_{BT})^3 + 6.28 (soc_{BT})^4 \\ r_{BT} = 0.99 + 1.06 soc_{BT} - 1.82 (soc_{BT})^2 + 1.11 (soc_{BT})^3 \end{cases} \quad (4.1)$$

#### 4.2.4 Modèle géométrique

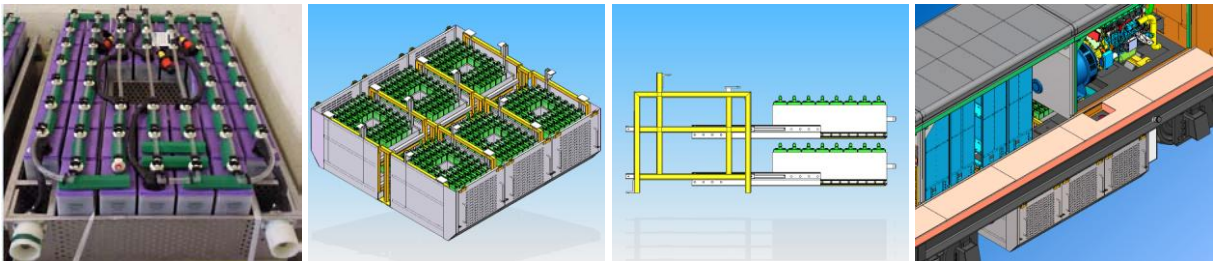
A partir des données fournies par les constructeurs, le volume du groupe Diesel en  $m^3$  peut être exprimé en fonction de sa puissance nominale  $P_{GDN}$  en Watt par la relation suivante :

$$\Omega_{GD} = 3 \times 10^{-5} P_{GDN} + 0.09 \quad (4.2)$$

Le volume des batteries, en tenant compte des systèmes de refroidissement et des modules électroniques associés (équilibre des états de charge, contrôle thermique,...), est donné par la relation suivante :

$$\Omega_{BT} = \lambda_{BT} \times NP_{BT} \times NS_{BT} \times \Omega_{BT0} \quad (4.3)$$

avec  $\Omega_{BT0} = 4.33 \times 10^{-3} m^3$  et où  $\lambda_{BT} = 1.8$  désigne le coefficient d'assemblage qui tient compte des espaces inter-cellules, des modules électroniques et des systèmes de refroidissement. La [Figure 4.6](#) est une illustration de l'assemblage de 12 tiroirs batteries sur la locomotive PLATHEE, contenant chacun 48 éléments de batterie en série.



**Figure 4.6.** Illustration de l'assemblage des éléments de batterie sur la locomotive PLATHEE

Le volume des convertisseurs batteries  $\Omega_{CBT}$  est fortement lié au volume des inductances de filtrage  $\Omega_{LBT}$ . Nous disposons de  $NP_{BT}$  convertisseurs dont chacun embarque  $n_b$  bras de hacheur. Le nombre total des inductances vaut alors  $n_b \times NP_{BT}$ .

Soit  $\Delta i_b$  l'ondulation de courant maximale par bras de hacheur,  $V_{bus}$  la tension du bus DC et  $F_{dec}$  la fréquence de découpage des convertisseurs. L'expression de chaque inductance de bras  $L_{BT}$  s'écrit :

$$L_{BT} = \frac{V_{bus}}{4F_{dec}\Delta i_b} \quad (4.4)$$

Soit  $i_{BT}$  le courant d'un seul bloc de batteries ( $NS_{BT}$  cellules en série). Le nombre de bras  $n_b$  est choisi comme étant le plus petit entier permettant d'avoir un courant traversant les IGBTs inférieur ou égal à 450 A.

$$\frac{i_{BTmax}}{n_b} + \frac{\Delta i_b}{2} \leq 450 \quad (4.5)$$

L'énergie maximale traversant chaque inductance vaut :

$$W_{max} = \frac{1}{2} L_{BT} \left( \frac{i_{BTmax}}{n_b} + \frac{\Delta i_b}{2} \right)^2 \quad (4.6)$$

Le volume d'une inductance est alors calculé à partir de  $W_{max}$  par le polynôme suivant :

$$\Omega_{LBT} = a_0 + a_1 W_{max} + a_2 W_{max}^2 \quad (4.7)$$

avec  $a_0 = 1.3392$ ,  $a_1 = 0.1549$ ,  $a_2 = 0.0014$ .

Le volume global des convertisseurs batteries s'écrit alors :

$$\Omega_{CBT} = \lambda_c n_b NP_{BT} \Omega_{LBT} \quad (4.8)$$

avec  $\lambda_c$  désigne un coefficient de correction tenant compte du volume des IGBTs et de l'électronique associée ( $\lambda_c = 1.2$ ) [AKL 08].

#### 4.2.5 Modèle de durée de vie des batteries

La durée de vie des batteries  $DV_{BT}$  est traduite par le nombre de cycles–batteries consommés au cours de la mission. Elle est exprimée par le produit du nombre de cycles<sup>19</sup>  $Nb_{cycles}$  consommés au niveau de la cellule par le nombre de batteries ( $NP_{BT} \times NS_{BT}$ ).

<sup>19</sup> Nous rappelons au lecteur que le calcul du nombre de cycle a été introduit au premier chapitre.

$$DV_{BT} = NP_{BT} \times NS_{BT} \times Nb_{cycles} \quad (4.9)$$

#### 4.2.6 Modèle de coût

L'estimation du coût du groupe Diesel en fonction de sa puissance nominale est donnée par la relation empirique suivante :

$$C_{GD} [\text{k€}] = 0.28 \times P_{GDN} + 14.5 \quad (4.10)$$

D'après l'expérience de la SNCF, en tenant compte du coût de maintenance et d'entretien, un groupe Diesel coûte trois fois son prix d'achat au bout de 10 ans. La relation précédente s'écrit alors :

$$C_{GD} [\text{k€/an}] = \frac{3}{10} (0.28 P_{GDN} + 14.5) = 0.084 P_{GDN} + 4.35 \quad (4.11)$$

Pour les batteries, en tenant compte des frais d'installation, des convertisseurs et de la maintenance, le coût d'un cycle–batterie Ni-Cd est estimé à 0.122 € [AKL 08]. En utilisant l'indicateur de durée de vie  $DV_{BT}$  évalué pour une mission typique de durée  $\Delta\tau$  et qui se répète  $\Delta\tau_{an}/\Delta\tau$  fois ( $\Delta\tau_{an}$  désigne une année d'exploitation typiquement 8 heures/jour), le coût annuel des batteries est exprimé par la relation suivante :

$$C_{BT} [\text{k€/an}] = 0.122 \times 10^{-3} \times DV_{BT} \times \frac{\Delta\tau_{an}}{\Delta\tau} \quad (4.12)$$

Soit  $C_{fuel0}$  le prix unitaire d'un litre de gasoil<sup>20</sup> en k€. Le coût de gasoil annuel s'écrit alors :

$$C_{fuel} [\text{k€/an}] = C_{fuel0} \times Q_{fuel} \times \frac{\Delta\tau_{an}}{\Delta\tau} \quad (4.13)$$

### 4.3 Modèle d'optimisation

Comme nous l'avons exposé au [chapitre 1](#), bien que les choix effectués à chaque étape de la démarche de conception systémique séquentielle soient soumis à une réflexion d'optimisation locale, rien ne garantit que le système conçu soit optimal dans sa globalité. En effet, plusieurs couplages peuvent exister entre les paramètres de conception liés aux

<sup>20</sup> D'après la SNCF, le prix d'un litre de gasoil est estimé à 1.35 € en 2008



dimensions du système voire à son architecture, d'une part, et les paramètres liés à la gestion du système, d'autre part. Il est alors plus judicieux de suivre une démarche de conception systémique simultanée par optimisation qui vise en général, à déterminer parallèlement (simultanément) l'architecture, le dimensionnement et les paramètres de la gestion énergétique.

#### 4.3.1 Processus de conception simultanée par optimisation

L'algorithme d'optimisation utilisé au sein du processus de conception simultanée est à base d'un algorithme génétique. Ce dernier est couplé aux différents modèles de la locomotive hybride présentés au début de ce chapitre (cf. Figure 4.7). Chaque configuration des paramètres de conception générés par l'algorithme génétique, est évaluée par une simulation des modèles de la locomotive en tenant compte de la mission « dimensionnante ». La solution optimale est celle qui respecte les contraintes de conception (encombrement, satisfaction de la mission) et minimise d'une part le coût global de possession du système intégrant, le coût d'investissement et le coût d'exploitation (maintenance, durée de vie des batteries et consommation de gasoil) et d'une autre part la quantité de CO<sub>2</sub> rejetée.

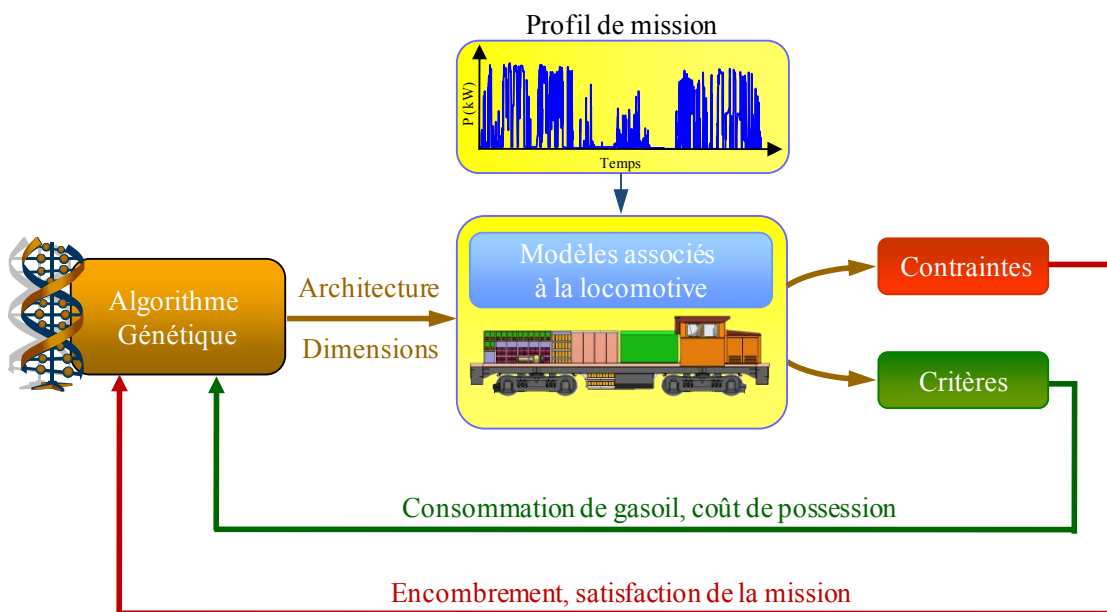


Figure 4.7. Processus de conception simultanée par optimisation

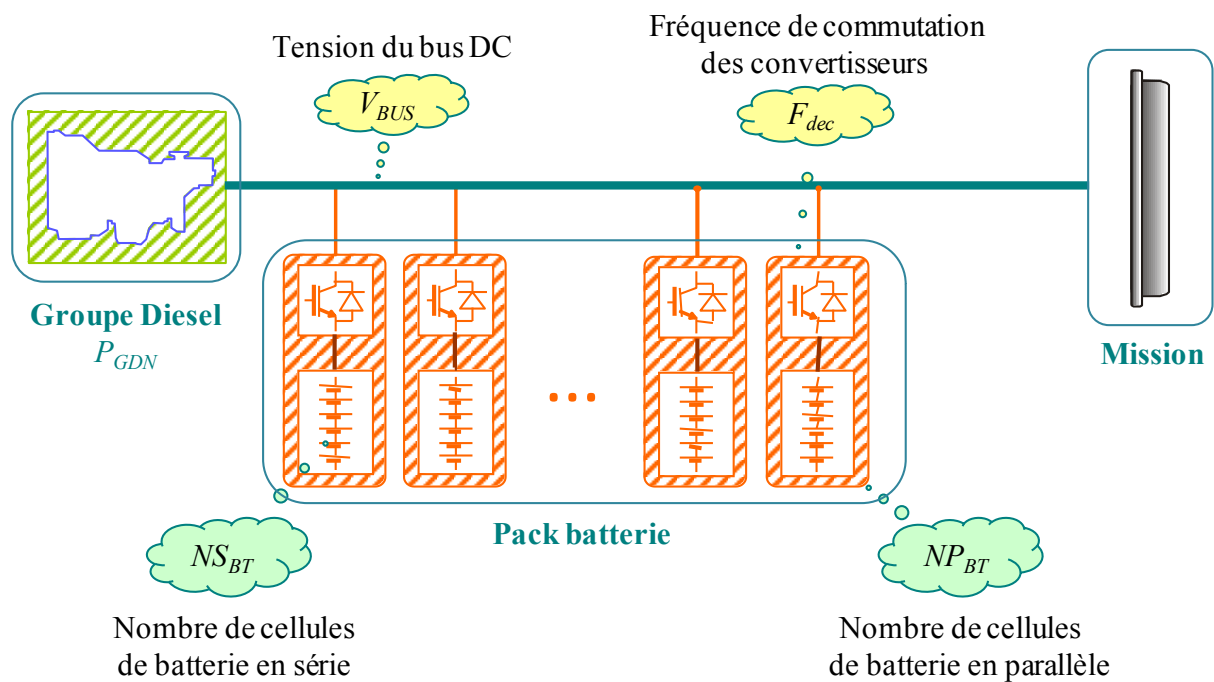
#### 4.3.2 Paramètres d'optimisation

Au niveau de cette étude, nous considérons comme fixe la taille de la source d'énergie principale. Autrement dit, la puissance nominale du groupe Diesel  $P_{GDN}$  n'est pas considérée

comme un paramètre d'optimisation. Elle est fixée au préalable en fonction des caractéristiques des missions de circulation réelles de la locomotive : notons que cette démarche et cette hypothèse ( $P_{GDN}$  constante) permettent de rompre le couplage qui existe entre mission et dimensions du système. Les paramètres d'optimisation nécessaires au dimensionnement de la locomotive hybride sont alors :

- $NS_{BT}$  : le nombre de batteries en série. Compte tenu de la tension bus maximale ( $V_{busmax} = 650$  V), des valeurs maximales du rapport cyclique des convertisseurs statiques ( $\alpha_{max} = 0.94$ ) et de la tension d'une cellule de batterie ( $v_{BTmax} = 1.34$  V), ce nombre est limité à 456 cellules ( $\alpha_{max} \times V_{bus} / v_{BTmax}$ ).
- $NP_{BT}$  : le nombre de blocs batteries en parallèle. Compte tenu de la contrainte de volume, ce nombre est limité à 32 blocs.
- $V_{bus}$  : la tension bus continu variant entre 50 V et 650 V.
- $F_{dec}$  : la fréquence de découpage variant entre 1 kHz et 10 kHz.

Les nombres de batteries en séries  $NS_{BT}$  ou en parallèle  $NP_{BT}$  sont considérés comme des variables discrètes tandis que la tension bus  $V_{bus}$  et la fréquence de découpage  $F_{dec}$  sont des variables continues. Une illustration de l'ensemble de ces paramètres est donnée par la Figure 4.8.



**Figure 4.8.** Illustration des paramètres d'optimisation

### 4.3.3 Contraintes d'optimisation

Afin de réduire le temps de calcul du processus de conception simultanée par optimisation, les contraintes sont scindées en deux catégories :

- *Contraintes pré simulation*

Sans avoir besoin de simuler le système au cours de sa mission, ces contraintes sont évaluées *a priori* à partir des paramètres d'une configuration donnée.

- $g_1$  : Contrainte liée au volume des sources d'énergie

Cette contrainte ne tient pas compte du volume des convertisseurs. Sachant que le volume disponible pour les sources d'énergie<sup>21</sup> est limité à  $\Omega_{SEmax}$ ,  $g_1$  s'écrit alors :

$$g_1 = \Omega_{BT} + \Omega_{GD} - \Omega_{SEmax} \leq 0 \quad (4.14)$$

- $g_2$  : Contrainte liée au nombre de batteries minimal

Pour une puissance nominale donnée du groupe Diesel, cette contrainte traduit le fait que le nombre de batteries doit être suffisant pour satisfaire la mission. Son expression est la suivante :

$$g_2 = N_{BTmin} - NP_{BT} \times NS_{BT} \leq 0 \quad (4.15)$$

Le nombre minimal de cellules de batteries ( $N_{BTmin}$ ) est calculé à partir d'une procédure de dimensionnement des batteries basée sur un modèle en flux de puissance et tenant compte à la fois de l'énergie utile du stockage et de la puissance maximale de charge et de décharge qui caractérisent la mission des batteries [AKL 08].

- $g_3$  : Contrainte liée à la tension batterie

Compte tenu de la structure élévatrice des hacheurs (des batteries vers le bus) et du rapport cyclique maximal  $\alpha_{max}$ , la tension maximale d'un bloc de batteries  $V_{BTmax}$  doit être inférieure à  $\alpha_{max} \times V_{bus}$ . Sachant que la tension maximale d'un élément de batterie est de 1.34 V, la contrainte  $g_3$  s'écrit :

$$g_3 = 1.34 \times NS_{BT} - \alpha_{max} V_{bus} \leq 0 \quad (4.16)$$

<sup>21</sup> Par exemple, pour la locomotive BB63000, le volume maximal disponible pour les sources d'énergie  $\Omega_{SEmax}$  est égal à 32 m<sup>3</sup>.

- *Contraintes post simulation*

- $g_4$  : Contrainte liée au volume du système énergétique

Cette contrainte tient compte du volume des convertisseurs associés aux blocs batteries. L'estimation de ce volume se fait à partir du courant d'un bloc de batteries ( $NS_{BT}$  cellules en série) et nécessite donc la simulation du système au cours de sa mission.

$$g_4 = \Omega_{GD} + \Omega_{BT} + \Omega_{CBT} - \Omega_{SEmax} \quad (4.17)$$

#### 4.3.4 Critère d'optimisation

L'objectif de la démarche d'hybridation consiste à minimiser le coût d'investissement et d'exploitation ainsi que le coût climatique (en tonnes CO<sub>2</sub>) par rapport à une locomotive tout Diesel de même puissance. Sachant que la quantité de CO<sub>2</sub> dégagée est proportionnelle à la quantité de gasoil consommée (1 litre de gasoil donne lieu à 2.66 kg de CO<sub>2</sub>), l'optimisation du coût climatique peut alors se traduire par l'optimisation du coût de la quantité de gasoil consommée [AKL 09]. Ceci nous permet de réduire le problème d'optimisation à un seul critère défini par le coût de possession annuel ( $C_f$ ). Il s'agit alors de minimiser le coût annuel intégrant à la fois le coût d'investissement et le coût d'exploitation (maintenance, durée de vie et consommation de gasoil).

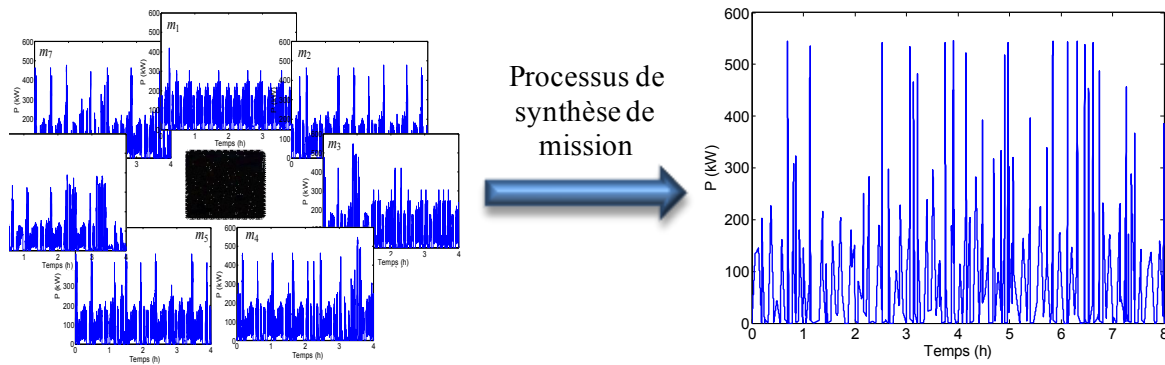
$$C_f = C_{GD} + C_{BT} + C_{fuel} \quad (4.18)$$

#### 4.4 Conception simultanée par optimisation de la locomotive hybride BB63000

Après avoir présenté les différents modèles associés à la locomotive ainsi que le processus de conception simultanée par optimisation, nous procédons dans cette partie au dimensionnement d'une locomotive hybride capable de satisfaire la famille des 7 missions de la locomotive BB63000 décrites au chapitre 3. Nous rappelons que la puissance nominale du groupe Diesel est fixée à 70 kW. Dans l'objectif de vérifier l'efficacité du processus de synthèse de mission représentative dans une démarche de conception par optimisation, nous dimensionnons la locomotive hybride, d'une part avec la mission représentative de l'ensemble des 7 missions ( $M_{rep}$ ) et d'une autre part en considérant toutes les missions du cahier des charges ( $M_{totale}$ ). Dans le premier cas, il s'agit de la mission fictive, de durée 8 heures, issue

du processus de synthèse de mission représentative du [chapitre 3](#) (cf. [Figure 4.9](#)). Dans le deuxième cas, la mission dimensionnante  $M_{totale}$ , de durée 28 heures, est obtenue par la concaténation des 7 missions de circulation réelles, mais en considérant que l'état de charge des batteries est ramené à sa valeur initiale après chaque mission de durée 4 heures.

Nous rappelons au lecteur que, dans cette étude, la conception simultanée est ici ramenée à un problème d'optimisation monocritère. En effet, le critère d'erreur est défini par le coût de possession annuel  $C_f$ , intégrant à la fois le coût d'investissement et le coût d'exploitation. Nous proposons d'utiliser le « Clearing » (CLR) pour dimensionner les éléments de la locomotive. Sur la base de 8 heures d'exploitation de la locomotive par jour (2920 heures par an), les résultats issus du processus de conception simultanée par optimisation sont résumés dans le [Tableau 4.1](#).



**Figure 4.9.** Rappel du résultat du processus de synthèse de la mission représentative  $M_{rep}$

Variables	Conception avec $M_{totale}$	Conception avec $M_{rep}$
$NP_{BT}$	3	3
$NS_{BT}$	455	441
$N_{BT} = NP_{BT} \times NS_{BT}$	1365	1323
$V_{bus}$ (V)	649	650
$F_{dec}$ (kHz)	4.7	4.5
$C_f$ (k€/an)	72.8	73.8
Temps de calcul/configuration (s)	7.9	2.2

**Tableau 4.1.** Résultat de conception avec les missions  $M_{rep}$  et  $M_{totale}$

La conception avec  $M_{totale}$  donne lieu à 1365 cellules de batteries, réparties en trois blocs parallèles, nécessaires pour satisfaire les 7 missions du cahier des charges. Dans le cas de la conception avec  $M_{rep}$ , nous obtenons 1323 cellules de batteries réparties identiquement en trois blocs. Soit une différence de 42 cellules de batteries ce qui reste négligeable vis-à-vis du

nombre de batteries embarquées. Les valeurs de la tension de bus et de la fréquence de découpage sont aussi très proches dans les deux cas.

En ce qui concerne le coût de possession annuel, nous obtenons une différence relativement faible de 1 k€ (~1.4%) par an. Même si les erreurs sur les indicateurs de caractérisation de la mission représentative sont aussi faibles<sup>22</sup> (cf. [Tableau 3.9](#)), le fait de simuler une année de mission ( $M_{rep}$  de durée 8 heures se répète 365 fois par an) induit un cumul d'erreur sur les coûts d'exploitation.

En fait, la différence majeure entre les deux approches se situe surtout en termes de temps de calcul : en effet, pour simuler une configuration donnée, la conception avec  $M_{rep}$  est 3.6 fois plus rapide que la conception avec  $M_{totale}$ . Ceci représente un gain de 5.7 s de temps de calcul par configuration testée sur un PC standard (*Core Duo* 2 GHz). Les résultats d'optimisation du [Tableau 4.1](#) sont obtenus avec une population de 100 individus et 500 générations. Il s'agit alors de 50 000 configurations testées pour lesquelles l'optimisation avec la mission  $M_{rep}$  nous permet de gagner 3.3 jours de temps de calcul. Nous montrons ainsi l'efficacité du processus de synthèse de mission représentative et compacte qui permet de réduire notablement le temps de calcul dans le processus de conception simultanée par optimisation [[JAA 10c](#)]. Nous présentons au niveau du [Tableau 4.2](#) les détails du coût de possession annuel.

Coûts	Conception avec $M_{totale}$	Conception avec $M_{rep}$
$C_{GD}$ (k€/an)	10.2	10.2
$C_{BT}$ (k€/an)	0.85	0.8
$C_{CVS}$ (k€/an)	0.78	0.78
$C_{fuel}$ (k€/an)	60.9	62
$C_f$ (k€/an)	72.8	73.8

**Tableau 4.2.** *Détails du coût de possession annuel*

Nous constatons que le coût annuel de gasoil est largement dominant pour cette famille de missions. En effet, il représente 84% du coût de possession annuel. En outre, nous signalons que l'écart sur ce coût de possession est dû principalement à la différence du coût annuel de gasoil et donc à la différence de la quantité de gasoil consommée entre les deux approches. Même si cet écart est faible, nous lui accordons dans la sous-section suivante, une explication particulière. Nous verrons plus tard que le phénomène responsable de cet écart peut produire

<sup>22</sup> Résultats du [paragraphe 3.5.1](#) du chapitre précédent

des écarts beaucoup plus significatifs. Enfin, le faible coût annuel des batteries nécessitera lui aussi des éléments d'éclaircissement.

#### 4.4.1 Analyse de la différence sur le coût annuel de gasoil

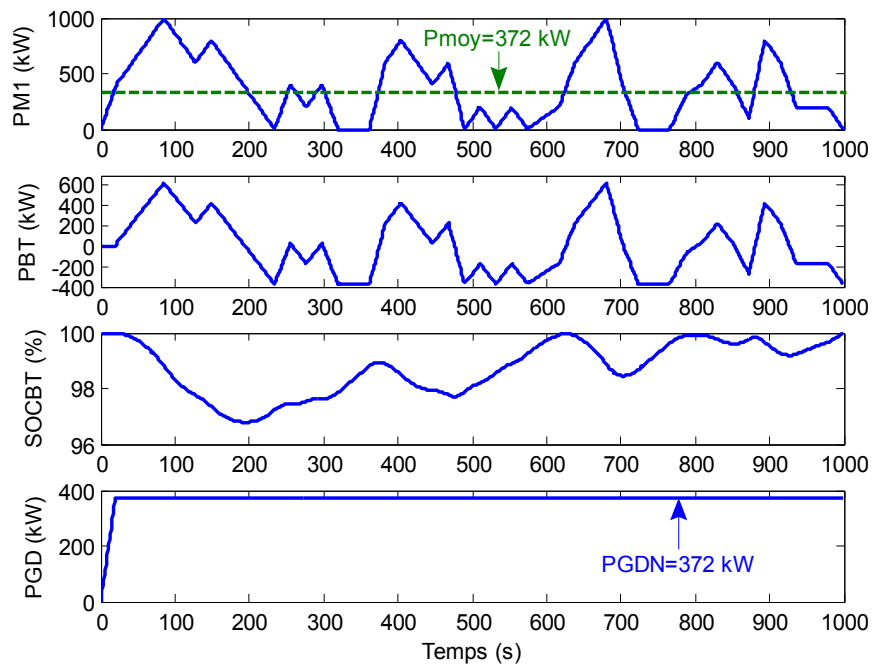
Une analyse de la consommation du groupe Diesel a montré que cette dernière dépend fortement du couplage existant entre la mission et la stratégie de gestion d'énergie mise en œuvre. En effet, le pilotage de marche/arrêt du groupe Diesel est conditionné par l'état de charge des batteries qui dépend fortement de la forme et de la variabilité de la mission. A titre d'exemple, dans le cas où la puissance de la mission est nulle (aucun effort de traction n'est appliqué), le groupe Diesel assure la recharge des batteries tant que leur état de charge est inférieur à un certain seuil.

Afin de mettre en évidence le phénomène d'écart de consommation entre la mission représentative  $M_{rep}$  et la mission totale  $M_{totale}$ , nous considérons deux missions  $M_1$  et  $M_2$  iso-durées caractérisées par les mêmes indicateurs de dimensionnement ( $P_{max}$ ,  $P_{moy}$ ,  $E_u$ ) et de même statistique (même fonction de répartition) mais de formes différentes. Soit une hybridation des missions  $M_1$  et  $M_2$  par un groupe Diesel de puissance nominale  $P_{moy} = 372$  kW et un nombre  $N_{BT}$  de batteries Ni-Cd suffisant pour satisfaire chacune des deux missions. La [Figure 4.10](#) et la [Figure 4.11](#) donnent respectivement, pour les deux missions  $M_1$  et  $M_2$ , les résultats de simulation suivants : la mission des batteries  $P_{BT}$ , leur état de charge  $SOC_{BT}$  et la puissance du groupe Diesel  $P_{GD}$ . Ces résultats montrent que le groupe Diesel ne suit pas la même mission dans les deux cas. En effet, lorsque les batteries sont complètement chargées ( $SOC_{BT} = 100\%$ ), la mission des batteries  $P_{BT}$  demeure nulle ce qui provoque l'arrêt du groupe Diesel comme c'est le cas de la mission  $M_2$  (cf. [Figure 4.11](#)).

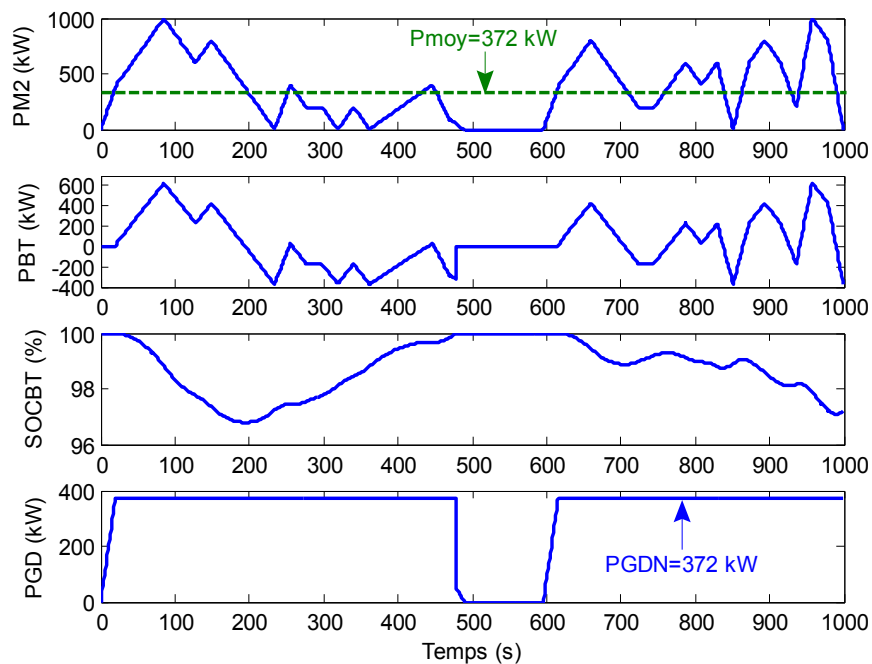
La [Figure 4.12](#) donne l'évolution de la consommation de gasoil pour les deux missions  $M_1$  et  $M_2$ . Nous constatons alors qu'un décalage entre les quantités de gasoil consommées au cours des missions est introduit par la phase d'arrêt du groupe Diesel.

Nous signalons que le phénomène de couplage entre la consommation de gasoil d'une part et la gestion énergétique d'autre part est susceptible de se produire avec la mission représentative  $M_{rep}$  comme avec l'ensemble des missions réelles. La [Figure 4.13](#) montre l'existence des zones de saturation de l'état de charge des batteries ( $SOC_{BT} = 100\%$ ) avec la

mission représentative, ce qui correspond à un état d'arrêt du groupe Diesel. Ceci nous permet alors d'expliquer l'écart du coût annuel de gasoil entre les deux approches de conception.



**Figure 4.10.** Résultats de simulation de la locomotive hybride avec la mission  $M_1$

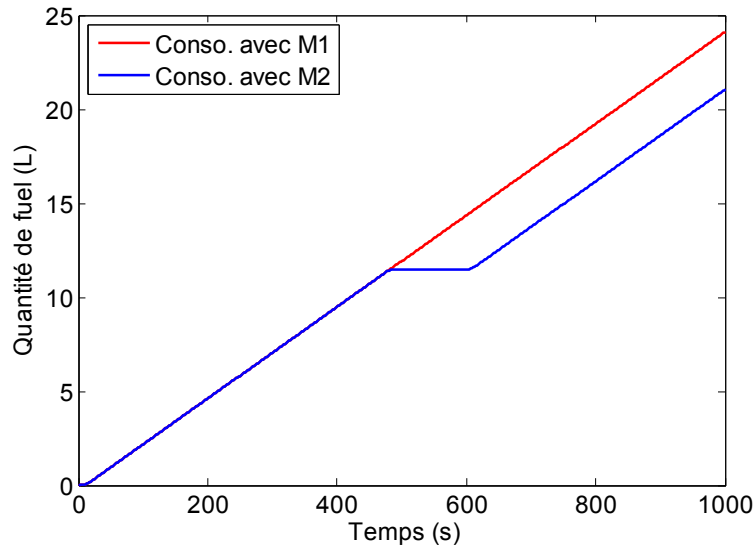


**Figure 4.11.** Résultats de simulation de la locomotive hybride avec la mission  $M_2$

En conclusion, même si la mission représentative respecte parfaitement les indicateurs de référence, rien ne garantit que la quantité de gasoil consommée avec cette dernière soit égale à celle consommée avec l'ensemble des missions réelles sur le même horizon temporel.



Néanmoins, ce phénomène n'influence pas le dimensionnement de la locomotive et le coût annuel de gasoil évalué à partir de la mission représentative reste une bonne approximation du coût de la consommation réelle.



**Figure 4.12.** Évolution de la consommation de gasoil pour les missions  $M_1$  et  $M_2$

#### 4.4.2 Analyse du coût annuel des batteries

Le coût annuel des batteries est pratiquement identique pour les deux approches de conception (cf. [Tableau 4.2](#)), mais son poids dans le coût de possession annuel de la locomotive semble être très faible. Nous donnons ci-dessous une explication de cette constatation.

Pour la famille de missions ainsi étudiée, le nombre de batteries minimal  $N_{BTmin}$  à respecter avec un groupe Diesel de 70 kW vaut 1176 cellules. Dans ce cas, le dimensionnement des batteries Ni-Cd est contraint par la puissance maximale de décharge<sup>23</sup>. Autrement dit, les batteries sont dimensionnées en puissance et sont alors surdimensionnées sur le plan énergétique. En effet, 1176 cellules de batteries embarquent 191 kWh alors que l'énergie utile maximale pour satisfaire toutes les missions est seulement de 36.4 kWh, l'équivalent de 225 cellules de batteries. Par la suite, les profondeurs de décharges du pack de batteries minimal (1176 cellules) sont très faibles.

Pour la conception avec la mission représentative  $M_{rep}$ , le nombre de batteries obtenu est égal à 1323 ( $3 \times 441$ ). Il en découle alors des phases de décharge encore moins profondes. La

<sup>23</sup> La puissance maximale de décharge des batteries est définie par :  $\max(P_{BT}) = \max(P_{mission} - P_{GDN})$

Figure 4.13 donnant l'état de charge des batteries dans le cas de la conception avec la mission représentative  $M_{rep}$ , montre que la profondeur de décharge maximale ne dépasse pas 17% de la capacité de stockage totale des 1323 cellules de batteries. Sachant que les poids des cycles peu profonds sur la durée de vie sont beaucoup plus faibles que ceux des grands cycles<sup>24</sup> : le nombre de cycles équivalent ne dépasse pas quelques cycles complets. Par exemple, un cycle de profondeur 10% ne représente que 0.45% d'un cycle complet. Autrement dit, il faut 223 cycles de profondeur 10% pour solliciter une cellule de batterie sur l'équivalent d'un cycle complet. En effet, avec la mission  $M_{rep}$  réalisée 365 fois pendant une année, chaque cellule de batteries est sollicitée seulement sur l'équivalent de 4.9 cycles complets. Il en résulte alors un faible coût annuel des batteries défini comme étant le produit du coût de 4.9 cycles par 1323 cellules de batteries ( $C_{BT} = 0.8$  k€/an). Comme conséquence, le coût annuel de la locomotive, pour cette famille de missions, est quasiment insensible au coût annuel des batteries.

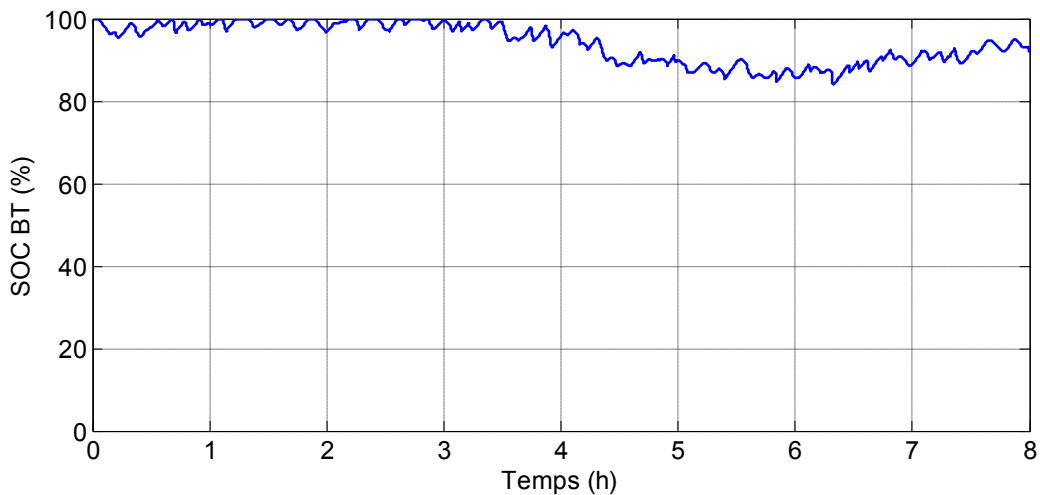


Figure 4.13. Etat de charge des batteries pour la mission représentative  $M_{rep}$

Que ce soit pour la mission représentative  $M_{rep}$  ou avec l'ensemble des missions réelles  $M_{totale}$ , nous constatons que l'optimiseur a tendance à augmenter légèrement le nombre de batteries par rapport au nombre minimal  $N_{BTmin}$ . Cette croissance a pour objectif d'amplifier d'avantage la capacité de stockage totale du pack de batteries et par la suite de réduire les profondeurs des cycles. Néanmoins, pour cet exemple de missions, la réduction des profondeurs des cycles n'est pas trop significative et le coût annuel des batteries sera beaucoup plus sensible à leur nombre. En outre, l'augmentation du nombre de cellules Ni-Cd

<sup>24</sup> Le nombre de cycles disponible en fonction de la profondeur de décharge est une fonction non linéaire fournie par le constructeur.

peut accroître le nombre de packs de batteries en parallèle  $NP_{BT}$  et donc le nombre de convertisseurs statiques, ce qui introduit des coûts supplémentaires au niveau du coût de possession annuel. La tendance de l'optimiseur à augmenter le nombre de batteries est alors régie par un compromis qui dépend d'une part de la mission du stockage, particulièrement de ses caractéristiques sur le plan énergie/puissance et de sa cyclabilité, et d'autre part de la technologie de stockage considérée (plan de Ragone, densité d'énergie massique et volumique, nombre de cycles admissibles). Nous venons alors de montrer le fort couplage existant entre la mission, le dimensionnement et la technologie de stockage considérée dans un contexte de conception systémique.

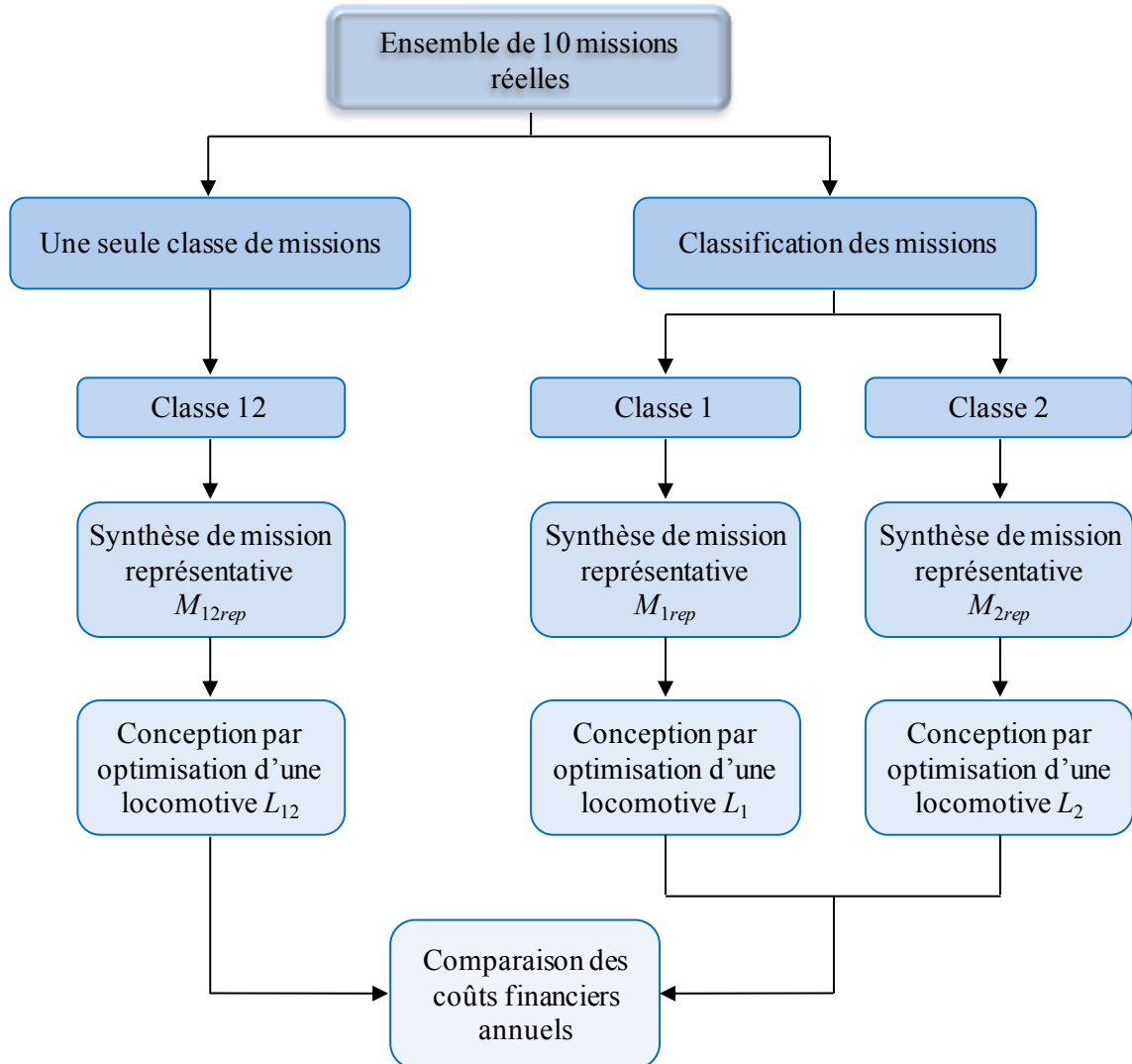
Pour conclure, l'analyse précédente a montré que le coût annuel des batteries dépend fortement de la nature de la mission en termes de cyclabilité. En outre, c'est le contenu énergétique de la mission qui définit le caractère du stockage en tant que source de puissance ou source d'énergie. A travers l'exemple de missions étudié ci-dessus, nous avons montré que dans le cas où les batteries sont dimensionnées en puissance et sont surdimensionnées sur le plan énergétique, les cycles sont peu profonds et par conséquent le coût annuel des batteries reste faible. A noter que cette tendance concorde avec les conclusions qualitatives souvent issues d'études en ferroviaire. Nous verrons dans la section suivante comment l'optimiseur « réagit » pour minimiser le coût annuel des batteries lorsque leur dimensionnement est contraint par l'énergie utile de leur part de mission).

#### **4.5 De la classification à la conception optimale**

Dans cette section, nous proposons, à travers un ensemble de missions ferroviaires, une illustration de la démarche systémique globale allant de la classification des profils d'une variable environnementale (cf. [chapitre 2](#)) jusqu'à la conception simultanée par optimisation en utilisant le processus de synthèse de profil représentatif et compact (cf. [chapitre 3](#)). L'idée est de concevoir une locomotive hybride optimisée pour chaque classe de missions et de comparer le coût de possession annuel des systèmes obtenus par rapport à la conception d'une seule locomotive capable de satisfaire toutes les classes de missions à la fois.

Cette démarche globale de conception systémique est illustrée à travers un exemple de 10 missions ferroviaires iso-durées (4 heures par mission) relatives à des locomotives Diesel. Cette approche revient en quelque sorte à permettre à l'exploitant, ici la SNCF, l'intérêt financier d'une segmentation de sa gamme de produit. Nous donnons à la [Figure 4.14](#) les

différentes étapes de l'étude comparative proposée. Nous signalons au lecteur que la classification de l'ensemble des missions donne lieu à deux classes différentes. Nous revenons dans les parties suivantes sur les détails de la classification des missions.



**Figure 4.14.** Démarche systémique globale de la classification à la conception simultanée par optimisation

Les indicateurs de dimensionnement de toutes les missions sont donnés par le [Tableau 4.3](#). Nous signalons que l'énergie utile de chaque mission est déterminée par rapport à sa propre puissance moyenne.

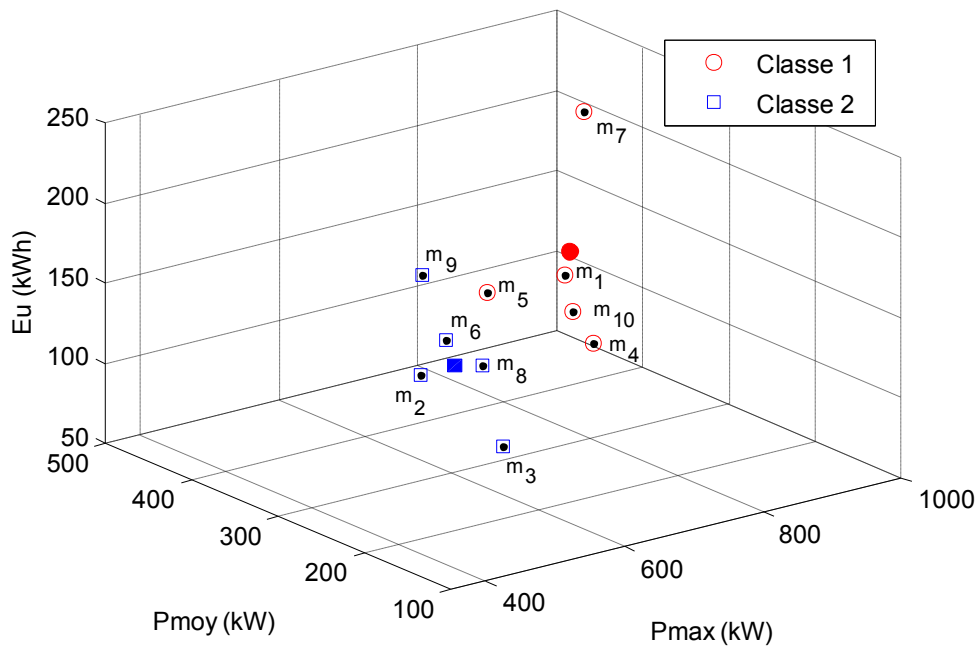
#### 4.5.1 Classification des missions

Dans cette première phase de la démarche systémique globale, nous procédons à la classification de l'ensemble des 10 missions ferroviaires selon les indicateurs de

dimensionnement  $\{P_{\max}, P_{\text{moy}}, E_u\}$ . Le résultat obtenu en appliquant l'algorithme génétique de classification développé au niveau du chapitre 2 est donné par la Figure 4.15. Les caractéristiques des classes trouvées sont également données par le Tableau 4.4.

Missions	$P_{\max}$ (kW)	$P_{\text{moy}}$ (kW)	$E_u$ (kWh)
$m_1$	939.3	441.8	104.9
$m_2$	407.9	153.6	187.4
$m_3$	601.3	241.5	80.1
$m_4$	980.5	442.2	57.3
$m_5$	750	380	128.1
$m_6$	490	248.6	135.4
$m_7$	885.6	376.3	227.6
$m_8$	501.4	185.3	154
$m_9$	396	171.6	224.9
$m_{10}$	850	359.7	110.3

**Tableau 4.3.** Indicateurs de dimensionnement de l'ensemble des missions



**Figure 4.15.** Classification des missions selon les indicateurs de dimensionnement

	Nombre de missions	Missions	Coordonnées du barycentre ( $P_{\max}$ (kW), $P_{\text{moy}}$ (kW), $E_u$ (kWh))
Classe 1 ( $C_1$ )	5	$\{m_1, m_4, m_5, m_7, m_{10}\}$	(881.1, 400, 125.7)
Classe 2 ( $C_2$ )	5	$\{m_2, m_3, m_6, m_8, m_9\}$	(479.3, 200.1, 156.4)

**Tableau 4.4.** Caractéristiques des classes obtenues après la classification

Deux classes de missions ont été identifiées. La première classe  $C_1$  est caractérisée par des puissances maximales et des puissances moyennes importantes ( $750 \text{ kW} \leq P_{\max} \leq 980.5 \text{ kW}$  et  $360 \text{ kW} \leq P_{\text{moy}} \leq 442 \text{ kW}$ ) tandis que la deuxième classe  $C_2$  regroupe les missions à puissances maximales et à puissances moyennes approximativement deux fois plus faibles ( $396 \text{ kW} \leq P_{\max} \leq 602 \text{ kW}$  et  $154 \text{ kW} \leq P_{\text{moy}} \leq 249 \text{ kW}$ ). En revanche l'énergie utile  $E_u$  des barycentres est assez comparable pour ces deux classes. Pour les deux classes de missions ainsi obtenues, l'énergie utile varie entre 57 kWh et 228 kWh.

#### 4.5.2 Synthèse des missions représentatives

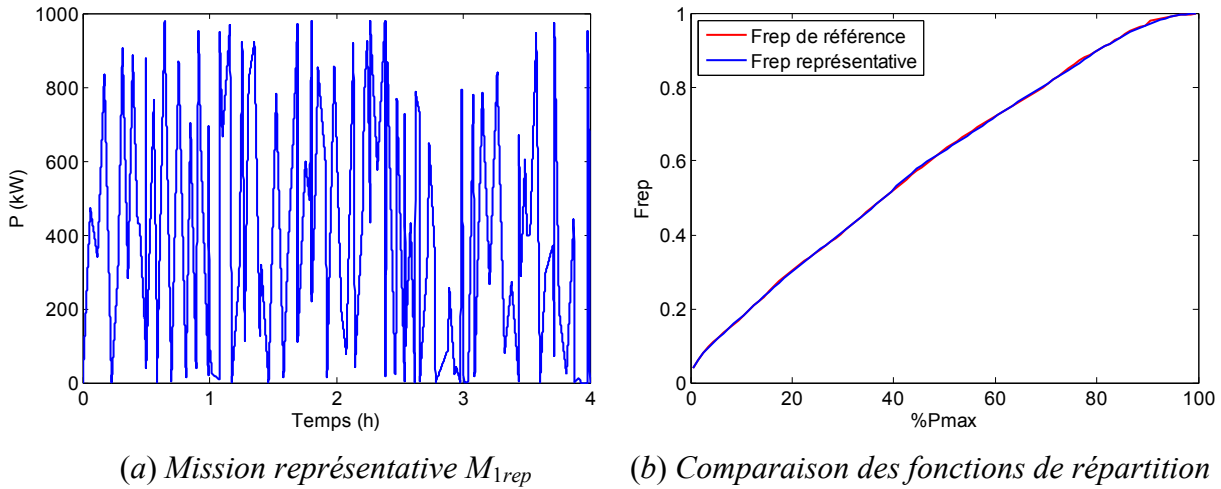
Au niveau de cette deuxième phase d'étude, nous appliquons le processus de synthèse de mission représentative et compacte pour chaque classe de missions ( $C_1$ ,  $C_2$  et  $C_{12}$  formée par l'ensemble de toutes les missions). En ce qui concerne le choix des indicateurs de référence, nous appliquons la stratégie dite « sans dégradation des missions » (cf. [chapitre 3](#)). La puissance moyenne de référence  $P_{\text{moy ref}}$  est déterminée par la valeur moyenne de la mission totale définie par la concaténation de toutes les missions de la classe en question. A titre de rappel, c'est à partir de la puissance moyenne de référence que nous dimensionnons le groupe Diesel de la locomotive hybride dédiée à satisfaire les missions de la classe étudiée. La puissance maximale  $P_{\max \text{ ref}}$  et l'énergie utile  $E_{u \text{ ref}}$  de référence sont fixées aux valeurs les plus contraignantes.

Les meilleurs résultats sont obtenus avec le motif de type segment, couplé au « Clearing » (CLR) comme algorithme génétique. La taille de la population est fixée à 100 individus et le nombre de générations est adapté en fonction du problème traité. Nous pouvons conclure sur la bonne convergence de l'algorithme d'optimisation lorsque le critère d'erreur demeure stable (pas d'amélioration) sur un nombre important de générations fixé *a priori*.

- *Mission représentative de la classe  $C_1$  :  $M_{1\text{rep}}$*

Nous rappelons que la première classe  $C_1$  est formée par 5 missions de durée 4 heures chacune. Compte tenu des contraintes de durées minimales relatives à la satisfaction de l'énergie utile de référence et au respect du nombre de cycles de référence par heure de mission exposées au [chapitre 3](#), nous choisissons de fixer la durée de la mission représentative à 4 heures, ce qui correspond à un gain de facteur 5 sur la durée totale des missions (20 heures). Le résultat de synthèse est donné par la [Figure 4.16.a](#). Il s'agit d'une mission

représentative  $M_{1rep}$  obtenue par la concaténation de 125 segments (251 paramètres) et qui respecte parfaitement les indicateurs de référence (cf. [Tableau 4.5](#) et [Figure 4.16.b](#)).



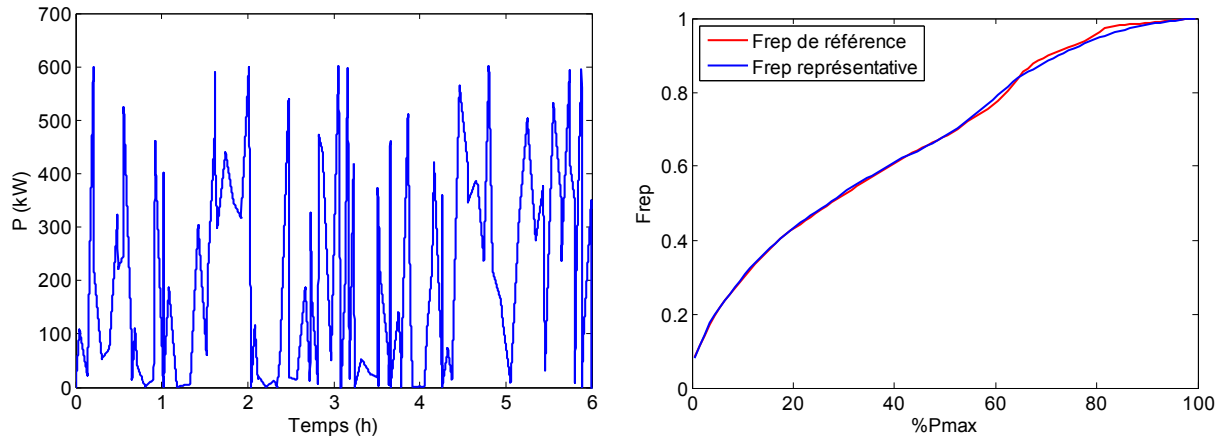
**Figure 4.16.** Résultat du processus de synthèse de la mission représentative  $M_{1rep}$

	Indicateurs de référence	Indicateurs de la mission représentative	Erreur (%)
$P_{max}$ (kW)	980	980	0.0
$P_{moy}$ (kW)	400	400	0.0
$E_u$ (kWh)	169.5	169.5	0.0
$Nb_{cyc}$ (/h)	0.155	0.155	0.0
$I_{stat}$	—	—	0.4

**Tableau 4.5.** Indicateurs de caractérisation de la mission représentative  $M_{1rep}$

- *Mission représentative de la classe  $C_2$  :  $M_{2rep}$*

La deuxième classe de missions est aussi formée par 5 missions de durée 4 heures chacune. En tenant compte de la contrainte de durée minimale à respecter, la durée de la mission représentative est fixée à 6 heures ce qui correspond à un gain de 3.3 sur la durée totale des missions (20 heures). Le résultat de synthèse est représenté par la [Figure 4.17.a](#). Il s'agit d'une mission représentative  $M_{2rep}$  issue de la concaténation de 126 segments (253 paramètres) et dont les indicateurs de caractérisation coïncident parfaitement avec les indicateurs de référence (cf. [Tableau 4.6](#) et [Figure 4.17.b](#)).

(a) Mission représentative  $M_{2rep}$ 

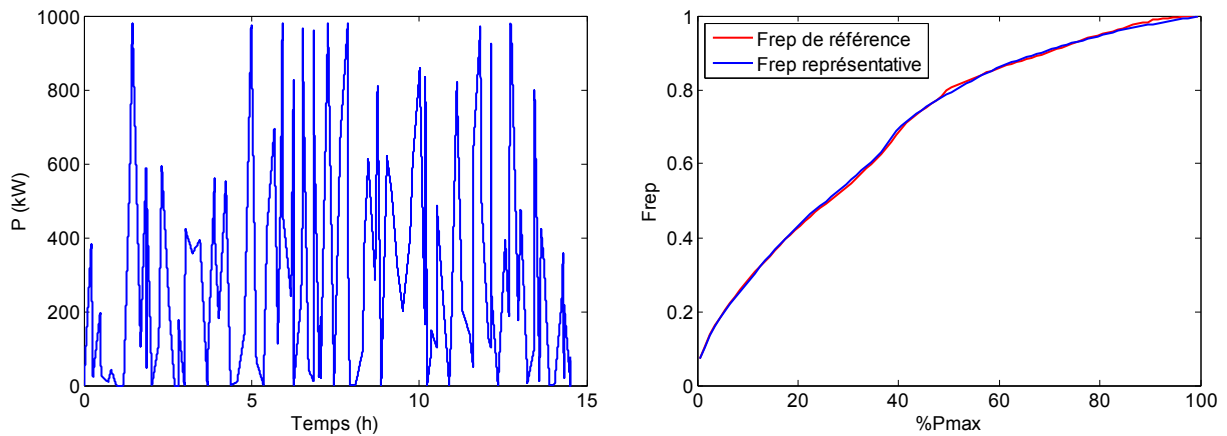
(b) Comparaison des fonctions de répartition

**Figure 4.17.** Résultat du processus de synthèse de la mission représentative  $M_{2rep}$ 

	Indicateurs de référence	Indicateurs de la mission représentative	Erreur (%)
$P_{max}$ (kW)	601	601	0.0
$P_{moy}$ (kW)	200	201	0.5
$E_u$ (kWh)	194.5	194.5	0.0
$Nb_{cyc}$ (/h)	0.11	0.11	0.0
$I_{stat}$	–	–	0.8

**Tableau 4.6.** Indicateurs de caractérisation de la mission représentative  $M_{2rep}$ 

- Mission représentative de la classe  $C_{12}$  :  $M_{12rep}$

(a) Mission représentative  $M_{12rep}$ 

(b) Comparaison des fonctions de répartition

**Figure 4.18.** Résultat du processus de synthèse de la mission représentative  $M_{12rep}$ 

La mission représentative  $M_{12rep}$  de la classe  $C_{12}$  qui intègre simultanément la classe  $C_1$  et la classe  $C_2$  est donnée par la [Figure 4.18.a](#). Il s'agit d'une mission fictive de durée



14.5 heures permettant ainsi un gain de 2.8 par rapport à la durée totale des missions (40 heures). Cette dernière est obtenue par la concaténation de 148 segments (297 paramètres). Une comparaison entre les indicateurs de caractérisation de cette mission et les indicateurs de référence est donnée par le [Tableau 4.7](#) et la [Figure 4.18.b](#).

	Indicateurs de référence	Indicateurs de la mission représentative	Erreur (%)
$P_{\max}$ (kW)	980.5	980.5	0.0
$P_{\text{moy}}$ (kW)	300	300	0.0
$E_u$ (kWh)	569.3	569.3	0.0
$Nb_{\text{cyc}}$ (/h)	0.04	0.04	0.0
$I_{\text{stat}}$	–	–	0.8

**Tableau 4.7.** Indicateurs de caractérisation de la mission représentative  $M_{12\text{rep}}$

### 4.5.3 Conception simultanée par optimisation

Après avoir déterminé les missions représentatives, nous procédons dans ce paragraphe à la conception simultanée de trois locomotives hybrides optimisées par classe de missions. Désignons par  $L_1$ ,  $L_2$  et  $L_{12}$  les locomotives hybrides dimensionnées respectivement pour les classes  $C_1$ ,  $C_2$  et  $C_{12}$  à partir des missions représentatives  $M_{1\text{rep}}$ ,  $M_{2\text{rep}}$  et  $M_{12\text{rep}}$ . Afin de valider la cohérence des résultats de conception et de mettre en évidence l'efficacité du processus de synthèse de mission représentative en termes de temps de calcul, nous dimensionnons aussi trois locomotives hybrides optimisées par classe de missions à partir des missions réelles. Autrement dit, il s'agit de concevoir pour chaque classe de missions, une locomotive hybride en considérant comme mission dimensionnante la concaténation de toutes les missions de la classe étudiée sous l'hypothèse que l'état de charge des batteries sera ramené à sa valeur initiale après chaque mission de durée 4 heures. Ainsi, soit  $L_{1\text{tot}}$ ,  $L_{2\text{tot}}$  et  $L_{12\text{tot}}$  les locomotives hybrides dimensionnées respectivement pour les classes  $C_1$ ,  $C_2$  et  $C_{12}$  à partir des missions réelles dimensionnantes  $M_{1\text{tot}}$ ,  $M_{2\text{tot}}$  et  $M_{12\text{tot}}$ .

Sous l'hypothèse que toutes les missions initiales se réalisent avec la même occurrence pendant une année et sur la base de 8 heures d'exploitation de chaque locomotive par jour, les résultats de conception sont donnés par le [Tableau 4.8](#). Nous donnons au niveau du [Tableau 4.9](#) les détails des coûts de possession annuels pour chaque classe de missions et pour les deux types de missions considérées dans le processus de conception (missions représentatives et missions réelles dimensionnantes).

Variables	classe $C_1$		classe $C_2$		classe $C_{12}$	
	$L_{1tot}$	$L_1$	$L_{2tot}$	$L_2$	$L_{12tot}$	$L_{12}$
$P_{GD}$ (kW)	400	400	200	200	300	300
$NP_{BT}$	10	10	12	12	11	11
$NS_{BT}$	437	421	444	440	429	421
$N_{BT}$	4370	4210	5328	5280	4719	4631
$V_{bus}$ (V)	640	624	636	630	614	602
$F_{dec}$ (kHz)	7.3	8.1	5.4	5.1	4.9	5.7
$C_f$ (k€/an)	397.2	421.8	202.8	191	318.5	350.4
$T_c$ (s) <sup>25</sup>	4.9	1	4.9	1.4	9	3.3

**Tableau 4.8.** *Résultat de conception des locomotives hybrides*

Variables	classe $C_1$		classe $C_2$		classe $C_{12}$	
	$L_{1tot}$	$L_1$	$L_{2tot}$	$L_2$	$L_{12tot}$	$L_{12}$
$C_{GD}$ (k€/an)	38	38	21.1	21.1	29.5	29.5
$C_{BT}$ (k€/an)	13.6	12.7	9.9	9.5	52.5	63.2
$C_{CVS}$ (k€/an)	2.6	2.6	3.12	3.12	2.86	2.86
$C_{fuel}$ (k€/an)	343	368.5	168.7	157.3	233.6	254.7
$C_f$ (k€/an)	397.2	421.8	202.8	191	318.5	350.4

**Tableau 4.9.** *Détails des coûts de possession annuels*

Les résultats de conception obtenus par classe de missions prouvent une fois de plus l'efficacité du processus de synthèse de mission représentative dans la démarche de conception systémique. En effet, nous obtenons pratiquement le même dimensionnement de la locomotive hybride en considérant d'une part la mission représentative fictive et d'une autre part la concaténation des missions réelles de la classe envisagée. Pour les trois classes de missions ainsi étudiées, nous relevons un gain notable en termes de temps de calcul entre les deux approches. En effet, les simulations précédentes sont obtenues au bout de 1000 générations avec une population de 100 individus. Dans ces conditions, la conception avec les missions représentatives nous permet un gain sur le temps d'optimisation qui s'élève jusqu'à 4.5 jours pour la classe  $C_1$ , 4 jours pour la classe  $C_2$  et 6.6 jours pour la classe  $C_{12}$ .

La différence des coûts de possession annuels obtenue entre la conception avec les missions représentatives fictives et celle avec les missions réelles est due principalement aux « erreurs » sur l'estimation du coût annuel de gasoil (cf. [Tableau 4.9](#)). Comme nous l'avons

<sup>25</sup> Temps de calcul par configuration

expliqué précédemment, l'état de marche/arrêt du groupe Diesel dépend de l'état de charge des batteries qui lui-même dépend de la forme de la mission même si tous les indicateurs de caractérisation sont parfaitement respectés. Toutefois, les coûts annuels de gasoil déterminés à partir des missions représentatives procurent une bonne estimation du coût de la consommation réelle. En effet, l'erreur sur le coût annuel de gasoil ne dépasse pas 9% pour les trois classes ainsi étudiées.

En outre, nous signalons que le coût annuel des batteries peut parfois présenter une erreur importante comme c'est le cas de la conception avec la mission représentative  $M_{12rep}$  de la classe  $C_{12}$ . Il s'agit d'une surestimation du coût annuel des batteries de 20% par rapport au coût déterminé avec les missions réelles. Ce taux d'erreur relativement élevé nécessite une analyse particulière. Nous signalons d'abord que pour satisfaire la mission représentative  $M_{12rep}$  ou toutes les missions de la classe  $C_{12}$ , le dimensionnement des batteries est contraint par le caractère énergétique des missions. Autrement dit, les batteries se comportent comme une « source d'énergie » sur la mission. Le nombre de batteries minimal obtenu dans ce cas est égal à 4395 cellules ce qui correspond à une capacité de stockage de 712 kWh et une puissance maximale de décharge de l'ordre de 1780 kW, largement supérieure à la puissance maximale des missions (980.5 kW). D'après les résultats de conception, nous constatons que l'optimiseur a tendance à augmenter le nombre de batteries pour réduire davantage leur coût annuel. En effet, si l'optimiseur a choisi de caler le nombre de batteries sur leur nombre minimal (4395), que ce soit avec la mission  $M_{12rep}$  ou avec les 10 missions réelles  $M_{12tot}$ , le coût annuel des batteries s'élève jusqu'à 68 k€/an. Nous signalons que dans ce cas de figure, les batteries sont sollicitées par un même nombre de cycles complets (127 cycles de 100%) mais pas forcément avec les mêmes profondeurs de cycles<sup>26</sup>. Avec la mission représentative  $M_{12rep}$ , l'optimiseur a augmenté le nombre de batteries de 236 cellules ce qui a provoqué une diminution du nombre de cycles consommés sur une année de missions de 15 cycles complets (112 cycles au lieu de 127 cycles avec 4395 batteries) et par la suite un passage du coût annuel des batteries de 68 k€/an à 63.2 k€/an. En ce qui concerne la conception avec les 10 missions réelles  $M_{12tot}$  de la classe  $C_{12}$ , l'optimiseur a choisi d'augmenter encore le nombre de batteries de 324 cellules ce qui a permis d'améliorer la durée de vie des batteries en réduisant leur sollicitation de 36 cycles complets sur une année d'exploitation. Ceci provoque un gain notable en termes de coût annuel des batteries qui passe de 68 k€/an à 52.5 k€/an.

---

<sup>26</sup> Nous rappelons que l'indicateur du nombre de cycle est un indicateur global

Bien qu'il s'agisse de la même tendance de l'optimiseur avec la mission représentative  $M_{12rep}$  ou avec les 10 missions réelles  $M_{12tot}$ , l'amélioration du coût annuel des batteries n'est pas la même entre les deux approches. En effet, compte tenu du caractère source d'énergie des batteries d'accumulateurs, il est incontestable que nous disposons des cycles de profondeurs importantes mais pas forcément répartis de la même façon. Par exemple en termes de cycles complets, un cycle de profondeur 80% (0.59 cycle complet) est équivalent à deux cycles de profondeur 60% chacun (0.297 cycle complet). Le fait d'augmenter alors le nombre de batteries, donc leur capacité de stockage totale, provoque une diminution des profondeurs des cycles. Par suite, le nombre de cycles complets n'est plus le même entre les deux approches de missions ( $M_{12rep}$  et  $M_{12tot}$ ).

#### 4.5.4 Comparaison des résultats de conception

- *Comparaison à base des missions réelles*

Nous rappelons au lecteur que la procédure de conception des locomotives hybrides avec les missions réelles a été mise en place pour valider notre processus de synthèse de mission représentative dans la démarche de conception simultanée par optimisation. Toutefois, nous profitons des résultats de conception pour établir une comparaison plus réaliste des résultats obtenus. Sur la base de 8 heures d'exploitation par jour de chaque locomotive, les 10 missions initiales de la classe  $C_{12}$  et de durée totale 40 heures sont répétées 73 fois pendant une année. Tandis que les 5 missions de la classe  $C_1$  de durée 20 heures ainsi que les 5 missions de la classe  $C_2$  de même durée, sont répétées respectivement 146 fois par an. Afin d'évaluer l'intérêt de la classification (ou segmentation), nous comparons le coût annuel moyen des deux locomotives  $L_{1tot}$  et  $L_{2tot}$  avec le coût annuel de la locomotive  $L_{12tot}$  :

- Coût moyen annuel de  $L_{1tot}$  et  $L_{2tot} = 0.5 \times (\text{Coût } (C_1)/L_{1tot} + \text{Coût } (C_2)/L_{2tot}) = 300 \text{ k€}/\text{an}$

- Coût annuel de  $L_{12tot} = 318.5 \text{ k€}/\text{an}$

Nous constatons que pour cet exemple de 10 missions, le fait de concevoir deux locomotives optimisées par classe de missions nous permet un gain de 18.5 k€/an (~6%/an) par rapport à une seule locomotive capable de satisfaire toutes les missions du cahier des charges. En revanche, bien qu'elle soit moins rentable sur le plan financier, la solution avec une seule locomotive équivalente  $L_{12tot}$  est plus avantageuse sur le plan écologique. En effet, l'exploitation annuelle de cette locomotive engendre environ 460 tonnes de  $\text{CO}_2$  contre 504 tonnes de  $\text{CO}_2$  dégagées lors d'une exploitation moyenne annuelle des deux locomotives

$L_{1tot}$  et  $L_{2tot}$ . Le choix de concevoir une locomotive  $L_{12tot}$  unique, ou deux locomotives différentes  $L_{1tot}$  et  $L_{2tot}$  est régi par un compromis sur le plan financier et le plan écologique : au delà du résultat de cette application particulière, l'intérêt de la démarche complète (classification – synthèse mission – conception simultanée) que nous avons proposée est de faciliter les choix du concepteur exploitant, ici la SNCF, en termes de segmentation de la gamme de produit puis de dimensionnement. Ces choix peuvent être mis directement en perspectives vis-à-vis des critères essentiels de conception (ici le coût financier et climatique).

- *Comparaison à base des missions représentatives*

Bien que l'estimation des coûts de possession annuels ainsi que les quantités de CO<sub>2</sub> rejetées soit plus précise et plus proches de la réalité avec les missions réelles concaténées, nous relevons les mêmes tendances sur le plan financier et le plan écologique avec les missions représentatives synthétisées. En effet, la conception avec les missions représentatives prouve aussi que le fait de concevoir une seule locomotive  $L_{12}$  reste moins avantageux sur le plan financier que la conception de deux locomotives  $L_1$  et  $L_2$  optimisées par classe de missions. Nous notons une différence de 44 k€/an entre les deux solutions.

- Coût moyen annuel de  $L_1$  et  $L_2 = 0.5 \times (\text{Coût } (C_1)/L_1 + \text{Coût } (C_2)/L_2 = 306.4 \text{ k€/an}$

- Coût annuel de  $L_{12} = 350.4 \text{ k€/an}$

En ce qui concerne les quantités de CO<sub>2</sub> émises lors d'une année de missions, la locomotive  $L_{12}$  (502 tonnes) présente, comme pour le cas précédent, un avantage de 16 tonnes/an par rapport à l'exploitation moyenne des deux locomotives  $L_1$  et  $L_2$  (518 tonnes).

Que ce soit avec les missions réelles ou les missions fictives synthétisées, l'étude précédente montre que la classification permet d'évaluer l'intérêt d'un dispositif spécifique, dédié et optimisé par classe de missions par rapport à un dispositif capable de satisfaire l'ensemble des classes. L'efficacité de cette démarche est aussi renforcée par l'utilisation du processus de synthèse de profil représentatif permettant ainsi un gain notable en termes de temps de simulation.

#### **4.6 Conception par optimisation de la source d'énergie principale**

Dans cette partie, nous procédons à la conception simultanée par optimisation de la locomotive hybride en considérant la puissance nominale du groupe Diesel comme un paramètre d'optimisation. Il s'agit alors de trouver, en plus du nombre de batteries en série

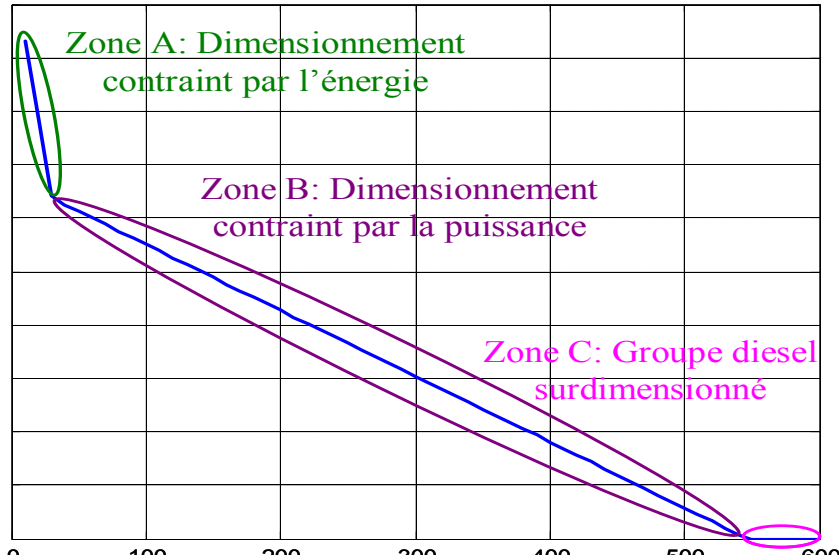
$NS_{BT}$  et en parallèle  $NP_{BT}$ , de la fréquence de découpage  $F_{dec}$  et de la tension bus  $V_{bus}$ , la bonne valeur de la puissance nominale du groupe Diesel  $P_{GDN}$ . Nous utilisons le même processus de conception simultanée par optimisation décrit dans la [section 4.3](#). La contrainte  $g_2$  (cf. équation (4.15)) qui traduit le nombre de batteries minimal pour une puissance nominale donnée du groupe Diesel est exprimée sous forme d'un graphe  $N_{BTmin} = f(P_{GDN})$ . En effet pour chaque valeur de  $P_{GDN}$  est associé un nombre de batteries minimal en dessous duquel la mission des batteries sera dégradée. Nous donnons à la [Figure 4.19](#) l'évolution du nombre de batteries minimal en fonction de la puissance nominale du groupe Diesel pour la mission  $M_{rep}$ , représentative de la famille des 7 missions de la locomotive BB63000 déterminée au [chapitre 3](#). Nous relevons l'existence de trois zones distinctes dont les caractéristiques sont fortement liées au contenu énergétique de la mission.

- **Zone A** : dans cette zone, les batteries se comportent comme une source d'énergie. En effet la puissance nominale du groupe Diesel est assez faible et le contenu énergétique de la mission des batteries est donc très important. Le dimensionnement des batteries (nombre de cellules Ni-Cd) est alors contraint par l'énergie utile.
- **Zone B** : contrairement à la zone A, le dimensionnement des batteries dans la zone B est contraint par la puissance maximale de décharge<sup>27</sup>. Ces dernières jouent alors le rôle d'une source de puissance.
- **Zone C** : cette zone traduit le fait que la puissance nominale du groupe Diesel est largement suffisante pour satisfaire la mission et qu'aucun élément de batterie n'est nécessaire.

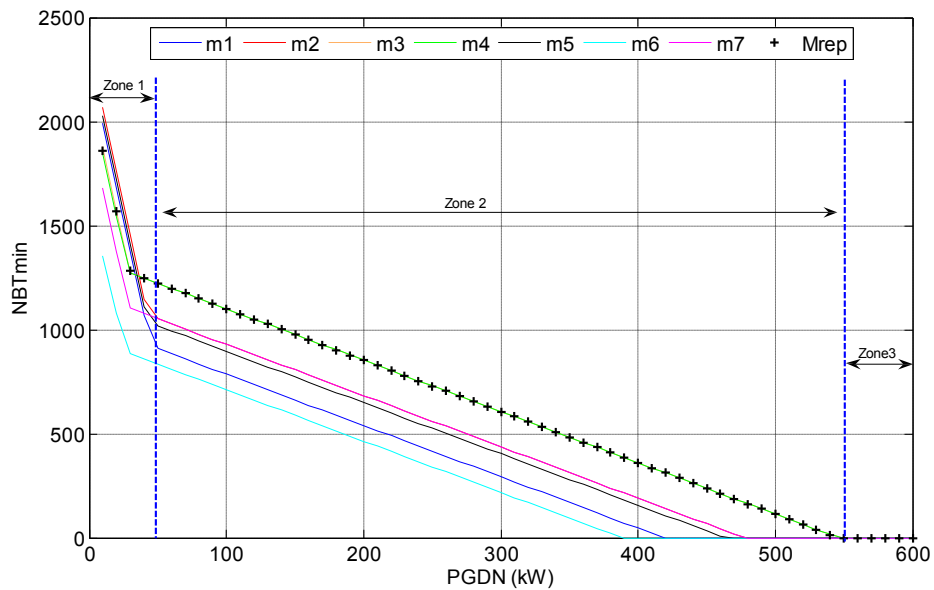
Nous signalons que le nombre de cellules de batteries est plus sensible à la variation de la puissance nominale du groupe Diesel quand ces dernières sont en mode « source d'énergie » (zone A). Ceci dépend d'une part des caractéristiques énergie/puissance de l'élément de batterie considéré et d'une autre part du contenu énergétique de la mission. A ce niveau, il est important de souligner, encore une fois, le fort couplage qui existe entre la mission et le choix technologique des briques énergétiques à intégrer dans le système vis-à-vis d'un certain nombre de contraintes de conception telles que les limites de masse et de volume associées aux sources d'énergie [[JAA 09b](#)], [[JAA 10d](#)].

---

<sup>27</sup> puissance crête de la mission des batteries



**Figure 4.19.** Nombre de batteries minimal en fonction de  $P_{GDN}$  pour la mission  $M_{rep}$



**Figure 4.20.** Nombre de batteries minimal en fonction de  $P_{GDN}$  pour la mission représentative  $M_{rep}$  et les 7 missions de la locomotive BB63000

Nous rappelons au lecteur que la mission représentative  $M_{rep}$  a été synthétisée pour une configuration bien déterminée. Il s'agit d'une puissance moyenne de 70 kW qui dimensionne le groupe Diesel, une énergie utile des batteries d'accumulateurs de 36.4 kWh et une puissance maximale de 546 kW. Dans le cas où nous considérons la puissance nominale du groupe Diesel  $P_{GDN}$  comme un paramètre d'optimisation dans le processus de conception simultanée, il est incontournable de poser la question suivante : pour chaque valeur de la puissance nominale du groupe Diesel  $P_{GDN}$ , le nombre de batteries minimal  $N_{BTmin}$  déterminé

à partir de la mission représentative  $M_{rep}$  correspond-t-il au nombre de batteries minimal pour satisfaire l'ensemble des 7 missions réelles du cahier des charges? Autrement dit, il s'agit de vérifier si la mission représentative  $M_{rep}$  reste également dimensionnante lorsque nous modifions la taille de la source d'énergie principale.

La Figure 4.20 donne l'évolution du nombre de batteries minimal  $N_{BTmin}$  en fonction de la puissance nominale du groupe Diesel  $P_{GDN}$  pour chacune des 7 missions réelles de la locomotive BB63000. Nous constatons, pour chaque mission, l'apparition des trois zones A, B et C de la Figure 4.19 décrivant le caractère source d'énergie/source de puissance des batteries d'accumulateurs. En comparant ce graphe avec celui donnant l'évolution du nombre de batteries minimal pour la mission représentative  $M_{rep}$ , il s'avère que cette dernière n'est plus dimensionnante dans la zone où les batteries, dimensionnées à partir de  $M_{rep}$ , jouent le rôle d'une source d'énergie (zone 1). En effet, lorsque nous modifions la taille du groupe Diesel, rien ne nous garantit que l'énergie utile de la mission représentative  $M_{rep}$ , déterminée à partir d'une puissance nominale du groupe Diesel  $P_{GDN} = 70$  kW, reste dominante. En revanche, lorsqu'il s'agit d'un dimensionnement contraint par la puissance maximale de décharge des batteries, la mission représentative  $M_{rep}$  demeure dominante en termes de nombre minimal de batteries à respecter. Effectivement, cette mission est synthétisée en vérifiant la plus grande valeur des puissances maximales de l'ensemble des missions réelles ( $P_{max\ ref} = 546$  kW). Par conséquent, la puissance maximale de décharge des batteries est dominante par rapport aux puissances crêtes de toutes les missions batteries. Ce phénomène est décrit par la zone 2 de la Figure 4.20 délimitée par les puissances nominales du groupe Diesel  $P_{GDNmin} = 40$  kW et  $P_{GDNmax} = 550$  kW. Dans cette zone, nous constatons que la courbe obtenue à partir de la mission représentative  $M_{rep}$  est superposée avec les courbes relatives aux deux missions réelles  $m_3$  et  $m_4$  caractérisées aussi par la plus grande valeur de puissance maximale sur l'ensemble des missions ( $P_{max} = 546$  kW).

La zone 3 montre qu'au-delà de  $P_{GDNmax} = 550$  kW, le groupe Diesel seul est capable de satisfaire l'ensemble des missions. Nous n'avons évidemment aucun intérêt à un tel dimensionnement dans une démarche d'hybridation.

L'analyse précédente montre l'insuffisance de la mission représentative pour le dimensionnement de la locomotive hybride lorsque la puissance nominale du groupe Diesel impose au stockage le caractère source d'énergie (zone A). En outre, même si la mission représentative permet de déterminer le nombre de batteries minimal à respecter sur une plage



de variation de la puissance nominale  $P_{GDN}$ , rien ne garantit qu'elle représente le même nombre de cycles pour une valeur de  $P_{GDN}$  différente de 70 kW. Ces limites ne mettent pas en cause l'approche de synthèse de profil représentatif d'une variable environnementale mais plutôt la façon avec laquelle nous avons défini les indicateurs de conception. En effet, en considérant la puissance nominale du groupe Diesel comme paramètre d'optimisation, nous ne traitons pas le même problème de conception qu'avec le cas où la taille du groupe Diesel est fixe. Il s'agit en fait d'un problème plus compliqué dont il est absolument nécessaire de revisiter et d'enrichir les indicateurs de conception afin que le contenu informationnel de la mission représentative soit plus pertinent.

Une perspective possible à ce travail consiste à concevoir une base de données de missions représentatives pour différentes tailles du groupe Diesel (à chaque valeur de  $P_{GDN}$ , nous appliquons le processus de synthèse de mission représentative avec une mise à jour des valeurs de référence des indicateurs de conception) et d'adapter, dans le processus de conception par optimisation, le choix de la mission représentative en fonction de la valeur de  $P_{GDN}$  choisie par l'optimiseur comme paramètre d'optimisation. Cette solution trouve son intérêt dans le cas où le temps de simulation nécessaire à la préparation de la base de données des missions représentatives en fonction de  $P_{GDN}$  reste bien inférieur au temps CPU pour la conception optimale du système en considérant comme mission représentative la concaténation de toutes les missions du cahier des charges.

#### **4.7 Conclusion**

Ce chapitre a été consacré à la validation du processus de synthèse de mission représentative et compacte par la voie d'une démarche de conception simultanée par optimisation d'une locomotive hybride. Après avoir présenté les modèles de conception de la locomotive hybride, ainsi que le processus de conception optimale et en se référant à une famille de 7 missions ferroviaires de la locomotive BB63000, nous avons dimensionné une locomotive hybride équivalente à partir de la mission représentative et compacte de cette famille de missions. Nous avons prouvé que l'intégration de la mission représentative dans le processus de conception par optimisation conduit aux mêmes résultats de dimensionnement que lorsque nous utilisons l'ensemble des missions réelles. L'utilisation de la mission représentative et l'efficacité du processus de synthèse de profil représentatif est justifiée par le gain notable en termes de temps de calcul qui s'élève jusqu'à 3.3 jours. Dans une deuxième

partie, nous avons prouvé l'efficacité d'une démarche systémique globale en tant qu'outil d'aide à la décision. Cette dernière démarche est basée sur une démarche en trois étapes : la classification des missions, la synthèse de missions représentatives par classe de missions et enfin sur la conception optimale d'une locomotive hybride par classe de missions. Cette démarche nous a permis d'évaluer l'intérêt de concevoir une seule locomotive hybride optimisée et capable de satisfaire l'ensemble des missions par rapport à plusieurs locomotives hybrides dédiées et optimisées par classe de missions : cette démarche constitue donc, pour le concepteur – exploitant, une « aide à la segmentation du marché associé à une aide à la conception des systèmes ». A la fin de ce chapitre, nous avons montré l'importance d'adapter la synthèse de la mission représentative en fonction du problème de conception via le choix adéquat et la pertinence des indicateurs. Cet exemple d'application à la traction ferroviaire hybride a notamment permis de mettre en évidence le couplage fort entre mission et dimensionnement, ce qui complexifie sensiblement l'approche.



## Conclusion générale

La conception optimale des systèmes énergétiques ne cesse de se complexifier, nécessitant une évolution vers une approche de conception de plus en plus intégrée, incluant plusieurs points de vue (architecture, dimensionnement, gestion) de la conception, adressant un grand nombre de critères et contraintes hétérogènes, et associant différents domaines disciplinaires. Ces systèmes doivent d'autre part satisfaire une finalité de conception (mission) et se trouvent immergés dans leur environnement avec lequel ils interagissent. Ces divers niveaux de couplages souvent intimes sont à la source de la complexité de l'approche systémique, rendant nécessaire la mise en place d'une démarche méthodologique facilitant la mise en œuvre du processus de conception. C'est dans ce contexte que le travail effectué dans cette thèse constitue une contribution scientifique à une phase amont de ce processus, à savoir l'intégration de la mission et des variables environnementales dans une démarche de conception systémique.

Dans une première phase d'étude et après avoir exposé un état de l'art sur la prise en compte de la mission et des variables d'environnement dans la conception des systèmes énergétiques, nous avons dressé les concepts de base de notre nouvelle approche de traitement de la mission et des données environnementales. Cette approche est en premier lieu basée une phase d'analyse, nécessitant une bonne expertise du système dans son environnement, afin de définir des indicateurs de caractérisation pertinents vis-à-vis des critères et contraintes de conception. Ces indicateurs, dont la formulation mathématique se définit par les caractéristiques intrinsèques du système, trouvent leurs valeurs de référence dans le contenu informationnel des profils réels et bases de données définissant la mission et les variables d'environnement à traiter.

Il ne s'agit donc pas ici de traiter ces signaux environnementaux selon leur nature intrinsèque (forme, contenu spectral, statistique,...), mais plutôt d'assurer la mise en

cohérence directe entre propriétés des signaux traités et caractéristiques (indicateurs) du système à concevoir.

Deux exemples bien distincts sont traités dans ce mémoire afin d'illustrer l'intérêt de nos contributions :

- Le premier s'inscrit dans le contexte du projet PLATHEE dans lequel des missions ferroviaires font parties intégrante du processus de conception simultanée par optimisation d'une locomotive hybride.
- Le deuxième concerne le traitement de la vitesse du vent pour la conception d'une chaîne éolienne.

Ces exemples ont volontairement été choisis pour leur différence en termes de nature et de domaine : le premier dans le domaine ferroviaire adresse plutôt la problématique du traitement de la mission, tandis que le second dans le domaine de la production d'énergie, traduit la question du traitement de variables environnementales.

Notre approche se décline en deux étapes essentielles qui se situent en phase amont du processus de conception systémique :

- La « classification » de profils de mission ou de variables environnementales selon des indicateurs pertinents vis-à-vis des critères et contraintes de conception, fait partie d'un ensemble d'outils d'aide à la décision. En effet, elle doit permettre au concepteur d'évaluer l'intérêt d'un dispositif spécifiquement optimisé pour une classe donnée, ceci par rapport à un dispositif capable de satisfaire simultanément un ensemble de classes. Cette classification peut ainsi aider le concepteur à segmenter sa gamme de produit selon leur positionnement sur le plan technico-économique. Dans ce sens, nous avons développé au deuxième chapitre un algorithme de classification basé sur l'optimisation d'un critère de partitionnement (l'indice silhouette) associé à un algorithme génétique avec nichage. Cet algorithme a été ensuite validé sur trois « benchmarks » tests puis sur deux types de variables environnementales : des profils de missions ferroviaires dédiés à trois systèmes d'alimentation distincts et des profils de vitesses de vent issus de deux sites géographiques différents. Les résultats obtenus montrent la capacité de l'approche à retrouver des classes bien distinctes, mais révèle aussi la possibilité de remettre en cause certaines « idées reçues ».

- La « synthèse de profil représentatif et compact » a consisté à générer un profil fictif dimensionnant et de durée minimale à partir d'une famille ou d'une classe de profils réels de caractéristiques différentes et ce en respectant un certain jeu d'indicateurs de conception de référence. Le développement de ce processus de synthèse a fait l'objet du troisième chapitre de ce manuscrit. La synthèse de profil est basée sur l'agrégation ou la concaténation de motifs élémentaires dont le nombre et les paramètres sont déterminés par optimisation à l'aide d'un algorithme évolutionnaire. Trois motifs élémentaires ont été proposés : le sinus, le sinus cardinal et le segment. Nous avons montré que le choix du motif élémentaire n'est pas évident et dépend de la complexité et de la variabilité du profil réel de la variable environnementale en question. En se référant aux deux exemples de profils réels de missions ferroviaires et de gisement éolien, nous avons montré la capacité du processus de synthèse à générer un profil représentatif de durée réduite et qui répond parfaitement aux indicateurs de conception de références imposés. Pour ces deux exemples ainsi traités, les meilleurs résultats sont obtenus avec le motif élémentaire de type segment, couplé au Clearing (CLR) comme algorithme d'optimisation. En ce qui concerne le taux de réduction ou de compactage de l'horizon initial des profils réels, ce dernier est nettement supérieur dans le cas du compactage de la vitesse du vent (~500) que dans le cas des missions ferroviaires (~3.5). En effet, ce taux de compactage est fortement lié à la nature des indicateurs de conception qui peuvent imposer des temps minimums substantiels comme c'est le cas de l'indicateur de l'énergie utile ou du nombre de cycles qui sont considérés comme pertinents pour caractériser les missions ferroviaires. Les indicateurs retenus pour le traitement de la vitesse du vent sont « rapidement synthétisables » et n'introduisent pas des contraintes sévères sur la durée du profil à synthétiser mise à part le nombre d'échantillons du signal qui doit être suffisant.

Enfin, au niveau du quatrième chapitre, nous avons présenté une mise en application de l'approche complète proposée dans ce manuscrit allant de la classification jusqu'à la conception simultanée en passant par la synthèse de mission : cet exemple concernait la conception optimale d'une locomotive hybride économe en énergie et respectueuse de l'environnement. Cette démarche d'hybridation s'est inscrite dans le cadre du projet PLATHEE et a consisté à remplacer le groupe thermique d'une locomotive tout Diesel par un autre de taille plus réduite et de compléter le manque de puissance par des batteries d'accumulateurs. A travers cet exemple d'hybridation, nous avons montré l'importance de la

démarche de classification des missions ferroviaires pour aider le concepteur à évaluer l'intérêt d'une locomotive spécifique, dédiée et optimisée par classe de missions par rapport à une seule locomotive capable de satisfaire l'ensemble des classes. Notre démarche constitue donc, pour le concepteur – exploitant, ici la SNCF, une « aide à la segmentation du marché associé à une aide à la conception des systèmes ». Ensuite, nous avons mis en exergue l'efficacité de l'approche de synthèse de profil représentatif et de durée compacte au niveau du processus de conception par optimisation. Cette approche a conduit à des gains notables en termes de temps de calcul qui s'élèvent jusqu'à quelques jours par rapport à une conception par optimisation qui considère, comme mission typique, la concaténation de l'ensemble des missions du cahier des charges.

Une première perspective de notre travail s'inscrit dans un contexte de réduction et de simplification de la complexité des problèmes d'optimisation rencontrés lors de la synthèse de profil représentatif et compact. En effet, avec les trois types de motifs élémentaires proposés (sinus, sinus cardinal, segment), nous avons traité des problèmes inverses dont la taille peut dépasser 500 paramètres à optimiser ce qui est pénalisant en termes de convergence et de temps de calcul. La recherche d'autres motifs élémentaires à faible nombre de paramètres et plus efficaces pour la synthèse de signaux à forte complexité nous semble indispensable. Dans ce sens, une collaboration est en cours avec le laboratoire MAPMO (Mathématiques – Analyse, Probabilités, Modélisation – Orléans) afin d'examiner une telle approche de synthèse de signal en utilisant les « ondelettes » comme motifs élémentaires.

Une deuxième perspective de notre travail consisterait à reprendre la conception optimale d'une chaîne éolienne passive, travail déjà réalisé dans notre groupe de recherche GENESYS ([TRA10]), mais cette fois-ci en considérant un profil de vitesse du vent compact représentatif de la base de données considérée. Nous envisageons également d'appliquer cette approche de signal compact pour le dimensionnement optimal d'un système hybride associant une chaîne éolienne et des composants de stockage d'énergie (thèse BELOUDA M. – en cours). Il est aussi possible d'étendre le champ d'application de notre approche de compactage du contenu informationnel des données réelles à d'autres systèmes énergétiques en traitant d'autres types de variables environnementales tels que les irradiances solaires pour des systèmes photovoltaïques, des relevés de températures pour des systèmes thermiques,...

Dans le contexte de la conception optimale d'une locomotive hybride, la synthèse de la mission représentative développée au [chapitre 3](#) et intégrée au processus de conception par

optimisation au niveau du [chapitre 4](#), est basée sur un *a priori* relatif à la puissance nominale du groupe Diesel que nous avons du fixer à la puissance moyenne calculée sur l'ensemble des missions concaténées du cahier des charges. Cette valeur moyenne est donc considérée comme indicateur de conception de référence à respecter par la mission représentative synthétisée. Le fait de prendre en compte cet *a priori*, amène à briser un couplage entre mission et système conçu, puisque le traitement de la mission s'effectue ici à partir d'un *pré-requis* sur le dimensionnement du dispositif. En effet, nous avons montré au [chapitre 4](#) que la mission représentative synthétisée pour une taille donnée du groupe Diesel n'autorise plus un dimensionnement optimal du système libérant la taille de sa source d'énergie principale (groupe Diesel); ce qui pourrait être un *degré de liberté* essentiel se trouve donc ici figé par l'approche actuelle. Une solution à la résolution de ce problème de conception serait de synthétiser non pas une mission unique, mais de constituer une base de données de missions fictives établie pour différentes tailles du groupe Diesel. Il deviendrait alors possible de considérer la puissance nominale du groupe Diesel comme paramètre d'optimisation. En effet, en fonction de la puissance choisie par l'optimiseur, nous utiliserions, d'une manière automatique, la mission représentative adéquate. Cette solution, visant à conserver le couplage entre système et mission, trouverait son intérêt dans le cas où le temps de traitement nécessaire à la préparation de la base de données des missions représentatives en fonction de la puissance nominale du groupe Diesel resterait bien inférieur au temps CPU pour la conception optimale du système en considérant comme mission représentative la concaténation de toutes les missions du cahier des charges.

De façon générale, ce dernier exemple particulier met en évidence un niveau de couplage supplémentaire entre mission/environnement et système à concevoir, interaction qui ajoute encore de la complexité au processus de conception systémique.





## Annexe A

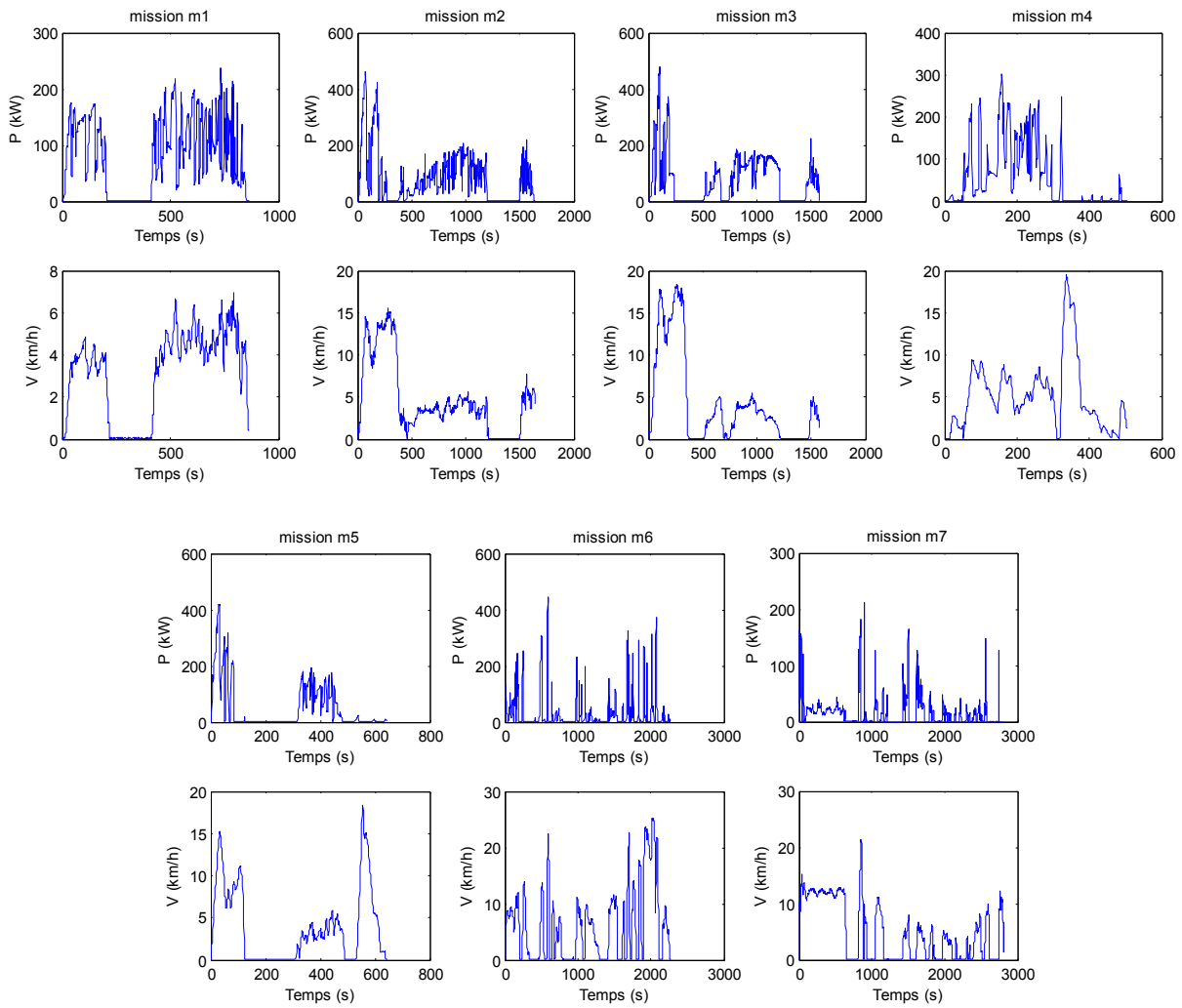
### Base de données des missions ferroviaires

#### A.1 Missions de la locomotive Diesel BB63000



##### A.1.1 *Missions triage*

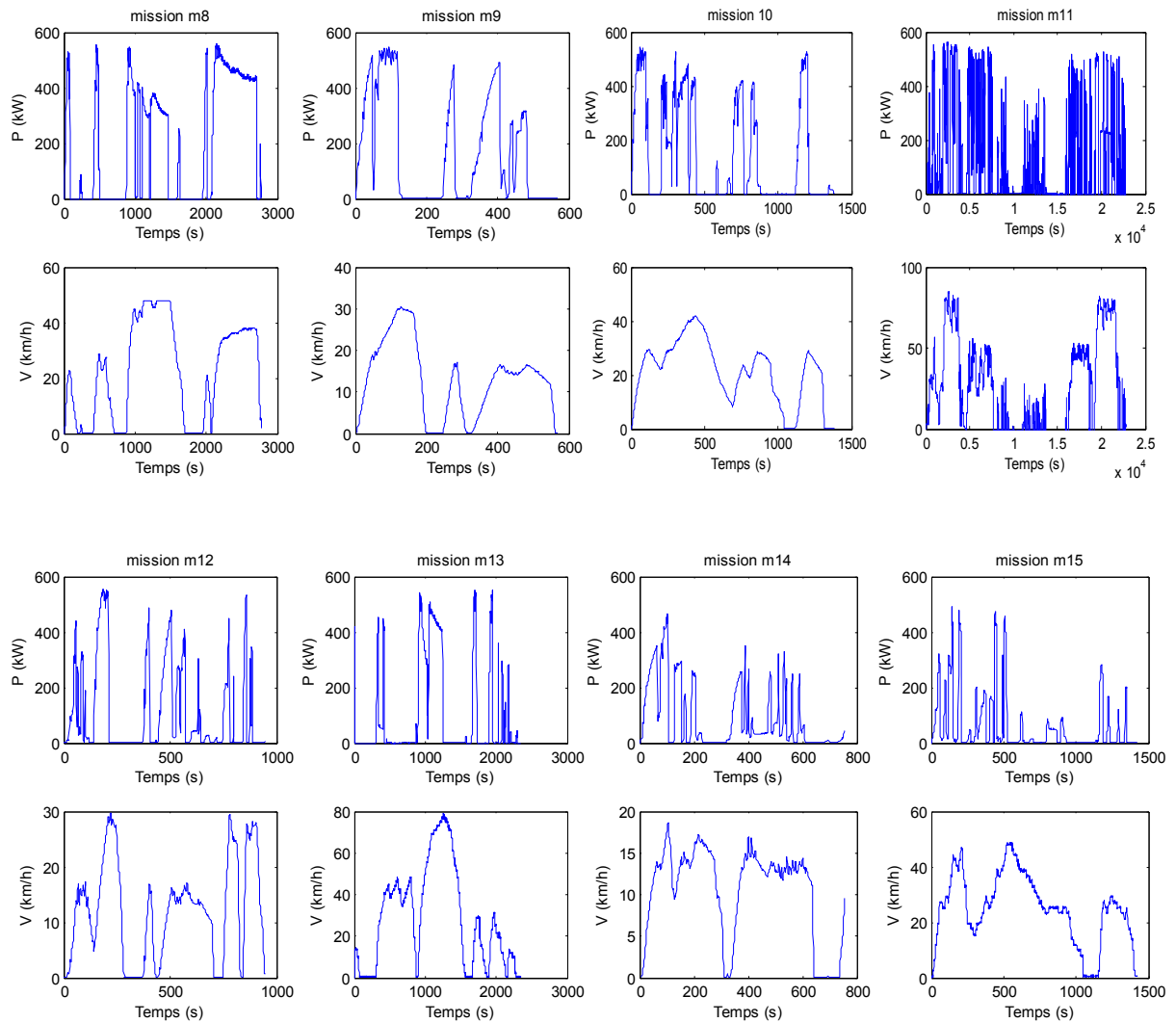
Nous disposons dans le projet PLATHEE de 7 missions ferroviaires de type triage (ou manœuvre). Nous donnons ci-dessous les chronogrammes temporels de la puissance instantanée et de la vitesse de la locomotive.



**Figure A.1.** Missions triage de la locomotive BB63000

### A.1.2 Missions desserte locale

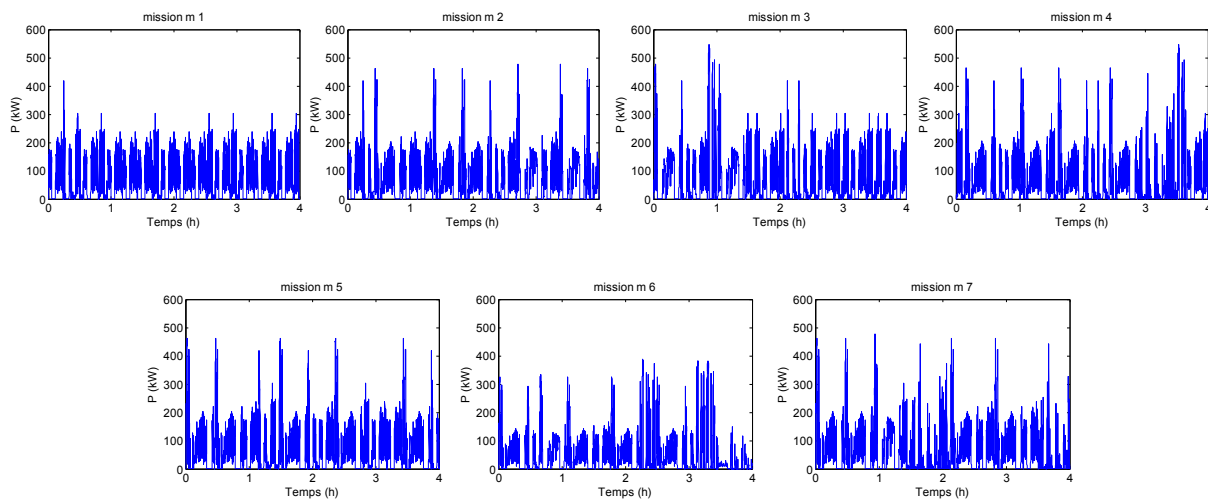
Nous disposons de 8 missions ferroviaires de type desserte locale. Nous donnons ci-dessous les chronogrammes temporels de la puissance instantanée et de la vitesse de la locomotive.



**Figure A.2.** Missions desserte locale de la locomotive BB63000

## A.2 Famille de 7 missions élémentaires de durée 4 heures

Les 7 missions ferroviaires utilisées dans le [chapitre 3](#) et le [chapitre 4](#) sont extraites à partir de la base de missions de circulations réelles présentées ci-dessus. En absence des données suffisantes à l'usage de la locomotive BB63000, notamment à l'occurrence de ses missions de durées très hétérogènes allant de quelques minutes jusqu'à quelques heures, nous proposons d'étudier un scénario de 7 missions iso-durées (4 heures par mission) et de même occurrence obtenues par un enchaînement de plusieurs missions réelles. Nous faisons l'hypothèse que l'état de charge du stockage, batteries Ni-Cd dans ce cas, est ramené à 100% après chaque séquence de missions de durée 4 heures. Ces missions sont représentées par la [Figure A.3](#).

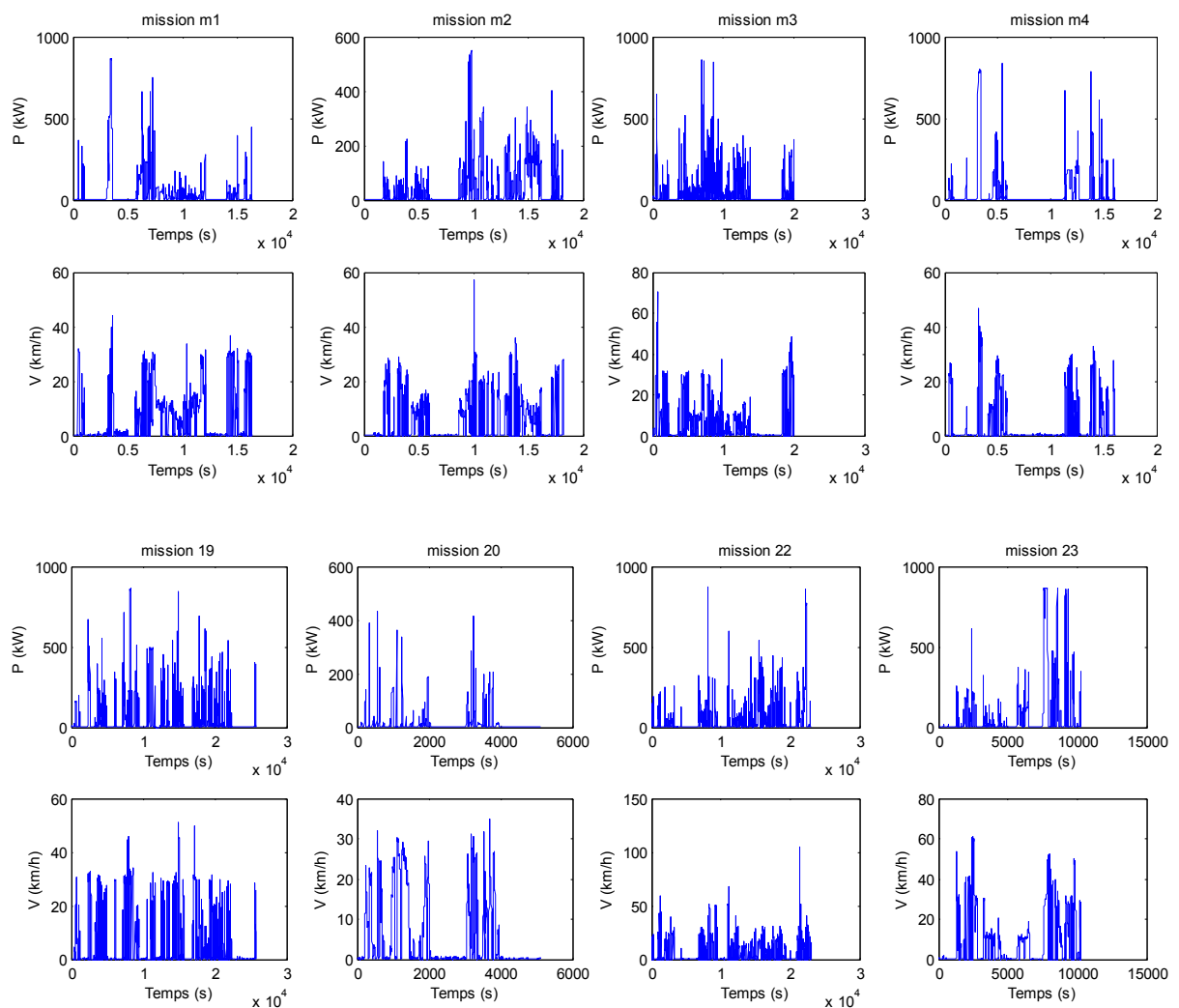


**Figure A.3.** Famille de 7 missions élémentaires de durée 4 heures

### A.3 Quelques missions de la locomotive BB460000



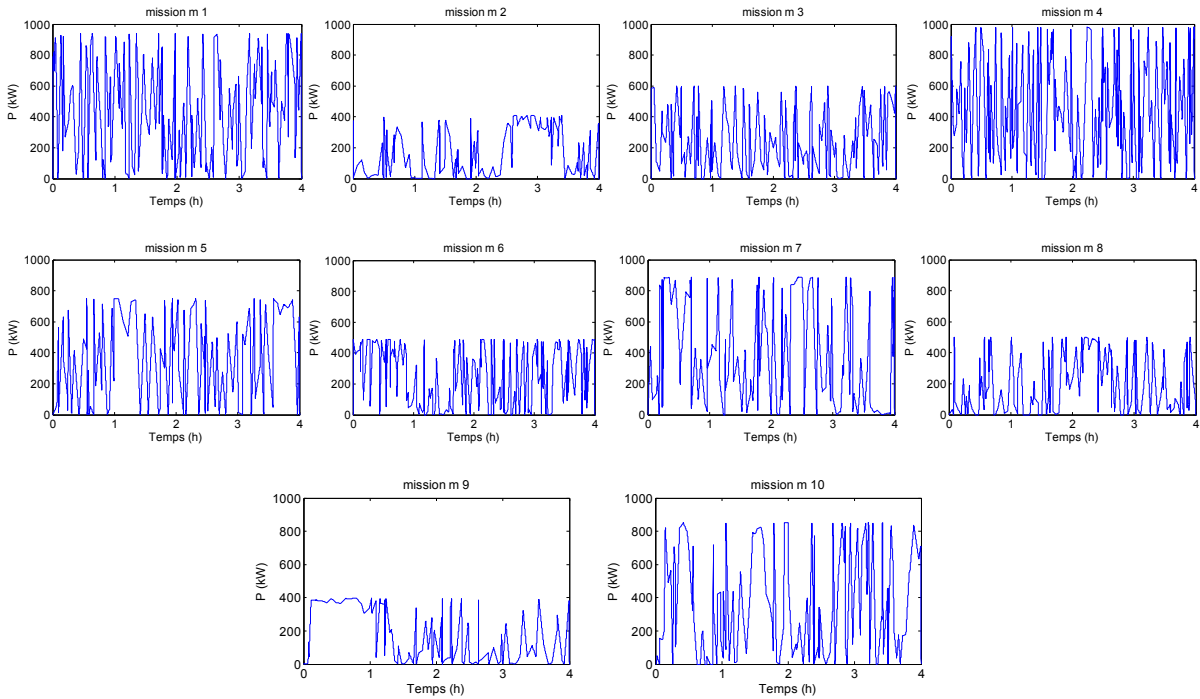
La locomotive BB460000 est une locomotive diesel dotée d'un groupe Diesel de 1 MW. Nous disposons de 27 missions effectuées sur la zone FRET de SOTTEVILLE et LE HAVRE. La [Figure A.4](#) donne quelques missions de la locomotive BB460000.



**Figure A.4.** Missions fret de la locomotive BB460000

#### A.4 Famille de 10 missions élémentaires de durée 4 heures

La famille des 10 missions utilisées dans le [chapitre 4](#) sont extraites à partir de la base de missions réelles des deux locomotives BB63000 et BB460000. Ces missions sont données par la [Figure A.5](#).



**Figure A.5.** Famille de 10 missions élémentaires de durée 4 heures

## Annexe B

### Hybridation des auxiliaires TGV POS



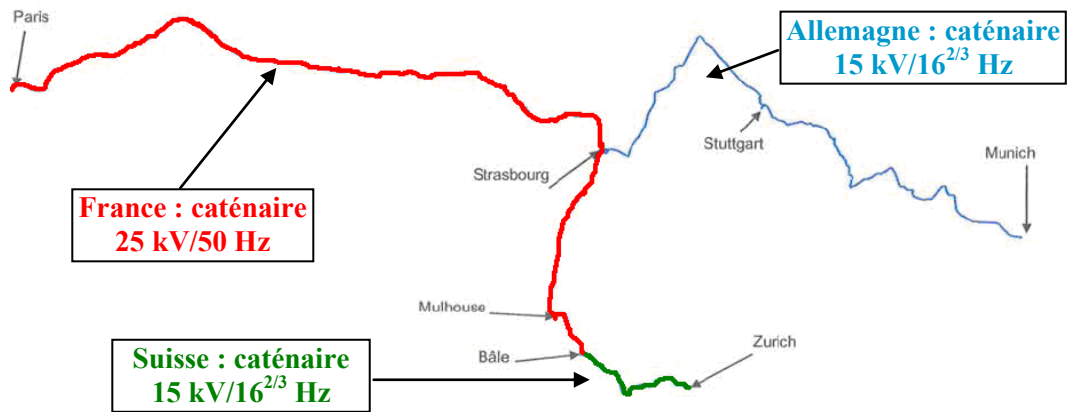
Dans le contexte d'une stratégie d'acquisition du savoir faire industriel en termes de rédaction du cahier des charges et d'analyse de futures offres de locomotives hybrides, la SNCF propose aujourd'hui d'étudier la possibilité d'hybrider les auxiliaires (confort des passagers) du TGV POS (*Paris-Ostfrankreich-Süddeuschland*, en français : *Paris – Est de la France – Allemagne du Sud*). Une campagne de mesure a eu lieu en printemps 2008 et a permis de relever plusieurs paramètres de la vie du TGV sur différents trajets le long de cette période. Dans cette annexe, nous donnons une description du TGV POS, des mesures effectuées et de la partie candidate à hybrider.

#### B.1 Description du TGV POS

##### B.1.1 Trajets du TGV POS

Le TGV POS effectue des trajets entre Paris et l'Allemagne en reliant 7 stations comme l'indique la figure ci-dessous :

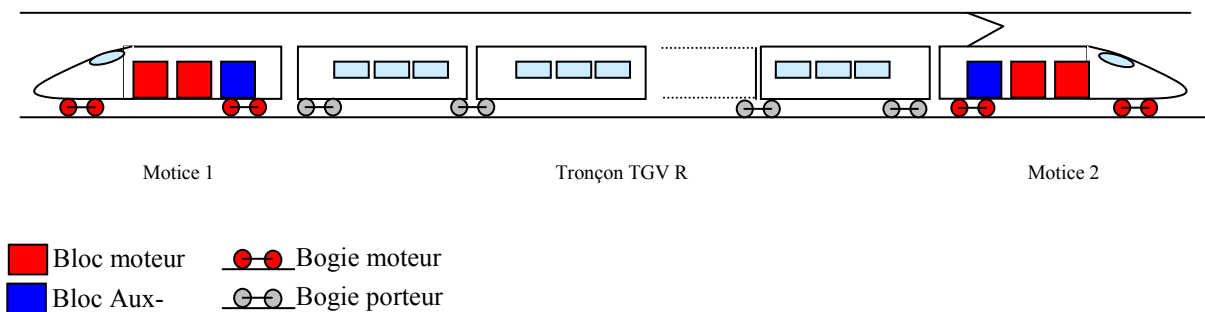




**Figure B.1.** Ensemble des trajets du TGV POS

### B.1.2 Architecture du TGV POS

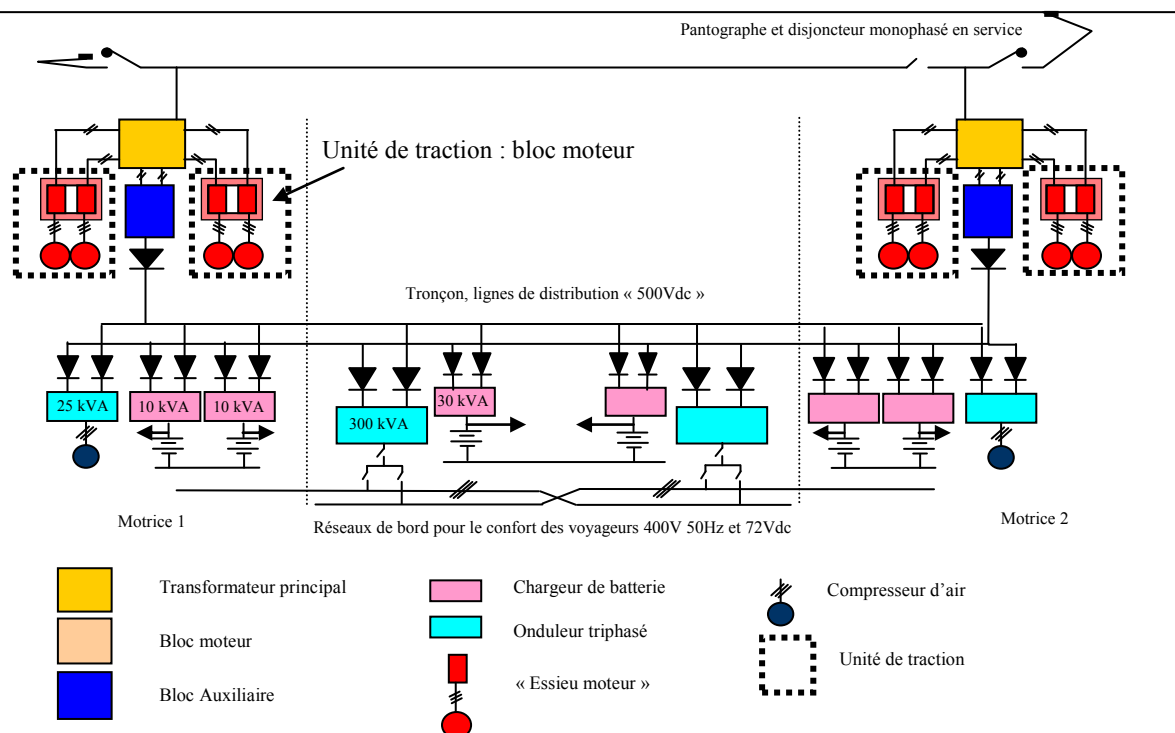
Un TGV POS est formé par 2 motrices (M1 et M2 alimentées par le même pantographe) et un tronçon généralement de 8 voitures concaténées. Chaque motrice contient 2 blocs moteurs et un bloc auxiliaire qui sera détaillé dans le paragraphe suivant.



**Figure B.2.** Architecture du TGV POS

### B.1.3 Distribution énergétique

Notons que les deux motrices du TGV POS sont identiques. L'étude se restreint alors à la motrice M1. Partant de la caténaire (25 kV/50 Hz en France et 15 kV/16.7 Hz en Allemagne et en Suisse), un transformateur principal abaisseur permet d'alimenter les deux blocs moteurs ainsi que le bloc auxiliaire. Ce dernier est un ensemble de convertisseurs permettant d'assurer l'alimentation d'une part des auxiliaires de la motrice et d'autre part des auxiliaires du tronçon.



**Figure B.3.** Distribution énergétique du TGV POS

Les auxiliaires de la motrice sont :

- Les charges assurant l'évacuation des calories produites par les organes électriques de la traction.
- Le compresseur d'air en fonctionnement intermittent alimenté par un onduleur triphasé 25 kVA.
- Le réseau de bord de la cabine.
- Les auxiliaires cabine (pupitre, éclairage, etc.) et les auxiliaires tractions alimentés par 2 chargeurs batteries 72 V/10 kVA chacun.

Les auxiliaires du tronçon dits aussi auxiliaires confort sont :

- Le chauffage : 25 kVA/voiture.
- Le système de ventilation.
- L'éclairage des voitures.
- Le système de climatisation.
- Les équipements de restauration.
- Les installations frigorifiques de conditionnement alimentaire.
- Les électroniques connectées au réseau de bord des voitures.

Tous les auxiliaires de la motrice et du tronçon, sauf le chauffage, sont connectés à la ligne continue 500 V (réellement 535 V). Il s'agit en réalité de 2 lignes issues chacune d'une motrice. Un système à diode est mis en place pour forcer le transfert d'énergie dans un seul sens.

#### B.1.4 Structure du bloc auxiliaires

Le bloc auxiliaires est connecté aux deux secondaires du transformateur principal. Chacun alimente un pont monophasé à commutation forcée (PMCF 1 et PMCF 2) qui permettent de définir la ligne «1800 V». Cette ligne est spécialement conçue pour alimenter les résistances de chauffage de toutes les voitures. Un hacheur abaisseur permet d'obtenir la ligne «500 V» qui alimente à la fois les auxiliaires de confort (à part le chauffage) et les auxiliaires de la motrice.

Le bloc auxiliaires est dimensionné autour de 1 MW, son schéma synoptique est donné par la figure suivante :

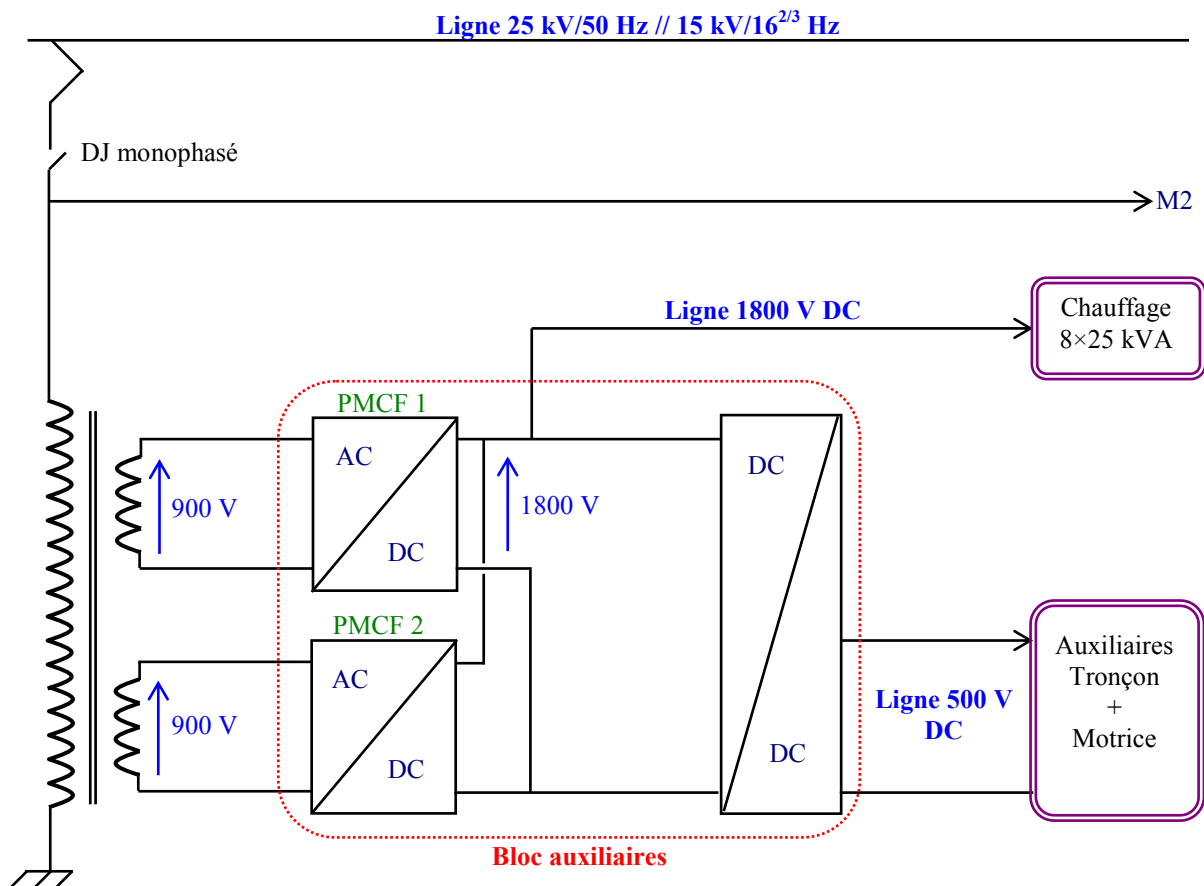
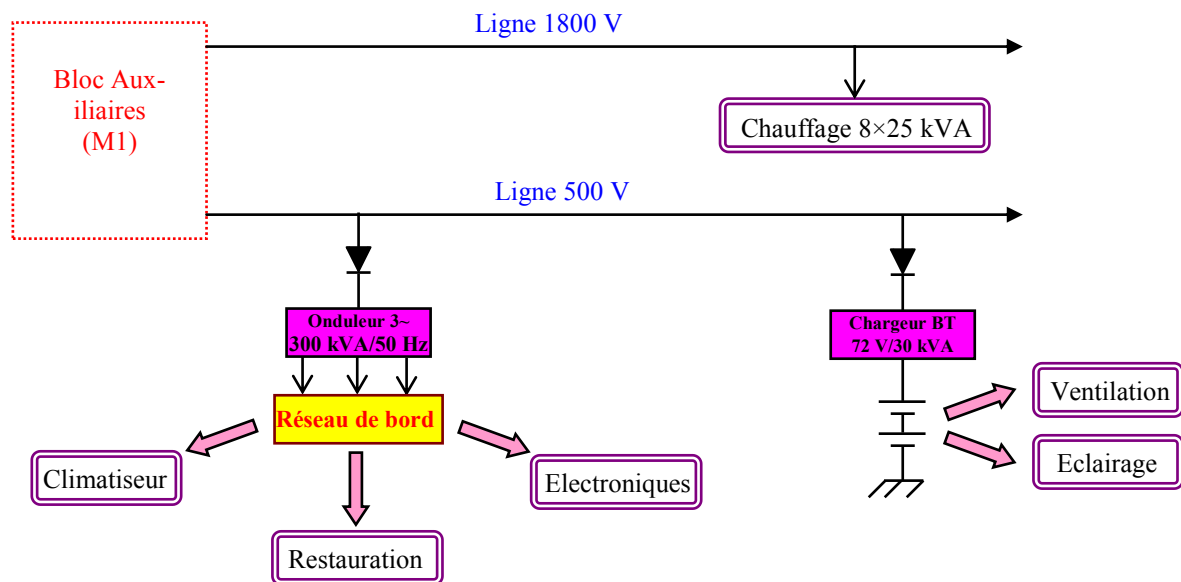


Figure B.4. Structure du bloc auxiliaires

### B.1.5 Répartition des auxiliaires tronçon

Les auxiliaires tronçon, auxquels nous attribuons un intérêt particulier en vue de leur hybridation, sont connectés au bloc auxiliaires (d'une motrice) de la façon suivante :



**Figure B.5.** Répartition des auxiliaires tronçon ou auxiliaires confort

- Le chauffage (8voitures×25 kVA) est alimenté directement par la ligne « 1800 V».
- La ventilation et l'éclairage sont assurés par un bloc de batteries (Ni-Cd) connecté à la ligne «500 V» à travers un chargeur batterie de 72 V/30 kVA.
- Le reste des auxiliaires (climatiseur, fours, micro-onde, électronique alimentée par le réseau de bord....) est connecté à la ligne «500 V» à travers un onduleur triphasé 300 kVA/50 Hz.

Nous signalons que le chargeur de batteries joue deux rôles : d'une part il assure l'alimentation de la ventilation et l'éclairage tant que la caténaire est présente (les batteries sont alors chargées), d'autre part, il assure la charge des batteries lorsque celles-ci sont déchargées. Parmi les causes principales de décharge des batteries, nous citons :

- La perte de la caténaire pour des phases de moyennes et longues durées.
- La perte de caténaire pour des phases de courtes durées (de l'ordre de la minute), notamment pour les instants de changement de phases par l'EDF en vue de l'équilibrage du réseau ou lors du passage de la frontière entre deux pays.

## B.2 Vers une démarche d'hybridation des auxiliaires du TGV POS

Dans le cadre du programme d'économie d'énergie, la SNCF veut étudier la possibilité de réduire la consommation des auxiliaires TGV POS tout en garantissant un certain niveau de confort. Le problème reste à définir le mot « confort » auprès des voyageurs. Un questionnaire « Qu'est ce que vous apparait utile comme confort sur un TGV POS » est lancé par la SNCF en 2006 et la tendance constatée est que les voyageurs veulent « tout ». Satisfaire le désir des voyageurs devient alors un compromis vis-à-vis des contraintes à respecter (masse, volume,...).

Parmi les problèmes rencontrés sur un TGV POS, nous notons :

✓ *Perte de la caténaire :*

Dans le cas de perte de caténaire pour des longues durées (problème du réseau EDF, phases de maintenance), les voyageurs perdent une très grande partie de leur confort. En effet, tous les auxiliaires du tronçon seront hors service à part la ventilation et l'éclairage. Cependant, ces deux derniers comforts ne sont que temporaires et dépendent directement de l'autonomie des batteries installées.

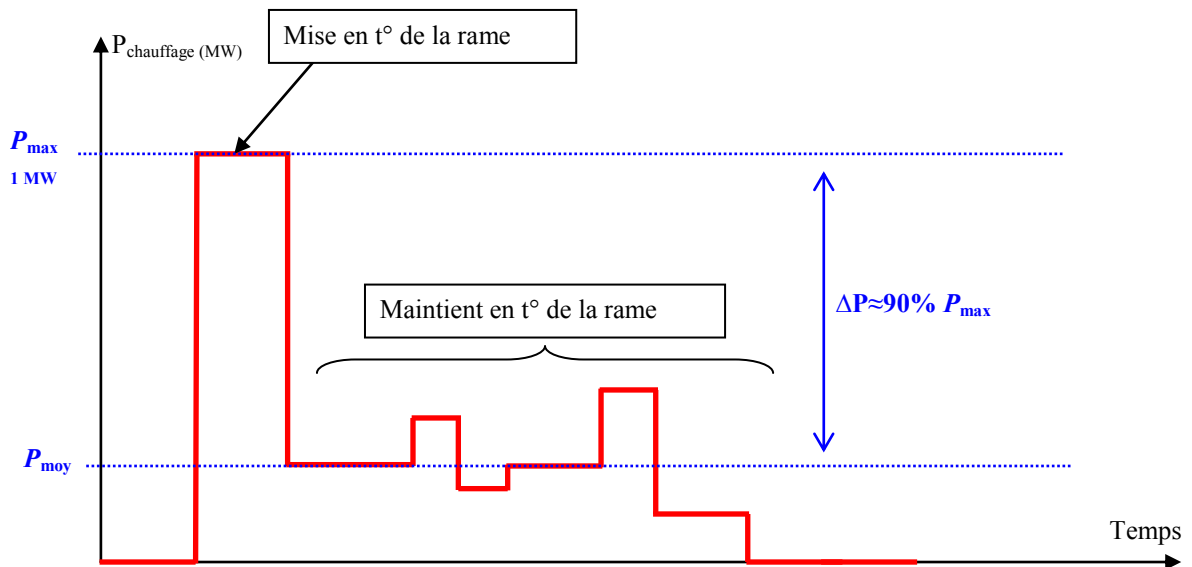
La perte de caténaire pour des phases de courtes durées (changement de phases EDF ou passage par la frontière de deux pays) ne pose pas un grand problème à part une décharge rapide des batteries.

Un problème très accentué est la perte du chauffage surtout en hiver pendant des durées plus au moins longues. La situation est encore grave en été lorsqu'il fait assez chaud (perte du climatiseur).

✓ *Consommation excessive lors de la mise en température de la rame :*

Une analyse préliminaire de la consommation en puissance du chauffage ou du climatiseur lors de la mise en température de la rame depuis l'arrêt montre une grande différence des niveaux de puissance relatifs à la mise en température d'une part et de son maintien d'une autre part (cf. [Figure B.6](#)).

L'analyse des problèmes cités ci-dessus laisse penser à suivre une démarche d'hybridation capable à la fois d'asservir le bilan énergétique des auxiliaires TGV et d'assurer un certain niveau de confort dans le cas de perte de la source d'énergie principale.



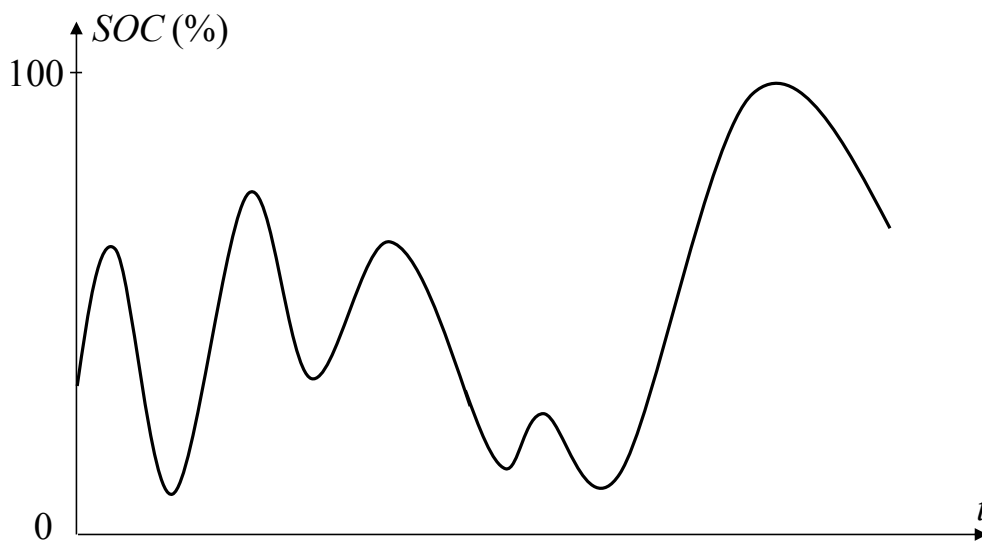
**Figure B.6.** Profil type de la puissance consommée par le chauffage

La valeur  $P_{\text{max}}$  est de l'ordre de 1 MW dans le cas extrême quand il fait très froid avec une seule motrice en service. Le bloc auxiliaires est alors dimensionné pour cette puissance. Un autre cas extrême quand il fait très chaud correspond à une consommation de puissance de l'ordre de 0.6 MW pour alimenter le climatiseur.



### Comptage des cycles avec la méthode de Rainflow

Dans cette annexe, nous donnons une description de la procédure de comptage des cycles par la méthode de Rainflow permettant d'estimer la durée de vie d'un élément de stockage d'énergie. Cette méthode admet en entrée la courbe d'état de charge de l'élément de stockage ( $SOC(t)$ ) et renvoie le nombre de cycles par profondeur de décharge. Dans ce qui suit, nous appliquons la méthode de Rainflow sur la courbe  $SOC(t)$  de la [Figure C.1](#).



**Figure C.1.** Exemple d'une courbe d'état de charge

#### C.1 Traitement préliminaire de la courbe d'état de charge $SOC(t)$

La première étape de l'algorithme de Rainflow consiste à extraire tous les extrema de la courbe d'état de charge de l'élément de stockage  $SOC(t)$  comme le montre la [Figure C.2](#).



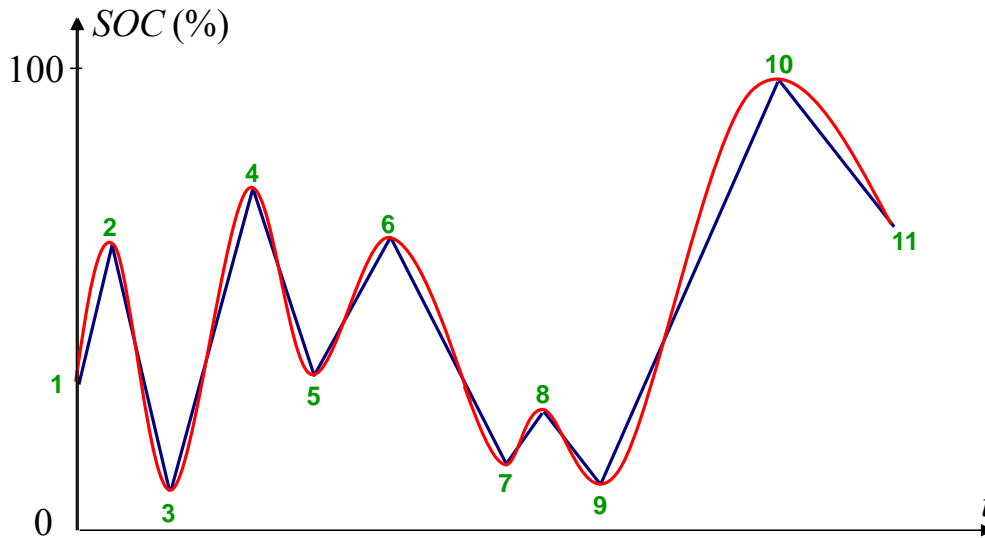


Figure C.2. Extraction des extremums de la courbe  $SOC(t)$

## C.2 Définition des classes des profondeurs de décharge

La deuxième étape consiste à définir des classes ou des intervalles des profondeurs de décharge (cf. Figure C.3). Dans cet exemple d'illustration, nous choisissons des classes de largeur 10% de profondeur de décharge. Ensuite, il faut définir les niveaux représentatifs des classes, généralement le centre de chaque classe comme le montre la Figure C.4. Enfin, nous ramenons tous les extrema de la Figure C.2 aux centres des classes correspondantes (cf. Figure C.5).

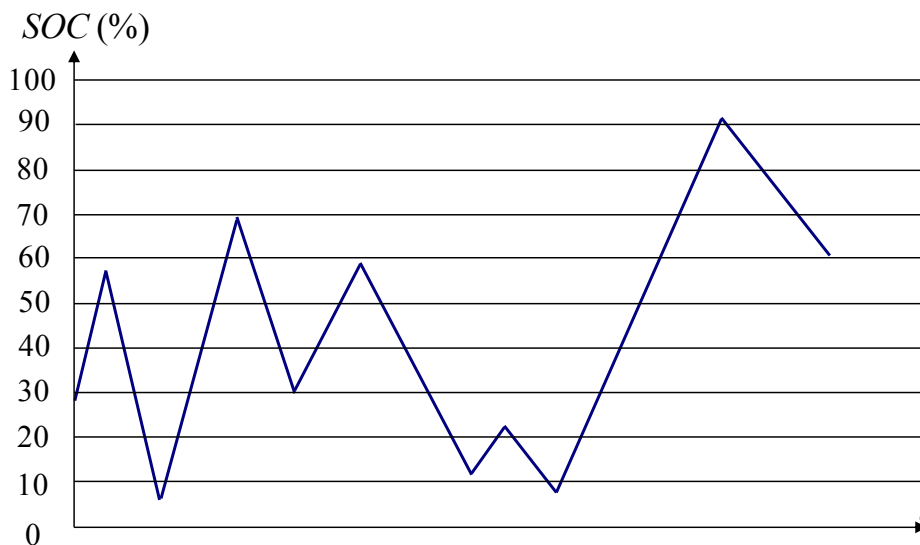
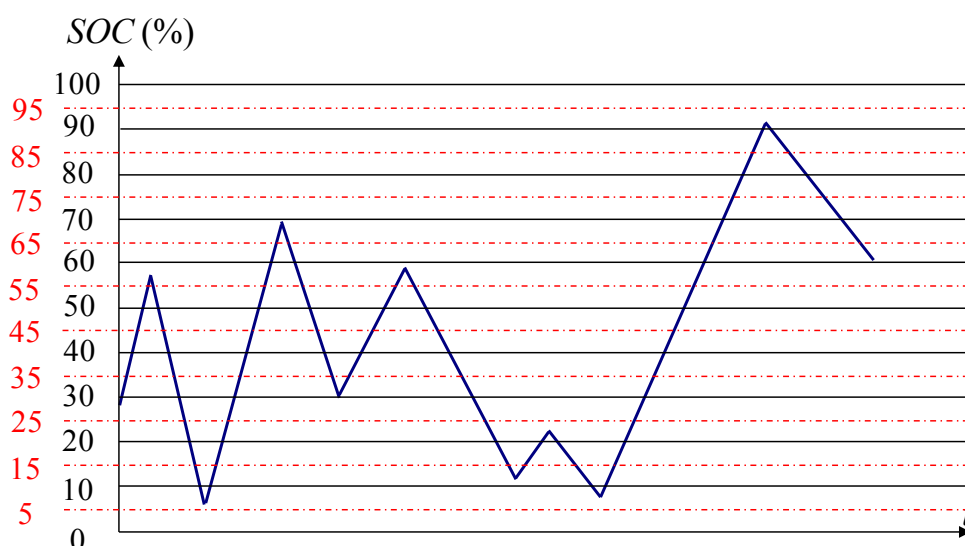
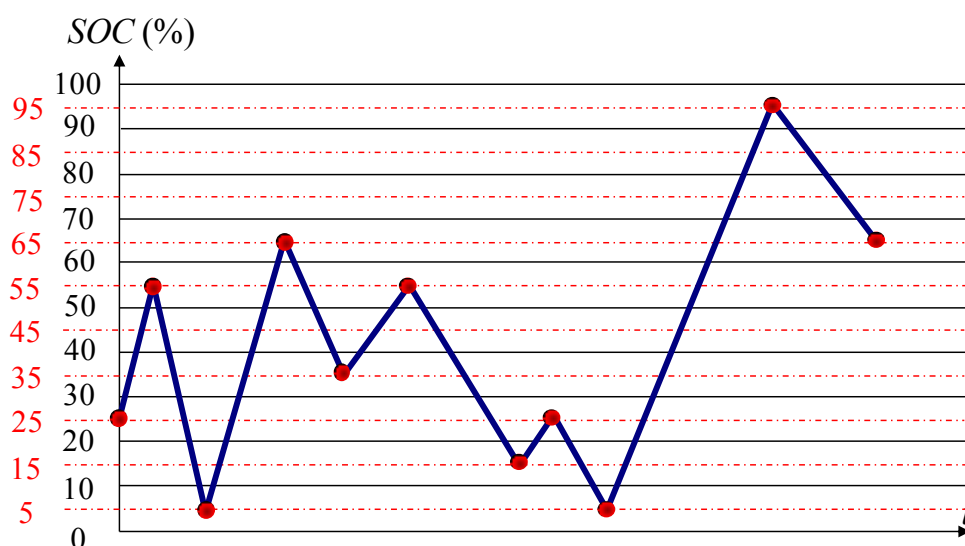


Figure C.3. Définition des classes ou intervalles des profondeurs de décharge



**Figure C.4.** Définition des niveaux représentatifs des classes



**Figure C.5.** Les extrema sont ramenés aux centres des classes

### C.3 Extraction des cycles

L'extraction des cycles est basée sur le calcul des étendues relatives à quatre extrema successifs ( $S_i, S_{i+1}, S_{i+2}, S_{i+3}$ ) du signal  $SOC(t)$  placés aux centres des classes (cf. [Figure C.6](#))

$$\Delta S_i = |SOC(S_i) - SOC(S_{i+1})|, \quad \Delta S_{i+1} = |SOC(S_{i+1}) - SOC(S_{i+2})|, \quad \Delta S_{i+2} = |SOC(S_{i+2}) - SOC(S_{i+3})|$$

Si  $\Delta S_{i+1} \leq \Delta S_i$  et  $\Delta S_{i+1} \leq \Delta S_{i+2}$  alors :

- Le cycle de profondeur  $\Delta S_{i+1}$  est extrait
- Les extrema  $S_{i+1}$  et  $S_{i+2}$  sont éliminés

La procédure s'applique de nouveau sur le signal obtenu après élimination de ( $S_{i+1}$  et  $S_{i+2}$ ) avec un balayage de gauche à droite en faisant glisser les quatre extrema. Dès qu'un cycle est extrait, la procédure démarre à nouveau.

➤ Première itération :

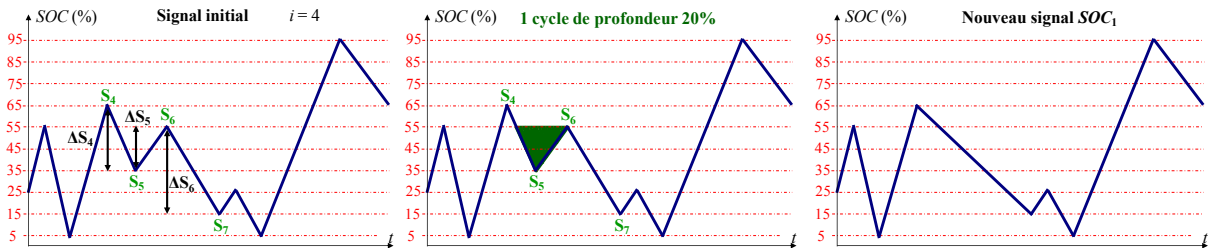


Figure C.6. Première itération : extraction d'un cycle de profondeur 20%

➤ Deuxième itération :

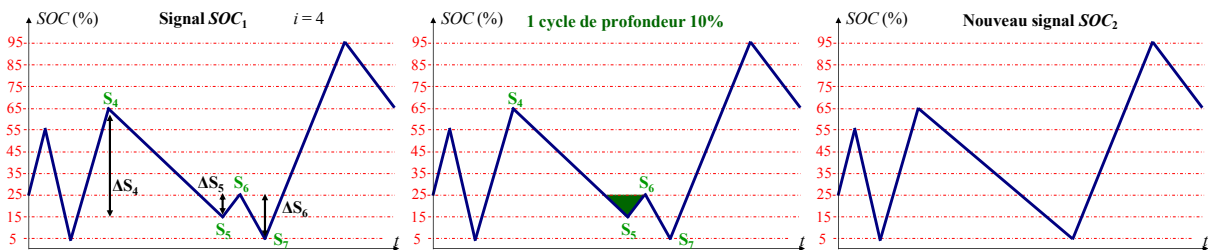


Figure C.7. Deuxième itération : extraction d'un cycle de profondeur 10%

➤ Troisième itération :

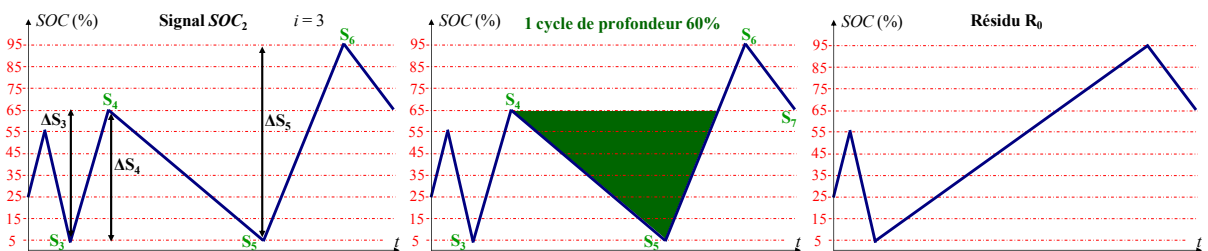


Figure C.8. Troisième itération : extraction d'un cycle de profondeur 60%

Après la troisième itération, nous obtenons un signal qualifié de « résidu » et dont la méthode de Rainflow précédente est incapable d'identifier les cycles de ce signal. Toutefois, en dupliquant ce résidu (en rajoutant le même signal à sa suite moyennant quelques précautions de raccordement des extrema) et en appliquant la méthode de Rainflow au signal

obtenu formé par les deux résidus, le même résidu initial est obtenu à la fin du comptage. Les cycles extraits (lors du traitement des deux résidus) correspondent donc aux cycles du résidu initial.

➤ Traitement du résidu :

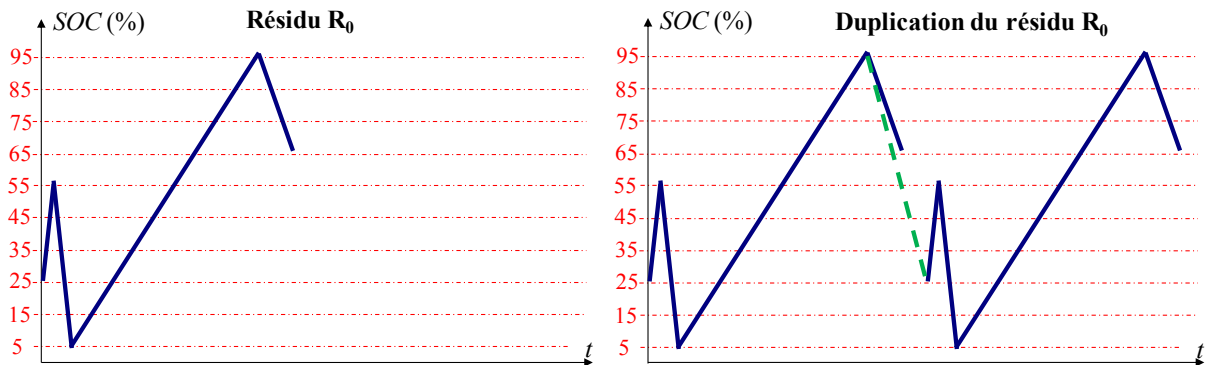


Figure C.9. Duplication du résidu

• Extraction des cycles du résidu (première itération) :

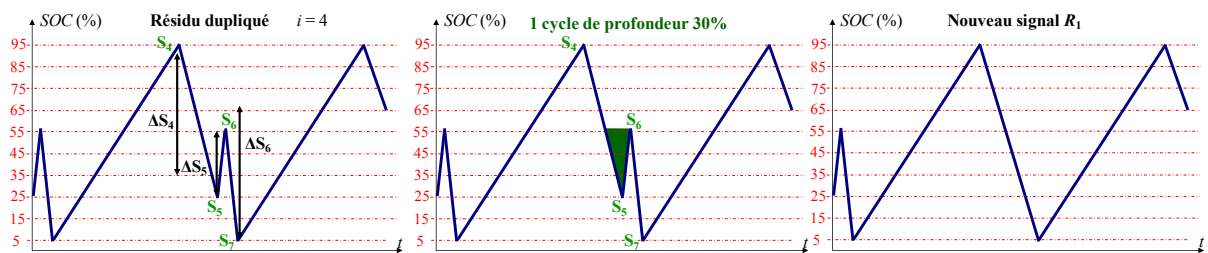


Figure C.10. Extraction des cycles du résidu : première itération

• Extraction des cycles du résidu (deuxième itération) :

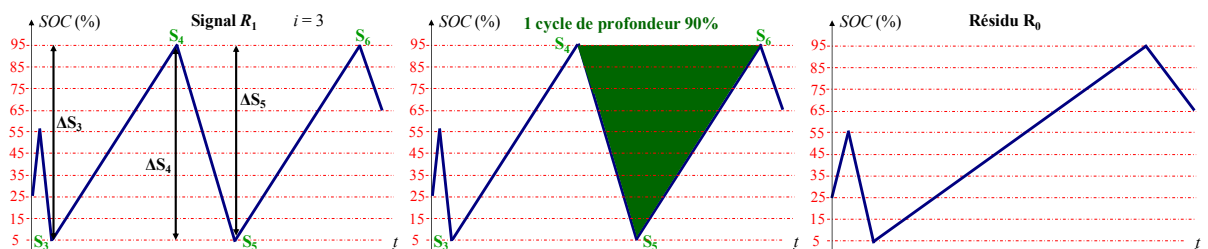


Figure C.11. Extraction des cycles du résidu : deuxième itération

En conclusion, les cycles de la courbe d'état de charge de la [Figure C.1](#) sont :

- Un cycle de profondeur 20%
- Un cycle de profondeur 10%
- Un cycle de profondeur 60%
- Un cycle de profondeur 30%
- Un cycle de profondeur 90%

## Annexe D

# Algorithmes évolutionnaires

### D.1 Principe

Les méthodes évolutionnaires sont des techniques d'optimisation stochastiques qui reposent sur une analogie avec la théorie de l'évolution naturelle des espèces de *Darwin*. Dans la nature, les espèces sont capables de s'adapter à leur milieu sous la pression de la sélection naturelle, et d'évoluer à partir de leur héritage génétique de manière à accroître leur adaptation au fil des générations. C'est une métaphore du paradigme darwinien qui est utilisée dans les algorithmes évolutionnaires pour la résolution de problèmes d'optimisation. La structure générale d'un algorithme évolutionnaire est donnée à la [Figure D.1](#). L'algorithme évolutionnaire est une suite de générations reproduisant les principes de l'évolution naturelle en agissant sur une population d'individus à l'aide d'opérateurs de deux types :

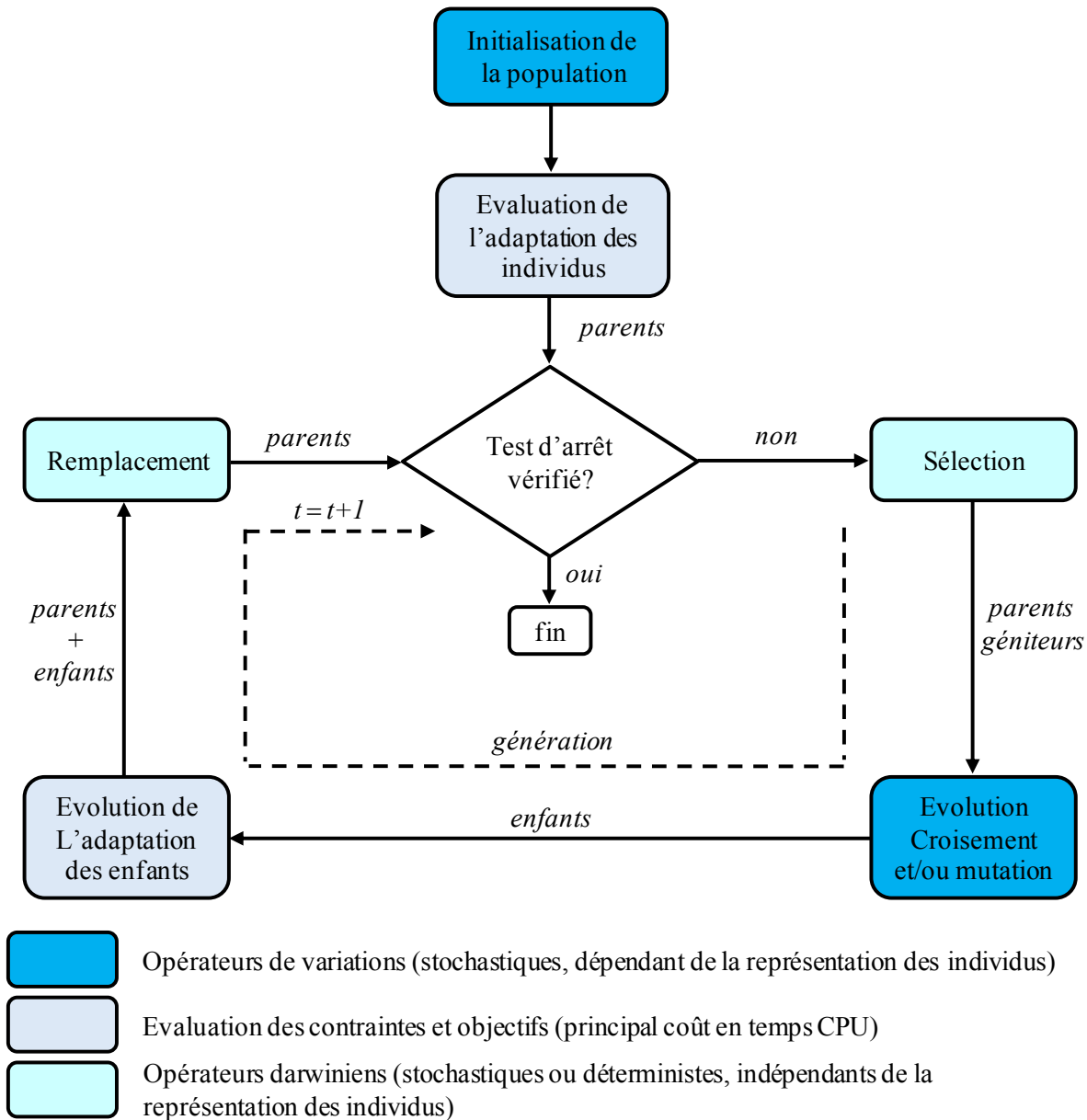
- des opérateurs darwiniens qui transposent les mécanismes de sélection naturelle en orientant la recherche vers les solutions les mieux adaptées,
- des opérateurs d'évolution (croisement<sup>28</sup> et/ou mutation), qualifiés aussi d'opérateurs de variation, dont le rôle est de favoriser l'émergence de nouvelles solutions.

La population initiale est obtenue par exploration aléatoire (ou compte tenu de connaissances *a priori*) de l'espace de recherche. Les parents géniteurs sont alors sélectionnés parmi la population à l'aide de procédures stochastiques ou déterministes. Vient ensuite une étape d'évolution où des opérateurs de variation sont appliqués aux parents géniteurs pour créer et explorer de nouvelles solutions (*enfants*). Enfin, un processus de remplacement est appliqué pour déterminer les « survivants » d'une génération. L'algorithme peut être

---

<sup>28</sup> On parle aussi de recombinaison ou hybridation pour qualifier l'opérateur de croisement

*générationnel* (les enfants remplaçant automatiquement les parents) ou *élitiste* (un ou plusieurs parents étant conservés de façon déterministe ou stochastique). Dans ce dernier cas, l'algorithme évolutionnaire est qualifié de quasi-stationnaire (*steady state EA*) dans la mesure où seule une partie de la population est renouvelée à chaque génération. Il est à noter que les tailles des populations de parents, géniteurs et enfants peuvent être identiques ou distinctes selon le type d'algorithme [KEI 02].



**Figure D.1.** Structure générale d'un Algorithme Evolutionnaire

Parmi les méthodes évolutionnaires les plus connues, le lecteur pourra se référer aux algorithmes génétiques [HOL 75], [GOL 89], [MIC 96], aux stratégies de dévolution [REC 73], [SCH 65] à la programmation évolutionnaire [FOG 91] et à l'évolution différentielle [STO 96].

## D.2 Les Stratégies d'Evolution : ES

Les Stratégies d'Evolution (ES) sont des Algorithmes Evolutionnaires qui ont été développés parallèlement dans les années 70 par deux chercheurs allemands, *Ingo Rechenberg* [REC 73] et *Hans-Paul Schwefel* [SCH 65]. Dans sa version de base, l'algorithme utilise un ensemble de  $\mu$  parents pour produire  $\lambda$  enfants ( $\lambda > \mu$ ) par mutation gaussienne (ajout d'une variable aléatoire suivant une loi normale). Une étape de sélection est ensuite appliquée pour déterminer les parents de la génération suivante. Ceux-ci sont déterminés soit à partir des  $\mu$  meilleurs enfants générés au cours de la génération (sélection générationnelle de type  $\mu, \lambda$ ) soit à partir des  $\mu$  meilleurs individus parmi l'ensemble parents + enfants de la génération courante (sélection élitiste  $\mu + \lambda$ ). A l'origine, il n'y avait aucune étape de recombinaison entre parents ce qui fait de l'opérateur de mutation le principal moteur d'évolution dans les stratégies d'évolution. Par la suite, il a été montré que l'introduction d'un opérateur de recombinaison s'avérait nécessaire pour améliorer les performances de l'algorithme, notamment pour des problèmes multimodaux [SCH 87]. L'autre caractéristique spécifique de l'algorithme est la présence d'écart-types  $\sigma_i$  associés aux paramètres  $x_i$  à optimiser dans le génotype des individus. Ces écart-types, eux-mêmes soumis aux opérateurs d'évolution (recombinaison + mutation) et de sélection, sont utilisés pour muter les enfants obtenus par recombinaison. Cette procédure confère aux stratégies d'évolution des propriétés d'auto-adaptation rendant l'algorithme particulièrement robuste dans un contexte d'optimisation dynamique [SAR 06]. Le pseudo-code détaillé d'une stratégie d'évolution est donné à la Figure D.2. L'opérateur de recombinaison  $r$  peut être implanté de différentes manières. On distingue plus particulièrement deux types de recombinaison : la recombinaison discrète et la recombinaison intermédiaire<sup>29</sup>, conformément à l'équation ci-après :

$$\alpha'_i = \begin{cases} \alpha_i(\mathbf{I}_1) \text{ ou } \alpha_i(\mathbf{I}_2) & \text{recombinaison discrète} \\ \alpha_i(\mathbf{I}_1) + U(0,1) \times (\alpha_i(\mathbf{I}_2) - \alpha_i(\mathbf{I}_1)) & \text{recombinaison intermédiaire} \end{cases}$$

où les variables correspondent  $\alpha_i$  aux paramètres ou aux écart-types associés,  $\mathbf{I}_1$  et  $\mathbf{I}_2$  représentent deux individus parents pris au hasard dans la population et  $U(0,1)$  désigne une variable aléatoire à densité uniforme prise dans l'intervalle  $[0,1]$ . Il est généralement

<sup>29</sup> On parle plutôt de croisement discret et croisement continu (ou croisement arithmétique) pour qualifier les recombinaisons discrète et continue dans les algorithmes génétiques



recommandé d'appliquer une recombinaison discrète sur les variables  $x_i$  à optimiser et une recombinaison intermédiaire sur les écart-types  $\sigma_i$  associés [SCH 87].

$g := 0$

*Initialise*  $P(0) := \{\mathbf{I}_1(0), \mathbf{I}_2(0), \dots, \mathbf{I}_\mu(0)\} \in \mathbb{R}^n$

où  $\mathbf{I}_k = \{\mathbf{x}_k, \boldsymbol{\sigma}_k\} = \{x_i, \sigma_i\} \quad \forall i = 1, \dots, n$

*Evalue*  $P(0) : \{\Phi(\mathbf{I}_1(0)), \Phi(\mathbf{I}_2(0)), \dots, \Phi(\mathbf{I}_\mu(0))\}$

où  $\Phi(\mathbf{I}_k) = f(\mathbf{x}_k)$

**Pour**  $g = 1 \dots G$

*Recombine*  $\mathbf{I}'_k(g) := r(P(g)) \quad \forall k = 1, \dots, \lambda$

*Mute*  $\mathbf{I}''_k(g) := m(\mathbf{I}'_k(g)) \quad \forall k = 1, \dots, \lambda$

*Evalue*  $P''(g) : \{\Phi(\mathbf{I}'_1(g)), \Phi(\mathbf{I}'_2(g)), \dots, \Phi(\mathbf{I}'_\lambda(g))\}$

*Sélectionne*  $P(g+1) := s_{(\mu, \lambda)}(P''(g)) \quad // \text{ sélection } \mu, \lambda$

ou  $P(g+1) := s_{(\mu+\lambda)}(P(g) \cup P''(g)) \quad // \text{ sélection } \mu+\lambda$

**Fin pour**

Notation	Signification
$n$	Nombre de paramètres à optimiser
$f$	Fonction à optimiser
$P$	Population de parents
$P''$	Population d'enfants
$\mathbf{I}_k$	Individu parent
$\mathbf{I}''_k$	Individu enfant (muté)
$\mathbf{x}_k$	Vecteur des paramètres $x_i$ à optimiser
$\boldsymbol{\sigma}_k$	Vecteur des écarts-types $\sigma_i$ associés aux paramètres $x_i$
$\mu$	Nombre de parents
$\lambda$	Nombre d'enfants
$\Phi$	Opérateur d'évaluation
$r$	Opérateur de recombinaison
$m$	Opérateur de mutation
$s_{(\mu, \lambda)}$	Opérateur de sélection $(\mu, \lambda)$
$s_{(\mu+\lambda)}$	Opérateur de sélection $(\mu+\lambda)$

**Figure D.2.** Pseudo-code d'une stratégie d'évolution

Enfin, l'opérateur de mutation  $m$  consiste d'abord à muter les écarts-types puis les paramètres de l'individu de la façon suivante :

$$\sigma'_i = \sigma_i \cdot \exp(\tau' \cdot N(0,1) + \tau \cdot N_i(0,1))$$

$$x'_i = x_i + \sigma'_i \cdot N_i(0,1)$$

où  $N(0,1)$  représente une variable aléatoire à densité gaussienne de moyenne nulle et d'écart-type unitaire,  $N_i(0,1)$  précise que cette variable est « renouvelée » pour chaque composante  $i$  des vecteurs relatifs aux paramètres et aux écarts-types. Les paramètres  $\tau$  et  $\tau'$  sont proportionnels à des valeurs empiriques déterminées dans [SCH 77], valeurs conférant à l'algorithme une bonne robustesse :

$$\begin{aligned}\tau &= \alpha \left( \sqrt{2\sqrt{n}} \right)^{-1} \\ \tau' &= \alpha \left( \sqrt{2n} \right)^{-1}\end{aligned}$$

### D.3 Les Algorithmes Génétiques standards : AG

Les Algorithmes Génétiques se sont développés parallèlement aux stratégies d'évolution aux Etats-Unis dans les années 70. Sous leur version de base, ils n'utilisent pas directement les paramètres à optimiser mais un codage de ces paramètres représentant le *chromosome* des individus solutions du problème. Par analogie avec la théorie de l'évolution et de la génétique, les individus à travers leur chromosome sont soumis aux opérateurs de sélection, croisement et mutation. Contrairement aux stratégies d'évolution, le moteur principal d'évolution dans les algorithmes génétique est le croisement que l'on applique avec une probabilité  $p_c$  élevée (typiquement  $p_c \in [0.7-1.0]$ ). Le taux de mutation des variables du chromosome est par contre limité pour bénéficier des effets positifs liés au croisement (stabilisation de la recherche dans les zones prometteuses) et éviter une exploration purement aléatoire. L'autre différence notable par rapport aux stratégies d'évolution est relative à la taille de la population. Pour plus de détails par rapport aux algorithmes génétiques, le lecteur pourra se référer à [GOL 89], [MIC 96], [HOL 75], [SAR 99].

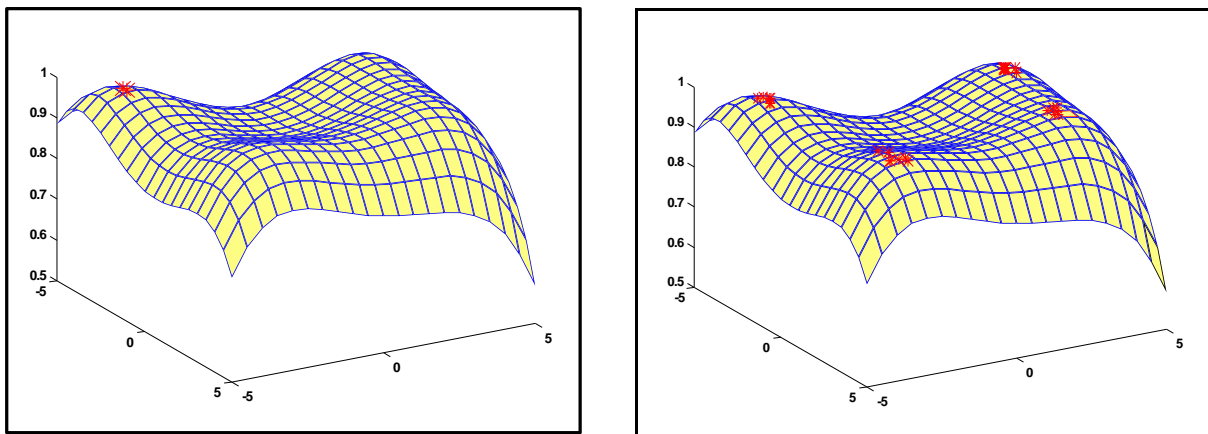
L'inconvénient majeur des stratégies d'évolution et algorithmes génétiques classiques est lié au phénomène de dérive génétique que l'on constate au fil des générations. En effet, l'opérateur de sélection finit par concentrer les individus de la population en une unique solution. Cette concentration nuit aux capacités d'évolution puisque l'opérateur de recombinaison est alors incapable de faire apparaître des solutions nouvelles<sup>30</sup>. En cas de convergence prématurée vers une solution ne représentant pas l'optimum global du problème, seul l'opérateur de mutation pourra permettre de s'extraire de l'attracteur local. Ce phénomène de dérive génétique, un peu moins prononcé dans les stratégies d'évolution en

<sup>30</sup> La recombinaison de deux parents identiques conduit à des enfants en tout point identiques à leur parent.

raison d'une plus grande importance de l'opérateur de mutation ainsi que des propriétés d'auto-adaptation, reste toutefois limitatif, en particulier pour résoudre des problèmes d'optimisation fortement multimodaux.

#### D.4 Les Algorithmes Génétiques avec méthodes de nichage

Les méthodes de nichage ont été développées pour réduire la dérive génétique. Elles préservent constamment la diversité dans la population assurant ainsi la capacité d'évolution des individus au cours de la recherche. En maintenant plusieurs solutions en parallèle, elles permettent en outre l'identification de solutions optimales multiples (cf. Figure D.3). Nous donnons ci-dessous le principe de deux méthodes de nichage.



(a) Algorithme génétique sans nichage

(b) Algorithme génétique avec nichage

**Figure D.3.** Exemples de solutions obtenues sans et avec méthode de nichage

##### D.4.1 La Sélection par Tournoi Restreint : RTS

La sélection par tournoi (*Tournament Selection*) est une méthode de sélection qui permet de choisir les individus qui pourront se reproduire pour créer la génération suivante. A chaque génération, une compétition a lieu entre des individus pris aléatoirement dans la population et les vainqueurs de ces tournois forment les parents de la nouvelle génération. Ce mécanisme est particulièrement bien approprié à l'optimisation unimodale mais se révèle inefficace lorsque l'on désire maintenir plusieurs solutions, en raison de l'importante dérive génétique qu'il introduit. De ce fait, *Harik* a adapté la sélection par tournoi à l'optimisation multimodale et a proposé une technique de nichage appelée sélection par tournoi restreint (*Restricted Tournament Selection*) dont le principe est le suivant [HAR 95] :

- ✓ 2 individus  $p_1$  et  $p_2$  sont sélectionnés aléatoirement dans la population.
- ✓ Ces individus, croisés et mutés, produisent deux enfants  $e_1$  et  $e_2$ .
- ✓  $w$  éléments sont pris au hasard dans la population globale. A partir de cet échantillon, deux sont sélectionnés,  $p_1'$  et  $p_2'$ , de telle sorte que leurs caractéristiques génétiques soient les plus proches des enfants respectifs  $e_1$  et  $e_2$  précédemment créés.
- ✓ Une compétition a ensuite lieu entre ( $p_1'$  et  $e_1$ ) et ( $p_2'$  et  $e_2$ ). Les vainqueurs des tournois sont réinsérés dans la population.

La procédure est répétée  $N/2$  fois par génération avec  $N$  désigne la taille initiale de la population.

#### **D.4.2 La méthode d'éclaircissement : CLR**

La méthode d'éclaircissement est voisine du partage standard [GOL 87] excepté qu'elle est basée sur le principe des ressources limitées dans l'environnement [PET 96]. Au lieu de partager les ressources entre tous les individus d'une même sous-population, l'éclaircissement les attribue uniquement aux meilleurs représentants. En pratique, la capacité  $k$  d'une niche spécifie le nombre maximal d'individus que la niche peut accepter. Après détermination de la performance des individus dans chaque niche, l'éclaircissement préserve les  $k$  meilleurs représentants des sous-populations respectives (individus *dominants*) et exclut les autres (individus *dominés*) de la population en réinitialisant leur adaptation. Comme pour la méthode de partage, le voisinage est défini de façon explicite. Les individus appartenant à une même niche sont séparés d'une distance inférieure à un seuil de similarité fixé. Le rayon de niche correspondant est aussi appelé rayon d'éclaircissement.

#### **D.5 Paramétrage des algorithmes évolutionnaires**

Nous donnons ci-dessous les principaux paramètres des trois algorithmes évolutionnaires utilisés dans la thèse.

**D.5.1 Stratégie d'Evolution standard : ES**

Paramètres	Valeur
Nombre de paramètres	A définir en fonction du problème
Valeur minimale du paramètre $i$	Contrainte de domaine du problème
Valeur maximale du paramètre $i$	Contrainte de domaine du problème
Population de parents $\mu$	15
Population d'enfants $\lambda$	100
Nombre maximal de générations	A définir en fonction du problème
Type de croisement des variables à optimiser	Croisement discret
Type de croisement des écarts-types associés	Continu global
Type de sélection	Sélection générationnelle ( $\mu, \lambda$ )

**Tableau D.1.** Paramètres de la Stratégie d'Evolution standard**D.5.2 Algorithme génétique avec Sélection par Tournoi Restreint : RTS**

Paramètres	Valeur
Nombre de paramètres	A définir en fonction du problème
Valeur minimale du paramètre $i$	Contrainte de domaine du problème
Valeur maximale du paramètre $i$	Contrainte de domaine du problème
Taille de la population	100
Nombre maximal de générations	A définir en fonction du problème
Probabilité de mutation $p_m$	0.1
Probabilité de mutation du gène d'auto-adaptation	0.05
Paramètre du croisement BLX- $\alpha$	0.5
Paramètre du croisement vSBX	1
Paramètre du croisement PNX	1
Paramètre de la mutation polynomiale	50
Type de croisement	auto-adaptatif
Facteur de surpeuplement	50
Norme utilisée pour le calcul de distance entre individus	norme Euclidienne

**Tableau D.2.** Paramètres de l'algorithme génétique avec Sélection par Tournoi Restreint

**D.5.3 Algorithme génétique avec méthode d'éclaircissement : CLR**

<b>Paramètres</b>	<b>Valeur</b>
Nombre de paramètres	A définir en fonction du problème
Valeur minimale du paramètre $i$	Contrainte de domaine du problème
Valeur maximale du paramètre $i$	Contrainte de domaine du problème
Taille de la population	100
Nombre maximal de générations	A définir en fonction du problème
Nombre maximal de générations sans amélioration	Nombre maximal de générations
Probabilité de mutation	0.05
Probabilité de mutation du gène d'auto-adaptation	0.05
Paramètre du croisement BLX- $\alpha$	0.5
Paramètre du croisement vSBX	1
Paramètre du croisement PNX	1
Paramètre de la mutation polynomiale	50
Type de croisement	auto-adaptatif
Capacité des niches	1
Rayon de niche (normalisé)	0.05
Norme utilisée pour le calcul de distance entre individus	norme infinie

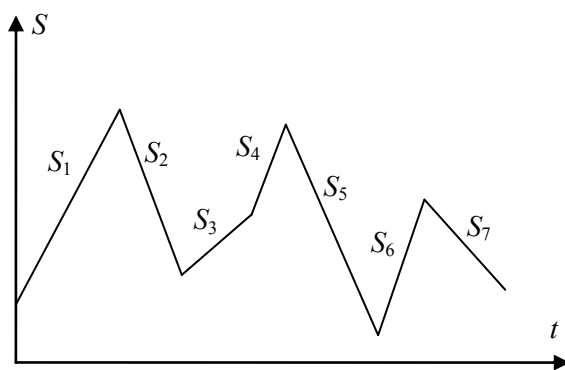
**Tableau D.3.** Paramètres de l'algorithme génétique avec éclaircissement



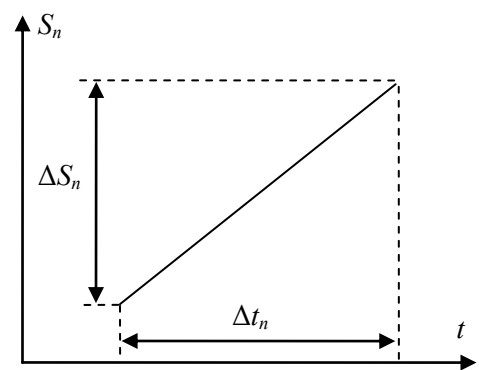
## Tests de paramétrage du motif élémentaire de type segment

### E.1 Paramétrage initial

Nous rappelons que le modèle de segment consiste à modéliser le profil d'une mission ou d'une variable environnementale par une concaténation ou une agrégation de  $N_m$  segments (cf. Figure E.1.a). Chaque segment est caractérisé par deux paramètres : une amplitude  $\Delta S_n$  et une durée  $\Delta t_n$ . Dans un premier paramétrage, nous définissons l'amplitude  $\Delta S_n$  comme étant la différence entre les deux extrémités du segment (cf. Figure E.1.b).



(a) Décomposition d'un profil en segments



(b) Paramètres d'un motif :  $\Delta S_n$  et  $\Delta t_n$

**Figure E.1.** Modélisation du profil en segments élémentaires

Afin de pouvoir représenter des segments croissants et décroissants, l'amplitude  $\Delta S_n$  d'un motif élémentaire doit varier entre une valeur minimale  $-\Delta S_{n \max}$  et une valeur maximale positive  $\Delta S_{n \max}$ . Le choix de cette dernière valeur est arbitraire. En effet, quelle que soit la



valeur maximale de l'amplitude d'un segment ( $\Delta S_{n \max}$ ), une phase de mise à l'échelle du signal entre une valeur minimale et une valeur maximale de référence est nécessaire après la concaténation de tous les segments. Par exemple, dans le cas de synthèse d'un signal représentatif d'un ensemble de missions ferroviaires, la mise à l'échelle sera effectuée entre la valeur minimale et celle maximale des missions initiales. La [Figure E.2](#) est une illustration des différentes étapes à suivre pour effectuer cette mise à l'échelle. Soit  $S''(t)$  le signal initial issu de la concaténation de  $N_m$  segments. Nous notons par  $\Delta S''$  la différence entre la valeur maximale et la valeur minimale de  $S''(t)$  (cf. [Figure E.2.a](#)).

$$\Delta S'' = \text{abs}(S''_{\max}) - \text{abs}(S''_{\min}) \quad (\text{E.1})$$

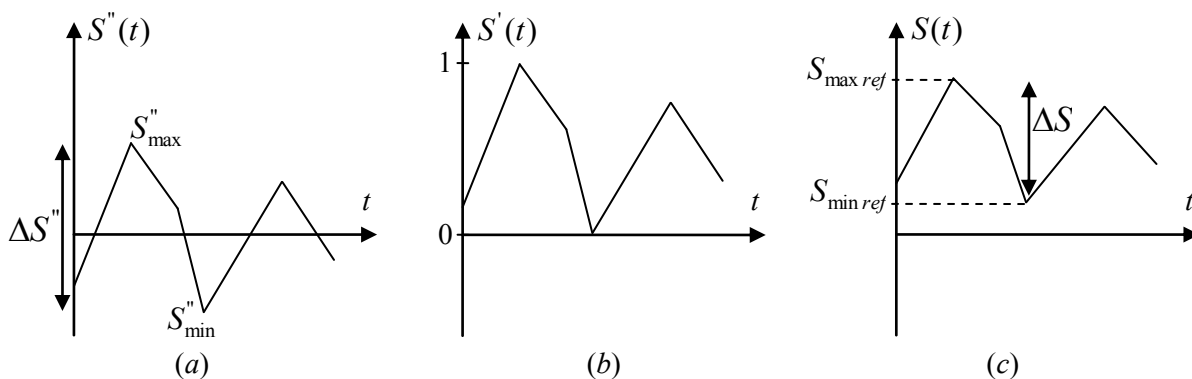
La première étape de la mise à l'échelle consiste à ramener le signal  $S''(t)$  dans l'intervalle  $[0, 1]$ . Nous obtenons le signal  $S'(t)$  de la [Figure E.2.b](#).

$$S'(t) = \frac{S''(t) - S''_{\min}}{\Delta S''} \quad (\text{E.2})$$

Ce signal  $S'(t)$  est ensuite ramené entre les valeurs crêtes de références  $S_{\min \text{ ref}}$  et  $S_{\max \text{ ref}}$  du signal à générer (cf. [Figure E.2.c](#)). Nous obtenons alors le signal final  $S(t)$  défini par :

$$S(t) = S'(t) \times \Delta S + S_{\min \text{ ref}} \quad (\text{E.3})$$

avec  $\Delta S = \text{abs}(S_{\max \text{ ref}}) - \text{abs}(S_{\min \text{ ref}})$

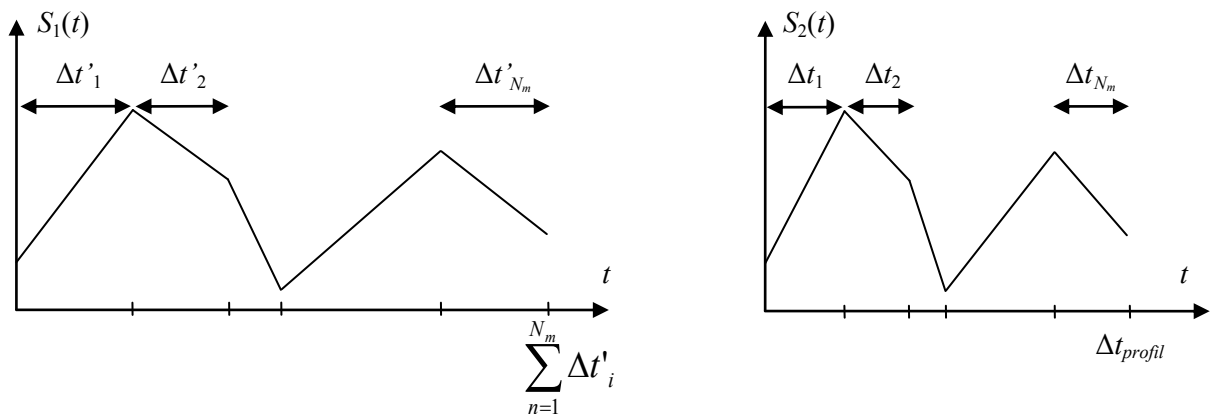


**Figure E.2.** Mise à l'échelle du signal selon l'axe vertical

Nous choisissons de faire varier les paramètres relatifs aux durées  $\Delta t_n$  de chaque motif élémentaire dans l'intervalle  $]0, \Delta t_{\text{profil}}]$ . Encore une fois, ce choix nécessite une opération de

mise à l'échelle selon l'axe temporel pour garantir la durée  $\Delta t_{profil}$ . Soit  $\Delta t'_i$  et  $\Delta t_i$  ( $i = \{1, \dots, N_m\}$ ) les durées de  $N_m$  motifs élémentaires respectivement avant et après la mise à l'échelle (cf. Figure E.3). Le passage du signal  $S_1(t)$  de la Figure E.3.a avant la mise à l'échelle au signal  $S_2(t)$  de la Figure E.3.b après la mise à l'échelle est assuré par la transformation suivante :

$$\Delta t_i = \frac{\Delta t'_i}{\sum_{n=1}^{N_m} \Delta t'_n} \times \Delta t_{profil} \quad (E.4)$$



(a) Signal avant la mise à l'échelle

(b) Signal après la mise à l'échelle

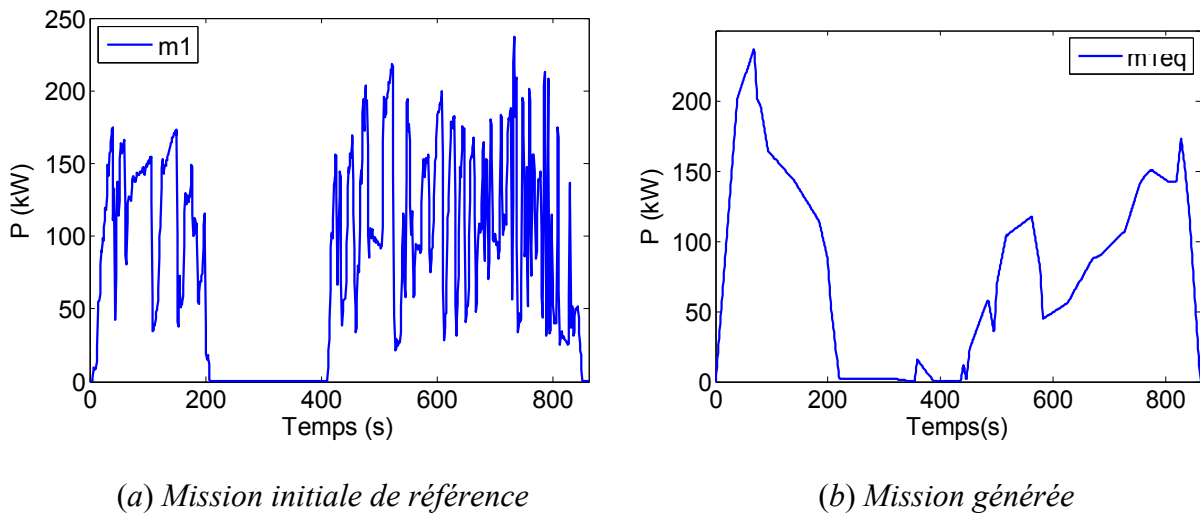
Figure E.3. Mise à l'échelle du signal selon l'axe temporel

## E.2 Application du paramétrage initial à la synthèse des missions ferroviaires

Dans ce paragraphe, nous appliquons la méthode de segment décrite précédemment pour la synthèse d'un profil équivalent à une mission de circulation réelle de la locomotive BB63000. Le profil à générer doit satisfaire l'ensemble des indicateurs de conception caractéristiques de la mission de référence à noter  $P_{max}$ ,  $P_{moy}$ ,  $E_u$ ,  $N_{cyc}$  et  $I_{stat}$ . Pour la validation du modèle de segment, nous considérons que le profil équivalent à générer possède la même durée que la mission ou le profil de référence. Deux exemples de missions ferroviaires sont ainsi traités (mission  $m_1$  et mission  $m_{10}$ ). L'algorithme génétique utilisé est le « Clearing » avec une population de 100 individus et un nombre de générations de 10000. Nous choisissons de faire varier l'amplitude des motifs élémentaires  $\Delta S_n$  entre  $-1$  et  $1$ .

En ce qui concerne le premier exemple, la mission réelle  $m_1$  est représentée par la Figure E.4.a. Le résultat du processus de synthèse de la mission équivalente  $m_{1eq}$  est donné par la Figure E.4.b. Cette dernière est obtenue avec seulement 40 motifs (segments). La

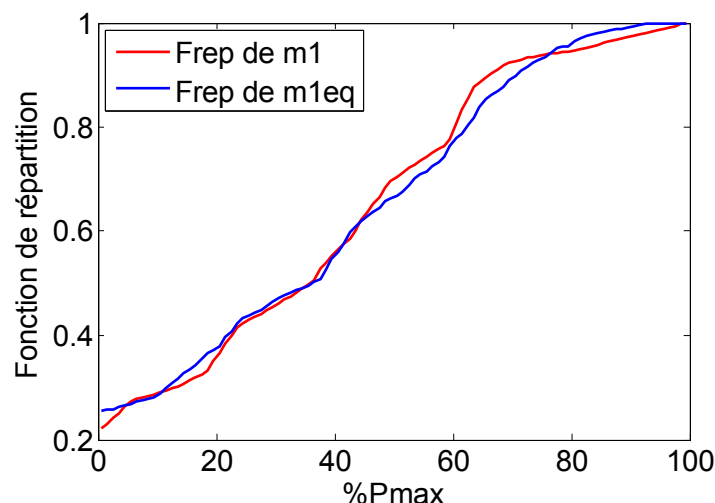
dimension du problème d'optimisation vaut alors 80 paramètres ( $2 \times 40$ ). Le [Tableau E.1](#) donne une comparaison entre les indicateurs de conception de la mission de référence et ceux de la mission régénérée  $m_{1eq}$ . Les fonctions de répartition de la mission de référence  $m_1$  et de la mission  $m_{1eq}$  sont représentées par la [Figure E.5](#). Nous constatons que la mission  $m_{1eq}$  ainsi obtenue respecte tous les indicateurs de référence imposés avec un taux d'erreur moyennement faible. Nous signalons que l'erreur nulle sur l'indicateur de puissance maximale  $P_{max}$  est une conséquence directe de l'opération d'étirement du signal selon l'axe vertical ( $S_{max\ ref} = P_{maxref}$ ).



**Figure E.4.** Résultat du processus de synthèse de mission équivalente

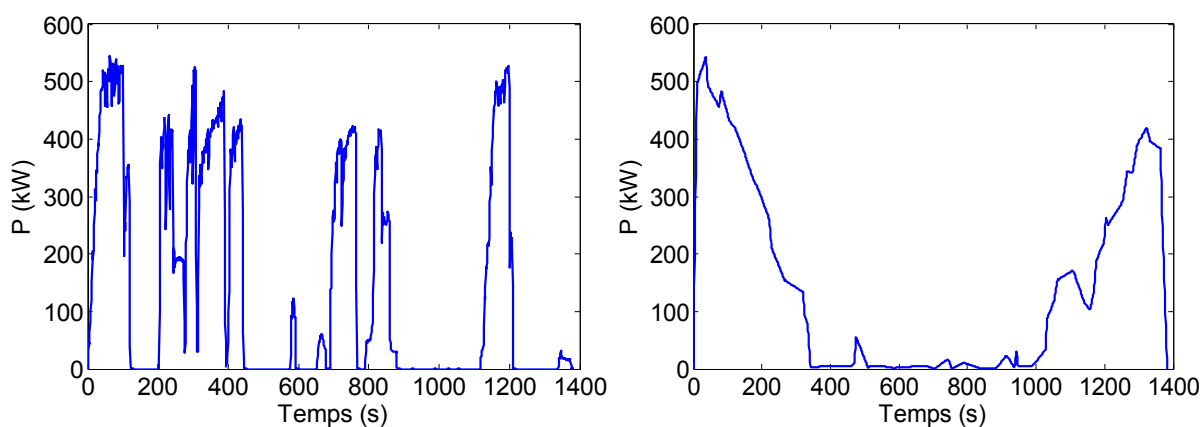
	Indicateurs de référence	Indicateurs de la mission équivalente ( $m_{1eq}$ )	Erreur (%)
$P_{max}$ (kW)	237	237	0.0
$P_{moy}$ (kW)	81.3	80.9	0.5
$E_u$ (kWh)	3.97	3.92	1.3
$Nb_{cycles}$	0.68	0.69	1.5
$I_{stat}$	—	—	3

**Tableau E.1.** Indicateurs de conception de la mission de référence et de la mission générée



**Figure E.5.** Indicateur statistique de la mission de référence et de la mission générée

La mission réelle  $m_{10}$  du deuxième exemple est représentée par la [Figure E.6.a](#). Le résultat du processus de synthèse de la mission équivalente  $m_{10eq}$  est donné par la [Figure E.6.b](#). Le nombre de motifs (segments) utilisé est égal à 60 et par la suite la dimension du problème d'optimisation vaut 120 paramètres ( $2 \times 60$ ). Une comparaison entre les indicateurs de référence et les indicateurs de la mission générée  $m_{10eq}$  est donnée par le [Tableau E.2](#) et la [Figure E.7](#).



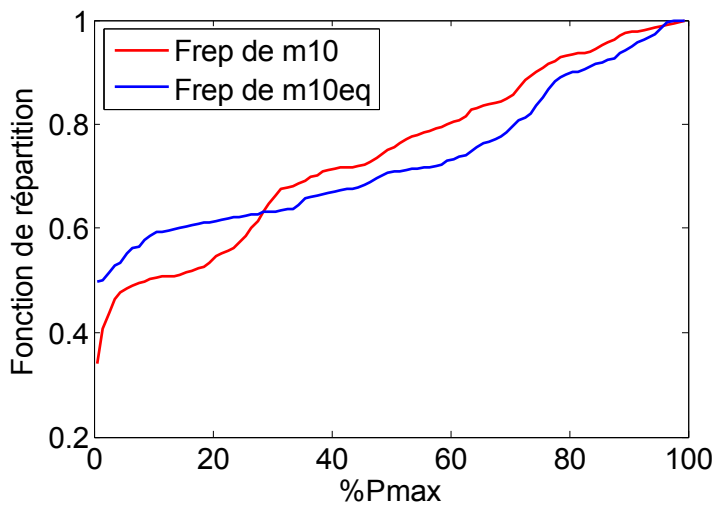
(a) Mission initiale de référence

(b) Mission générée

**Figure E.6.** Résultat du processus de synthèse de mission équivalente

	Indicateurs de référence	Indicateurs de la mission équivalente ( $m_{10eq}$ )	Erreur (%)
$P_{\max}$ (kW)	544	544	0.0
$P_{\text{moy}}$ (kW)	147	142.3	3.2
$E_u$ (kWh)	17.23	16.8	2.5
$Nb_{\text{cycles}}$	0.65	0.67	3.1
$I_{\text{stat}}$	–	–	7.8

**Tableau E.2.** Indicateurs de conception de la mission de référence et de la mission générée



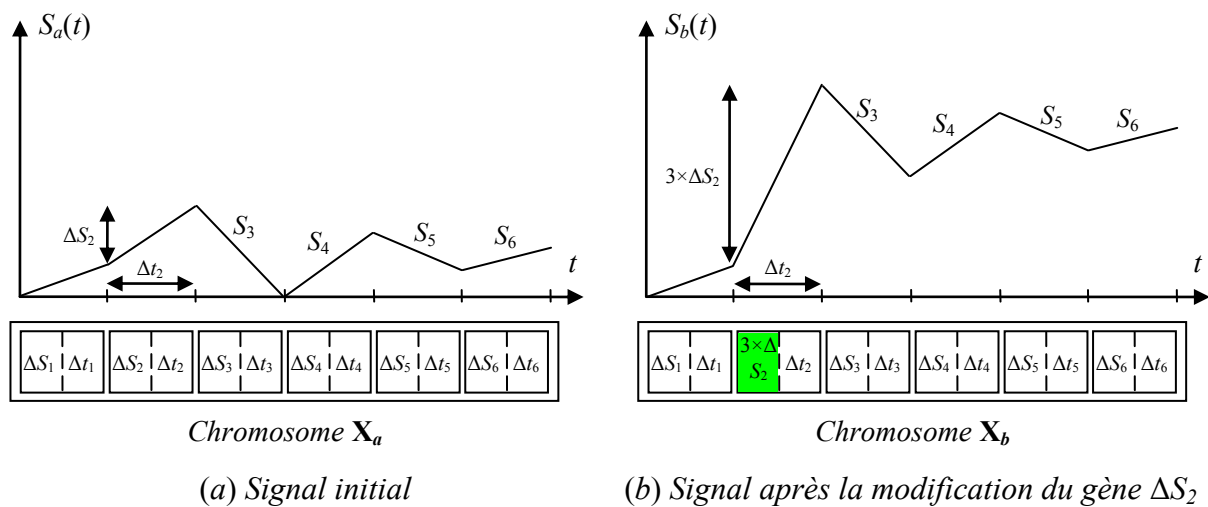
**Figure E.7.** Indicateur statistique de la mission de référence et de la mission générée

Nous constatons que les indicateurs de références ne sont pas parfaitement respectés et le taux d'erreur obtenu est relativement élevé. La [Figure E.7](#) montre un grand écart entre la fonction de répartition de la mission de référence et celle de la mission générée. Dans le paragraphe suivant, nous proposons une modification du paramétrage d'un segment élémentaire afin d'améliorer les performances de cette méthode.

### E.3 Nouveau paramétrage du segment

Les exemples précédents montrent que dans certains cas (par exemple le cas de la mission  $m_{10}$ ), le taux d'erreur entre les indicateurs de référence et ceux du signal généré est relativement élevé. Une analyse plus fine a montré que la limite ainsi constatée de la méthode de segment réside dans la manière avec laquelle nous avons défini les paramètres d'un motif élémentaire. En effet, le fait de définir l'amplitude  $\Delta S_n$  d'un segment comme étant la différence des niveaux de ses deux extrémités, engendre une forte influence des motifs de

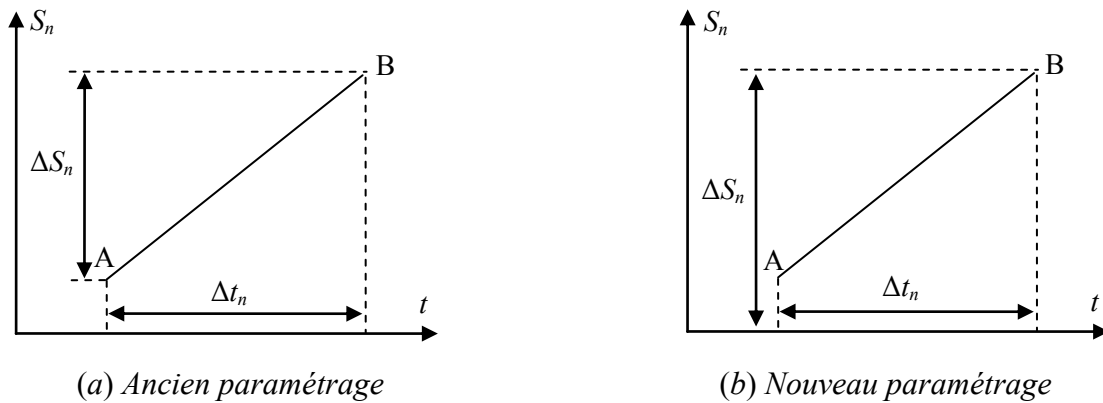
rangs inférieurs sur les motifs de rangs supérieurs. Nous nous référons à la [Figure E.8](#) pour une meilleure explication de ce phénomène. Soit un signal  $S_a$  obtenu par la concaténation de 6 motifs élémentaires et représenté par un chromosome  $\mathbf{X}_a$ . Supposons que suite à l'application d'un opérateur génétique, nous modifions, dans ce chromosome, le gène relatif à l'amplitude du motif de rang 2. Soit  $\mathbf{X}_b$  le chromosome ainsi obtenu et  $S_b$  son signal correspondant. Nous constatons alors que cette opération génétique, même si elle est effectuée d'une manière très locale, modifie toute la partie du signal initial  $S_a$  représentée par les motifs élémentaires de rang supérieur à 2 ( $S_3, \dots, S_6$ ).



**Figure E.8.** Exemple d'illustration de l'impact de paramétrage d'un segment élémentaire

La façon avec laquelle nous avons paramétré un segment élémentaire rend alors non contrôlable le niveau d'amplitude du signal généré par rapport aux différents gènes représentatifs dans le chromosome associé. A cela s'ajoute le phénomène d'étirement de l'amplitude du signal (selon l'axe vertical) qui pourra compliquer d'avantage le problème d'optimisation. Par la suite, nous proposons un deuxième paramétrage du segment plus adéquat. Il s'agit de définir l'amplitude  $\Delta S_n$  d'un segment de rang  $n$  comme étant le niveau de sa deuxième extrémité (point B) par rapport à une référence fixe (cf. [Figure E.9.b](#)). Cette dernière référence est égale à la valeur minimale  $S_{\min ref}$  que l'on désire avoir pour le signal généré. L'amplitude  $\Delta S_n$  d'un segment de rang  $n$  varie donc entre les deux valeurs extrêmes  $S_{\min ref}$  et  $S_{\max ref}$ . De cette façon, la procédure d'étirement vertical du signal n'est plus nécessaire et c'est à l'optimiseur de bien aiguiller les amplitudes  $\Delta S_n$  afin d'atteindre la valeur crête du signal si elle est imposée comme critère d'optimisation. De plus, ce nouveau paramétrage du motif élémentaire (segment) permet d'avoir un découplage entre les motifs.

Effectivement, l'optimiseur peut ajuster les paramètres d'un individu d'une manière locale sans avoir influencer le reste du signal.



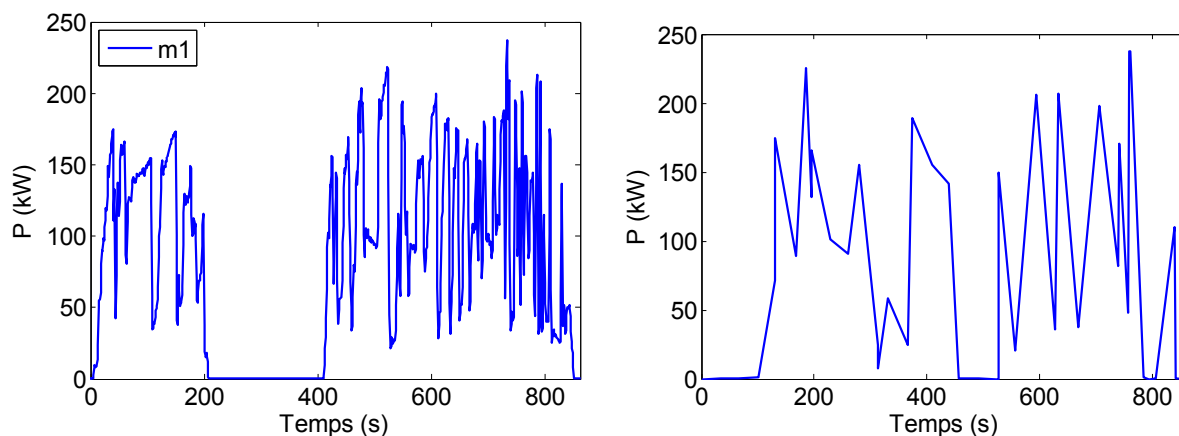
**Figure E.9.** Nouveau paramétrage d'un segment élémentaire

#### E.4 Application du nouveau paramétrage à la synthèse des missions ferroviaires

Au niveau de ce paragraphe, nous reprenons les mêmes exemples de missions ferroviaires ( $m_1$  et  $m_{10}$ ) traités précédemment avec le nouveau paramétrage de la méthode de segment. L'amplitude d'un segment élémentaire  $\Delta S_n$  varie entre  $S_{\min ref} = 0$  et  $S_{\max ref} = P_{\max ref}$ . Nous utilisons le même algorithme génétique (Clearing) avec une population de 100 individus et un nombre de générations égal à 10000. Afin de comparer les deux manières de paramétrage d'un motif élémentaire, nous avons gardé invariant, pour les deux missions testées, le nombre de motifs (40 segment pour  $m_1$  et 60 segments pour  $m_{10}$ ).

Le résultat du processus de synthèse de mission équivalente à la mission réelle  $m_1$  est donné par la [Figure E.10.b](#). Une comparaison entre les indicateurs de conception de la mission de référence  $m_1$  ainsi que celle obtenue est également donnée par le [Tableau E.3](#) et la [Figure E.11](#).

Avec le nouveau paramétrage du motif élémentaire, le signal de mission généré respecte parfaitement l'ensemble des indicateurs de référence. En effet, tous les indicateurs de la mission  $m_{1eq}$ , à part l'indicateur statistique, présentent une erreur nulle ([Tableau E.3](#)). La [Figure E.11](#) montre que la fonction de répartition du signal généré est quasiment confondue avec la fonction de répartition de référence. En effet, l'erreur moyenne entre ces deux fonctions de répartition est égale à 0.5% seulement.

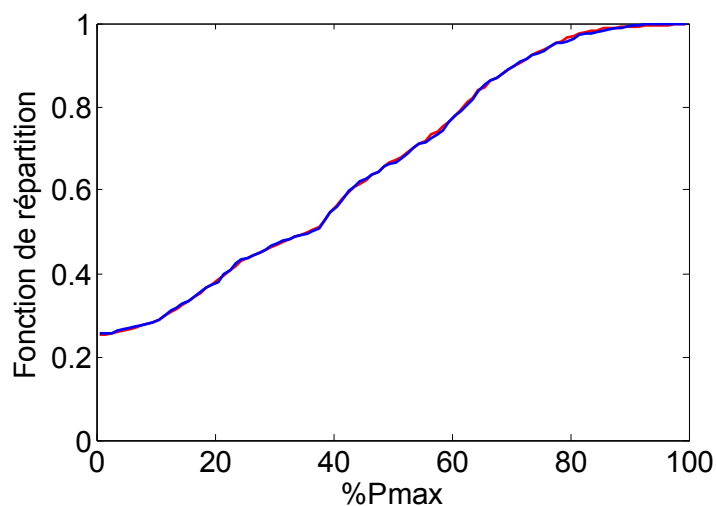


(a) Mission initiale de référence

(b) Mission générée

**Figure E.10.** Résultat du processus de synthèse de mission équivalente

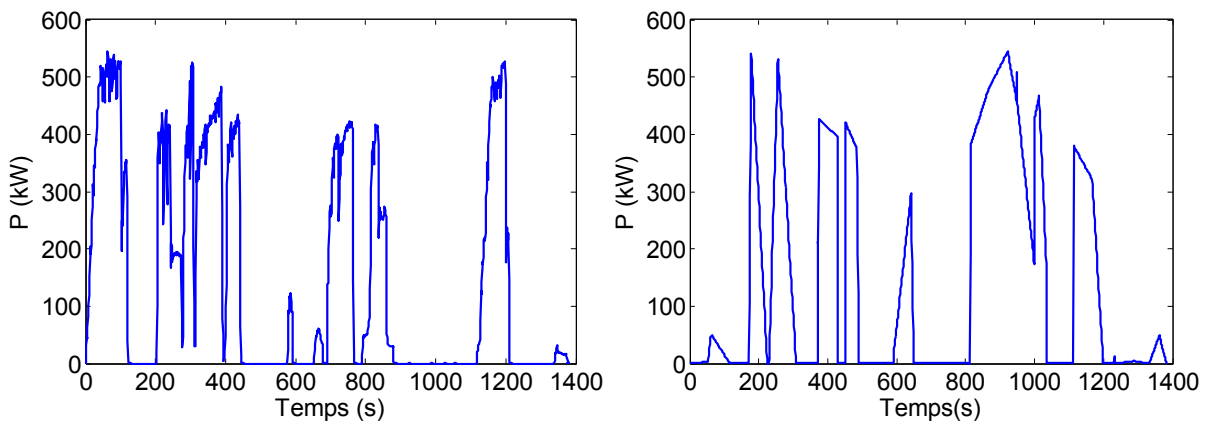
	Indicateurs de référence	Indicateurs de la mission équivalente ( $m_{1eq}$ )	Erreur (%)
$P_{max}$ (kW)	237	237	0.0
$P_{moy}$ (kW)	81.3	81.3	0.0
$E_u$ (kWh)	3.97	3.97	0.0
$Nb_{cycles}$	0.68	0.68	0.0
$I_{stat}$	–	–	0.5

**Tableau E.3.** Indicateurs de conception de la mission de référence et de la mission générée**Figure E.11.** Indicateur statistique de la mission de référence et de la mission générée

Le résultat de synthèse de mission équivalente à la mission  $m_{10}$  est donné par la Figure E.12.a. Le Tableau E.4 ainsi que la Figure E.13 montrent que les indicateurs de



caractérisation de la mission générée  $m_{10eq}$  coïncident avec les valeurs de référence avec une précision très élevée.

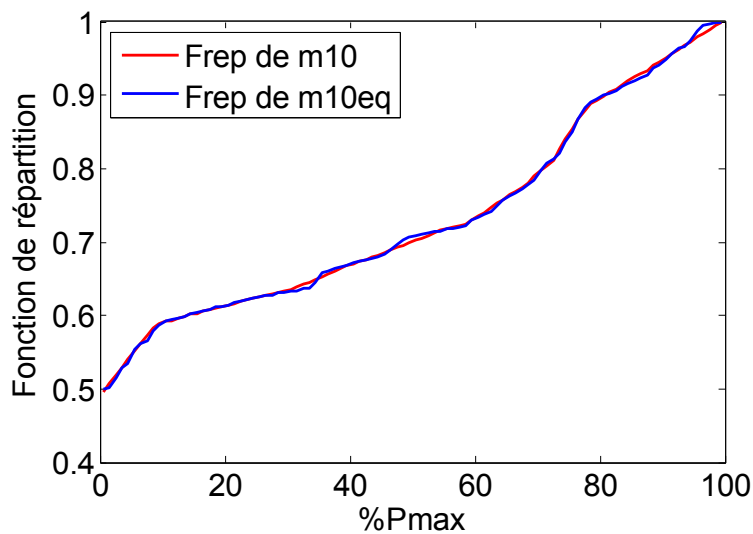


(a) Mission initiale de référence

(b) Mission générée

**Figure E.12.** Résultat du processus de synthèse de mission équivalente

	Indicateurs de référence	Indicateurs de la mission équivalente ( $m_{10eq}$ )	Erreur (%)
$P_{\max}$ (kW)	544	544	0.0
$P_{\text{moy}}$ (kW)	147	147	0.0
$E_u$ (kWh)	17.23	17.23	0.0
$Nb_{\text{cycles}}$	0.65	0.65	0.0
$I_{\text{stat}}$	—	—	0.4

**Tableau E.4.** Indicateurs de conception de la mission de référence et de la mission générée**Figure E.13.** Indicateur statistique de la mission de référence et de la mission générée

En comparant les résultats de synthèse de mission équivalente avec les deux techniques de paramétrage, nous montrons que les performances de la méthode de segment sont nettement meilleures avec le nouveau paramétrage.



## Annexe F

### Choix de la puissance nominale du groupe Diesel

Comme nous l'avons signalé dans le chapitre 3, le choix de la puissance moyenne de référence  $P_{\text{moy ref}}$  pour la synthèse d'une mission représentative n'est pas évident. C'est en fait cette valeur de puissance qui définit *a priori* la puissance nominale du groupe Diesel, valeur pour laquelle nous estimons que la consommation du gasoil et par la suite l'émission des particules polluantes sont optimales. Au niveau d'une classe de missions, fixer  $P_{\text{moy ref}}$  à la plus grande valeur des puissances moyennes obtenues pour chaque mission ( $P_{\text{moy ref}} = \max(P_{\text{moy } i})$ ) conduit à un surdimensionnement du groupe Diesel et donc à une faible taille de stockage. En effet, dans ce cas, le groupe Diesel ne fonctionne à sa puissance nominale que lorsqu'il s'agit de la mission de puissance moyenne  $P_{\text{moy ref}}$ . Pour le reste des missions, la consommation et la pollution ne sont alors plus optimales. Une deuxième solution consiste à fixer  $P_{\text{moy ref}}$  à la valeur moyenne d'une mission totale définie par la concaténation de toutes les missions. Le groupe Diesel est mieux utilisé sur l'ensemble mais cette configuration nécessite une taille énergétique du stockage (typiquement des batteries) plus importante.

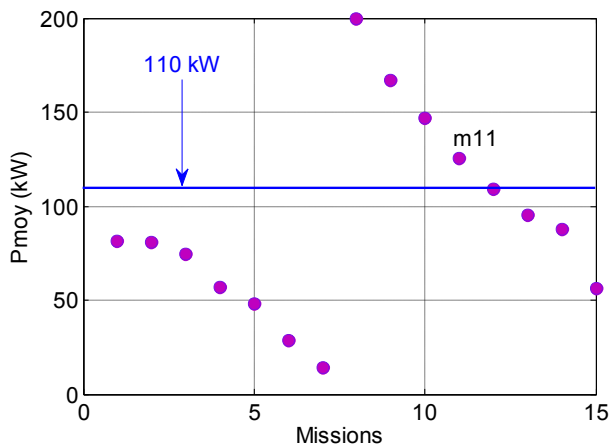
Dans cette annexe, nous étudions à travers deux exemples de missions de circulation réelles de deux locomotives Diesel candidates à une hybridation par des batteries Ni-Cd, l'évolution du coût de possession annuel  $C_f$  de la locomotive en fonction de la puissance nominale du groupe Diesel  $P_{GDN}$ .

$$C_f = C_{GD} + C_{BT} + C_{fuel} \quad (\text{F.1})$$

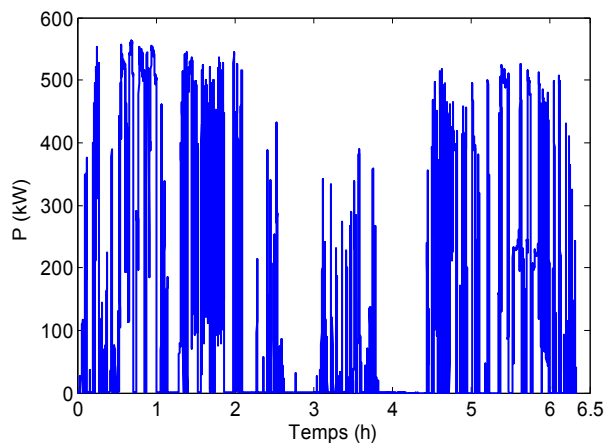
Ces deux exemples sont particulièrement intéressants par le fait qu'en fixant la puissance nominale du groupe Diesel entre la valeur moyenne sur la concaténation de toutes les missions et la plus grande des valeurs moyennes de chacune de ces missions, nous disposons d'une seule mission réelle et dominante sur le plan de la puissance maximale  $P_{\max}$  et de l'énergie utile  $E_u$ . Cette mission sera alors considérée comme mission dimensionnante et typique pour l'évaluation du coût de possession annuel des locomotives hybrides.

### F.1 Cas de la locomotive BB63000

Dans cet exemple, nous disposons de 15 missions dont les puissances moyennes varient entre 15 kW et 200 kW (cf. Figure F.1). La valeur moyenne sur la concaténation de toutes les missions est égale à 110 kW. Quelle que soit la puissance nominale du groupe Diesel entre 110 kW et 200 kW, la mission  $m_{11}$  (cf. Figure F.2.) est dominante sur le plan de la puissance maximale  $P_{\max}$  et de l'énergie utile  $E_u$ .

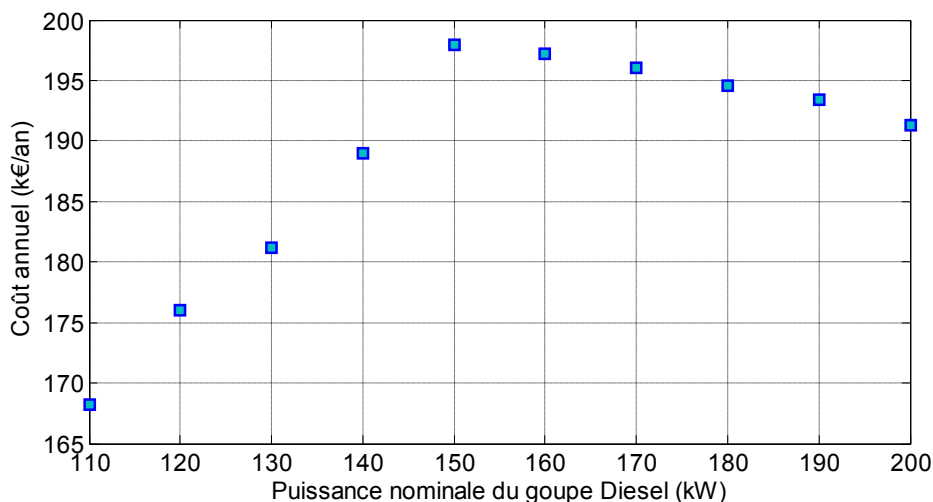


**Figure F.1.** Répartition des puissances moyennes des 15 missions



**Figure F.2.** Mission dimensionnante  $m_{11}$

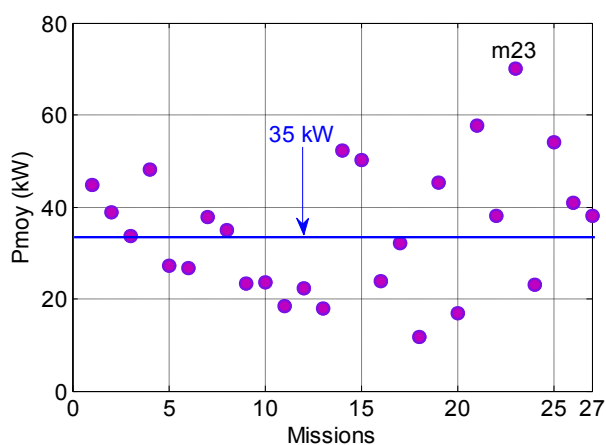
Pour différentes valeurs de puissance nominale du groupe diesel, nous dimensionnons alors la locomotive hybride à partir de la mission  $m_{11}$ . L'évolution du coût annuel de la locomotive hybride en fonction de la puissance nominale du groupe Diesel est donnée par la Figure F.3. Cette dernière montre que le coût de possession annuel minimal est obtenu avec la taille du groupe Diesel la plus faible (110 kW). Positionner  $P_{\text{moy ref}}$  à la valeur moyenne de la concaténation de l'ensemble des missions (110 kW) nous offre un gain de 25 k€ par an par rapport à la plus grande puissance moyenne des missions.



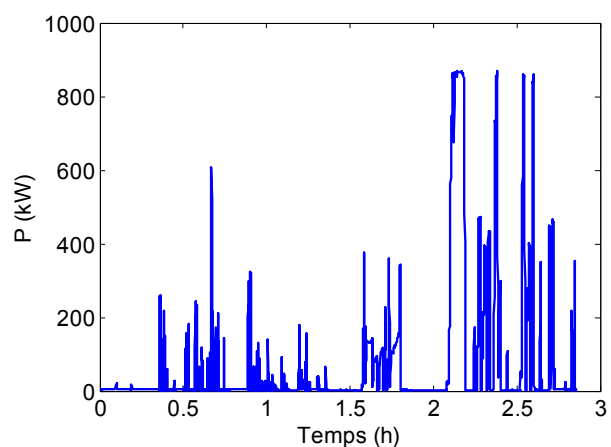
**Figure F.3.** *Évolution du coût de possession annuel de la locomotive en fonction de la puissance nominale du groupe Diesel*

## F.2 Cas de la locomotive BB460000

Le deuxième exemple contient les 27 missions de la locomotive BB460000, dont les puissances moyennes varient entre 12 kW et 70 kW (cf. Figure F.4). La valeur moyenne sur la concaténation de toutes les missions vaut 35 kW. La mission  $m_{23}$  (cf. Figure F.5.) est dominante sur le plan de la puissance maximale  $P_{\max}$  et le plan de l'énergie utile  $E_u$  pour tous les groupes Diesel de puissance nominale supérieure à 35 kW. L'évolution du coût de possession annuel de la locomotive hybride en fonction de la puissance nominale du groupe Diesel (cf. Figure F.6) montre, encore une fois, l'intérêt de dimensionner le groupe Diesel et donc de choisir la puissance moyenne de référence  $P_{\text{moy ref}}$  à la valeur moyenne sur la concaténation de toutes les missions.

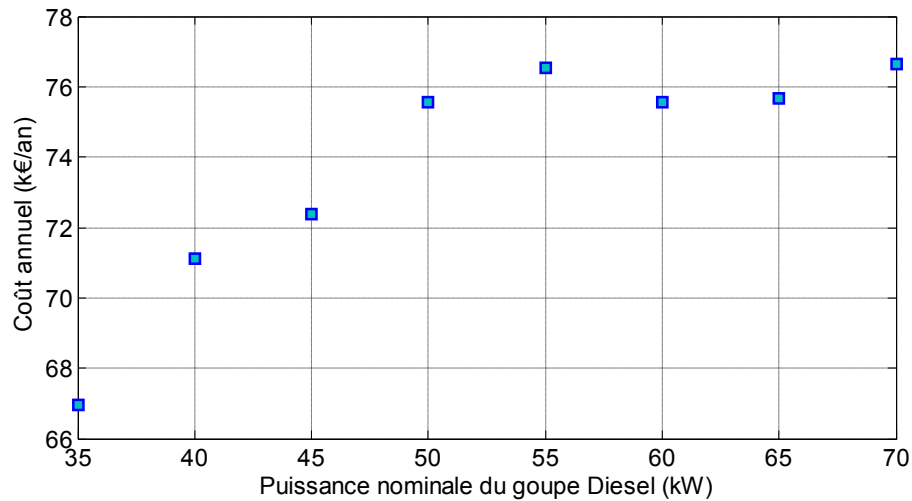


**Figure F.4.** *Répartition des puissances moyennes des 27 missions*



**Figure F.5.** *Mission dimensionnante  $m_{23}$*

Enfin, l'étude des deux exemples précédents de missions ferroviaires, nous a permis de conclure que la conception d'une locomotive hybride avec un groupe Diesel dimensionné à la puissance moyenne sur la concaténation de toutes les missions, est la plus avantageuse en termes de coût de possession annuel. Nous tiendrons compte de ce résultat pour le reste de ce chapitre.



**Figure F.6.** *Évolution du coût annuel de la locomotive en fonction de la puissance nominale du groupe Diesel*

# Bibliographie

- [ABD 07] ABDELLI A., Optimisation multicritère d'une chaîne éolienne « passive », Thèse de doctorat, Université de Toulouse, 2007.
- [AKL 07] AKLI C.R., ROBOAM X., SARENI B., JEUNESSE A., « Energy management and sizing of a hybrid locomotive », 12<sup>th</sup> *European Conference on Power Electronics and Applications (EPE'2007)*, Aalborg, Denmark, septembre 2007.
- [AKL 08] AKLI C.R., Conception systémique d'une locomotive hybride autonome, Thèse de doctorat, Université de Toulouse, 2008.
- [AKL 09] AKLI C.R., ROBOAM X., SARENI B., JEUNESSE A., « Integrated optimal design of a hybrid locomotive with multiobjective genetic algorithms », *International Journal of Applied Electromagnetics and Mechanics*, vol 4, n° 3-4, p. 151-162, 2009.
- [AND 97] ANDRE M., Driving patterns analysis and driving cycles, within the project : European Development of Hybrid Technology approaching efficient Zero Emission Mobility (HYZEM), Rapport INRETS, n° LEN 9709, p. 47, Bron, France.
- [AND 04] ANDRE M., « ARTEMIS European driving cycles for measuring car pollutant emissions », *Science of the Total Environment*, vol 334-335, p. 73-84, 2004.
- [AND 05] ANDRE M., Vehicule emission measurement collection of the ARTEMIS database, Rapport INRETS, n° LTE 0504, janvier 2005.
- [AST 03] ASTIER S., Des machines à aimants permanents aux systèmes énergétiques autonomes. Systémique, électricité et développement durable, HDR, Université de Toulouse, 2003.
- [BAE 08] BAEK S.H., CHO S.S., JOO W.S., « Fatigue life prediction based on the Rainflow cycle counting method for the end beam of a freight car bogie », *International Journal of Automotive Technology*, vol 9, n° 1, p. 95-101, 2008.
- [BAL 65] BALL G., HALL D., ISODATA a novel method of data analysis and pattern classification, Technical Report, Stanford Research Institute, 1965.
- [BER 73] BERTALANFFY L.V (1968), Théorie générale des systèmes, Traduction Jean Benoît Chabrol, Paris, Dunod, 1973.
- [BOL 03] BOLSHAKOVA N., AZUAJE F., « Cluster validation techniques for genome expression data », *Signal Processing*, vol 83, n° 4, p. 825-833, 2003.



- [BOS 04] BOSCH, Automotive Handbook, 6<sup>ème</sup> édition, Bentley Publishers, 2004.
- [CAL 74] CALINSKI R.B., HARABAZ J., « A dendrite method for cluster analysis », *Commun. Statistics*, vol 3, p. 1-27, 1974.
- [CAL 04] CALIF R., BLONBOU R., DESHAIES B., « Wind velocity measurements analysis for time scales smaller than 1 hour : Application to wind energy forecasting », *Proceedings of the 24<sup>th</sup> AIAA/ASME Wind Energy Symposium*, 2004.
- [CAL 08] CALIF R., EMILION R., SOUBDHAN T., BLONBOU R., « Wind speed PDF classification using Dirichlet mixtures », *Wind Energy*, 2008.
- [CHA 99] CHAPOULIE P., Modélisation systémique pour la conception de véhicules électriques multisources. Application aux véhicules équipés de générateurs photovoltaïques ou de supercondensateurs, Thèse de doctorat, Université de Toulouse, 1999.
- [DAV 79] DAVIES D.L., BOULDIN D.W., « A cluster separation measure », *IEEE Trans. Pattern Anal. Machine Intell.*, vol PAMI-1, p. 224-227, 1979.
- [DID 72] DID E., « Optimisation en classification automatique et reconnaissance des formes », *Revue Française de Recherche Opérationnelle*, vol 3, p. 61-96, 1972.
- [DID 74] DID E., « Classification automatique séquentielle pour grands tableaux », *Revue Française de Recherche Opérationnelle*, vol 9, p. 1-21, 1974.
- [DRO 97] DROUILHET S., JOHNSON B.L., « A battery life prediction method for the hybrid power applications », 35<sup>th</sup> AIAA Aerospace Sciences Meeting and Exhibit, Reno, Etats-Unis, janvier 1997.
- [DUF 08] DUF-LOPEZ R., BERNAL-AGUSTIN J.L., « Multi-objective design of PV–wind–diesel–hydrogen–battery systems », *Renewable Energy*, p. 2559-2572, april, 2008.
- [EEC 99] EEC Directive 90/C81/01, Emission Test Cycles for the Certification of light duty vehicles in Europe, EEC Emission Cycles, 1999. Available at : <http://www.dieselnet.com/standards/cycles>
- [EHS 99] EHSANI M., GAO Y., BUTLER K., « Application of Electrically Peaking Hybrid (ELPH) Propulsion System to a Full-Size Passenger Car with Simulated Design verification », *IEEE Trans. on Vehicular Technology*, vol 48, n° 6, p. 1779-1787, 1999.
- [FOG 91] FOGEL D.B., « System identification trough simulated evolution », *Ginn Press*, Needham Heights, 1991.
- [FOR 65] FORGY E.W., « Cluster analysis of multivariate data : efficiency versus interpretability of classifications », *Biometric Society Meetings*, Riverside, California, 1965.
- [FPT] see <http://www.fptpowertrain.com/eng/home.htm>
- [FRÄ 06] FRÄNTI P., VIRMAJOKI O., « Iterative shrinking method for clustering problems », *Pattern Recognition*, vol 39, n° 5, p. 761-765, 2006.

- [GOL 87] GOLDBERG D.E., RICHARDSON J., « Genetic algorithms with sharing for multimodal function optimization », *Proceedings of the 2<sup>nd</sup> International Conference on Genetic Algorithms*, p. 41-49, 1987.
- [GOL 89] GOLDBERG D.E., *Genetic Algorithms in Search Optimization & Machine Learning*, Addison Wesley, 1989.
- [GRA 06] GRAHAM L.A., « Greenhouse gas emissions from light duty vehicles under a variety of driving conditions », in *Proc. IEEE EIC Climate Change Technol.*, p. 1-8, 2006.
- [GRI 84] GRIFFITHS A., ROBINSON L.A., WILLETT P., « Hierarchic Agglomerative Clustering Methods for Automatic Document Classification », *Journal of Documentation*, vol 40, p. 175-205, 1984.
- [HAR 95] HARIK G., « Finding Multimodal Solutions Using Restricted Tournament Selection », *Proceedings of the Sixth International Conference on Genetic Algorithms*, San Francisco, CA, USA : Morgan Kaufmann, 1995.
- [HAU 00] HAU E., *Wind Turbines – Fundamentals Technologies Application Economics*, Second edition, Springer, 2000.
- [HIR 89] HIROSHI I., « Spectra of surface wind speed and air temperature over the ocean in the mesoscale frequency range in JASIN-1978 », *Boundary-Layer Meteorology*, vol 47, n° 1-4, p. 71-84, 1989.
- [HOL 75] HOLLAND J., *Adaptation in Natural and Artificial Systems*, the University of Michigan Press, Ann Arbor, 1975.
- [HOP] see <http://www.hoppecke.com>
- [JAA 09a] JAAFAR A., AKLI C.R., SARENI B., ROBOAM X., JEUNESSE A., « Sizing and Energy Management of a Hybrid Locomotive Based on Flywheel and Accumulators », *IEEE Transactions on Vehicular Technology*, vol 58, n° 8, october 2009.
- [JAA 09b] JAAFAR A., SARENI B., ROBOAM X., Hybridation des missions de la locomotive BB460000 sur la zone FRET de SOTTEVILLE et LE HAVRE, Rapport de contrat LAPLACE-SNCF, Tâche 1 : identification et étude des concepts d'hybridation, 2009.
- [JAA 10a] JAAFAR A., SARENI B., ROBOAM X., « Signal synthesis by means of evolutionary algorithms », *selected paper from OIPE'2010, International Journal for Inverse Problems in Science and Engineering*, 2011.
- [JAA 10b] JAAFAR A., SARENI B., ROBOAM X., « Clustering analysis of railway driving missions with niching genetic algorithms », *selected paper from OIPE'2010, International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, 2011.
- [JAA 10c] JAAFAR A., SARENI B., ROBOAM X., THIOUNN M., « Traitement de la mission pour l'éco conception d'une locomotive hybride par optimisation », *Conférence Francophone sur l'Eco-conception en Génie Electrique (ConFrEGE 2010)*, Toulouse, France, 2010.

- [JAA 10d] JAAFAR A., SARENI B., ROBOAM X., THIUNN M., « Sizing of a Hybrid Locomotive Based on Accumulators and Ultracapacitors », *IEEE Vehicle Power and Propulsion Conference (VPPC 2010)*, Lille, France, 2010.
- [JAI 88] JAIN A., DUBES R., *Algorithms for Clustering Data*, Prentice Hall Advanced Reference Series, Upper Saddle River, USA, 1988.
- [JAI 99] JAIN A.K., MURTY M.N., FLYNN P.J., « Data Clustering : A Review », *ACM Computing Surveys*, vol 31, n° 3, p. 264-323, 1999.
- [JOA 07] JOAQUIN M.A., ANGEL A.B., « Wind Power Variability Model Part I– Foundations », *9<sup>th</sup> International Conference Electrical Power Quality and Utilisation*, Barcelona, october 2007.
- [JOH 01] JOHNSON G.L., *Wind energy systems*, Electronic edition, december, 2001.
- [JOH 67] JOHNSON S.C., « Hierarchical clustering schemes », *Psychometrika*, vol 32, p. 241-254, 1967.
- [JOU 10] JOURDAIN F., « Phénoménologie et modélisation des transferts de polluants dans l'atmosphère », *Journées de modélisation CNRS – CEA Cadarache*, Saint-Paul-lez-Durance, France, février 2010.
- [JOU 99] JOURMARD R., ANDRE M., VIDON R., TASSEL P., PRUVOST C., Influence du cycle de conduite sur les émissions unitaires de polluants des voitures particulières, Rapport INRETS, n° LTE 9902, décembre 1999.
- [KAU 90] KAUFMAN L., ROUSSEEUW P.J., « Finding Groups in Data : An Introduction to Cluster Analysis », New York : Wiley, 1990.
- [KEI 02] KEIJZER M., MERLO J.J, ROMERO G., SCHOENAUER M., « Evolving objects: a general purpose evolutionary computation library », *Evolution Artificielle (EA-01)*, LCNS 2310, 2002.
- [KEL 92] KELLER J.K., « Simulation of Wind with „K’ Parameter », *Wind Engineering*, vol 16, n° 6, p. 307-312, 1992.
- [KRE 09] KREUWAN S., *Modelling and optimal design in railway applications*, Thèse de doctorat, École Centrale de Lille, 2009.
- [LAN 06] LANGLOIS O., *Conception d’un réseau de secours électrique pour l’aéronautique*, Thèse de doctorat, Université de Toulouse, 2006.
- [LEB 00] LEBART L., MORINEAU A., TABARD N., *Statistique Exploratoire Multidimensionnelle*, Dunod, Paris, 2000.
- [LEM 94] LEMOIGNE J.L, *La théorie du système général – théorie de la modélisation*, PUF, Paris, 1996.
- [LEM 95] LEMOIGNE J.L, *La modélisation des systèmes complexes*, Dunod, 1995.
- [LEM 96] LEMOIGNE J.L, *Sciences des systèmes*, Encyclopedia Universalis, 1996.

- [MAC 67] MACQUEEN J., « Some methods for classification and analysis of multivariate observations », in *Proc. 5<sup>th</sup> Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, California, 1967.
- [MAL 00] MALLAT S., *Une exploration des signaux en ondelettes*, les éditions de l'école polytechnique, Palaiseau, 2000.
- [MAR 08] MARTINE O.L. HANSEN, *Aerodynamics of Wind Turbines*, Second edition published by Earthscan in the UK and USA, 2008.
- [MAU 00] MAULIK U., BANDYOPADHYAY S., « Genetic algorithm-based clustering technique », *Pattern Recognition*, vol 33, n° 9, p. 1455-1465, 2000.
- [MES 98] MESSINE F., NOGAREDE B., LAGOUANELLE J.L., « Optimal design of electromechanical actuators: a new method based on global optimization », *IEEE Transactions on Magnetics*, vol 34, n° 1, p. 299-308, 1998.
- [MES 06] MESTER V., GILLON F., BRISSET S., BROCHET P., « Global optimal design of a wheel traction motor by a systemic approach of the electric drive train », *IEEE Vehicle Power and Propulsion Conference (VPPC 2006)*, Windsor, UK, 2006.
- [MES 08] MESTER V., GILLON F., BRISSET S., BROCHET P., MIRCEA M.R., « Multidisciplinary Design Optimisation applied to an Electric Drive train », *International Review of Electrical Engineering (IREE)*, vol 3, n° 4, p. 749-757, 2008.
- [MIC 96] MICHALEWICZ Z., *Genetic Algorithms + Data Structures = Evolution Programs*, 3<sup>ème</sup> édition, Springer-Verlag, Berlin Heidelberg New York, 1996.
- [MIL 85] MILLIGAN G., COOPER M., « An examination of procedures for determining the number of clusters in a data set », *Psychometrika*, vol 50, n° 2, p. 159-179, 1985.
- [MON 00] MONMARCHE N., *Algorithmes de fourmis artificielles : applications à la classification et à l'optimisation*, Thèse de doctorat, Université François Rabelais Tours, 2000.
- [MOU 09] MOUSSOUNI-MESSAD F., *Méthodologie et algorithmes adaptés à l'optimisation multi-niveaux et multi-objectif de systèmes complexes*, Thèse de doctorat, Ecole Centrale de Lille, 2009.
- [NGU 08] NGUYEN-HUU H., SARENI B., WURTZ F., RETIÈRE N., ROBOAM X., « Comparison of self-adaptive evolutionary algorithms for multimodal optimization », 10<sup>th</sup> *International Workshop on Optimization and Inverse Problems in Electromagnetism (OIPE'2008)*, Ilmenau, Germany, septembre 2008.
- [PAL 95] PAL N.R., BEZDEK J.C., « On cluster validity for the fuzzy c-means model », *IEEE Trans. Fuzzy Syst.*, vol 3, n° 3, p. 370-379, 1995.
- [PET 96] PETROWSKI A., « A clearing procedure as a niching method for genetic algorithms », *Proceedings of the IEEE International Conference on Evolutionary Computation*, Nagoya, Japan, p. 798-803, mai 1996.
- [PLA 09] « PLATHEE locomotive hybride », Descriptif technique n° 02-0352, février 2009.

- [REC 73] RECHENBERG I., « Evolution Strategies – Optimierung technischer Systeme nach Prinzipien der biologischen Evolution », *Frommann-Holzboog*, Stuttgart, 1973.
- [REG 03] REGNIER J., Conception de systèmes hétérogènes en Génie Électrique par optimisation évolutionnaire multicritère, Thèse de doctorat, Université de Toulouse, 2003.
- [RIC 04] RICARDEAU F., Sûreté de fonctionnement en électronique de puissance, HDR, Université de Toulouse, 2004.
- [ROB 00] ROBOAM X., De la modélisation à la commande et à la caractérisation des Systèmes Electromécaniques à propulsion asynchrone, HDR, Université de Toulouse, 2000.
- [ROS 75] ROSNAY J., *Le microscope, vers une vision globale*, Seuil, 1975.
- [ROU 87] ROUSESEUW P.J., « Silhouettes : a graphical aid to the interpretation and validation of cluster analysis », *Journal of Computational and Mathematics*, vol 20, n° 1, p. 53-65, 1987.
- [SAR 98] SARENI B., KRÄHENBÜHL L., « Fitness sharing and niching methods revisited », *IEEE Trans. on Evolutionary Computation*, vol 3, n° 3, p. 97-106, 1998.
- [SAR 99] SARENI B., Méthodes d'optimisation multimodales associées à la modélisation numérique en électromagnétisme, Thèse de doctorat, École Centrale de Lyon, 1999.
- [SAR 06] SARENI B., Conception simultanée par optimisation des systèmes d'énergie électrique, HDR, Université de Toulouse, 2006.
- [SCH 65] SCHWEFEL H.P., *Kybernetische Evolution als Strategie der experimentellen Forschung in der Strömungstechnik*, Diplomarbeit, Technische Universität Berlin, 1965.
- [SCH 77] SCHWEFEL H.P., « Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie », *Interdisciplinary Systems Research*, vol 26, Birkhäuser, Basel, 1977.
- [SCH 87] SCHWEFEL H.P., « Collective phenomena in evolutionary systems », *In Preprints of the 31<sup>st</sup> Annual Meeting of the International Society for General System Research*, Budapest, vol 2, p. 1025-1033, 1987.
- [SCH 95] SCHWEFEL H.P., « Evolution and Optimum seeking », Wiley, 1995.
- [SHE 05] SHENG W., SWIFT S., ZHANG L., LIU X., « A weighted sum validity function for clustering with a hybrid niching genetic algorithm », *IEEE Transactions on Systems, Man, and Cybernetics-Part B : Cybernetics*, vol 35, n° 6, p. 1156-1167, 2005.
- [SNE 57] SNEATH P.H.A., « The application of computer to taxonomy », *Journal General Microbiology*, vol 17, p. 201-226, 1957.
- [STO 96] STORN R., PRICE K., « Minimizing the real functions of the ICEC'96 contest by Differential Evolution », *International Conference on Evolutionary Computation*, Nagoya, Japan, 1996.

- [STR 90] STRARRON D.A., STENGELZ R.F., « Stochastic Prediction Techniques for Wind Shear Hazard Assessment », *Journal of Guidance, Control, and Dynamics*, vol 15, n° 5, p. 1224-1229, 1990.
- [THI 08] THIOUNN M., « PLATHEE – A Platform for Energy Efficiency and Environmentally Friendly Hybrid Trains », 8<sup>th</sup> *World Congress on Railway Research*, COEX, Seoul, Korea, may, 2008.
- [TRA10] TRAN D.H., Conception Optimale Intégrée d'une chaîne éolienne « passive » : Analyse de robustesse, validation expérimentale, Thèse de doctorat, Université de Toulouse, 2010.
- [WAR 63] WARD J.H., « Hierarchical Grouping to Optimize and Objective Function », *Journal of the American Statistical Association*, vol 58, p. 236-244, 1963.
- [WIN 06] see <http://www.wind-works.org/articles/IECWindTurbineClasses.html>
- [WON 82] WONG M.A., « A hybrid clustering method for identifying high density clusters », *Journal of the American Statistical Association*, vol 77, p. 841-847, 1982.
- [XU 05] XU R., WUNSCH 2 D.C., « Survey of Clustering Algorithms », *IEEE Trans. on Neural Networks*, vol 16, n° 3, p. 645-678, 2005.