ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

MANUSCRIPT-BASED THESIS
PRESENTED TO
ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

IN PARTIAL FULFILLEMENT OF THE REQUIREMENTS FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
Ph. D.

BY
Guillaume CORRIVEAU

IMPROVING EVOLUTIONARY ALGORITHMS BY MEANS OF AN ADAPTIVE PARAMETER CONTROL APPROACH

MONTREAL, DECEMBER 7 2012

Guillaume Corriveau, 2012

This Creative Commons licence allows readers to download this work and share it with others as long as the author is credited. The content of this work can't be modified in any way or used commercially.

BOARD OF EXAMINERS (THESIS PH.D.)

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS

Mr. Raynald Guilbault, Thesis Supervisor
Mechanical Engineering Department at École de technologie supérieure

Mr. Antoine Tahan, Thesis Co-supervisor
Mechanical Engineering Department at École de technologie supérieure

Mr. Robert Sabourin, Thesis Co-supervisor
Automated Production Engineering Department at École de technologie supérieure

Mr. Simon Joncas, President of the Board of Examiners
Automated Production Engineering Department at École de technologie supérieure

Mr. Tony Wong, Member of the jury
Automated Production Engineering Department at École de technologie supérieure

Mr. Sébastien Adam, External Evaluator
LITIS laboratory
Université de Rouen

THIS THESIS WAS PRENSENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND PUBLIC

ON OCTOBER 30, 2012

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

FOREWORD

Since this research is shaped as a manuscript-based thesis, there is some duplication across chapters. In fact, four of the five chapters have already been submitted as self-content papers. Nonetheless, special effort has been paid throughout the writing of these manuscripts to ensuring continuity to the best extent possible. Consequently, from the author's perspective, these repetitions do not detract from the purpose of the thesis, but rather help to strengthen important concepts.

Guillaume Corriveau, (September 4, 2012)

ACKNOWLEDGMENT

I would like to thanks my thesis advisor, Raynald Guilbault, Antoine Tahan and Robert Sabourin, for their technical, financial, and moral support throughout these years. Their guidance was always fruitful and their interest in this research project always encouraged me to push my own limit. I am also grateful for the freedom they provided to me. Many thank to Simon Joncas, Tony Wong, and Sébastien Adam for their participation to the board of examiners.

Also, I am grateful to the Fonds de recherche du Québec – Nature et technologies (FQRNT), Rio Tinto Alcan, and the École de technologie supérieure for their financial contribution to this research project.

Special thank go to Cynthia Orman for her assistance on the language correction. I am beholden to her as well as to Raynald Guilbault for the language quality of this thesis.

In addition, I would like to express my gratitude to my employer, Bombardier aerospace, and more specifically to my immediate superior, Franck Dervault, for the constant support in the last two years. They have shown understanding that led me the required flexibility to successfully complete this research project.

I would like to thank my parents who inspired me with their guidance and open mind. Finally, I am indebted to my beloved partner, Marjorie Leduc, for her patience, support, and encouragement during all these years. I cannot express in word all the gratitude I have regarding the sacrifices she made to follow me in this very demanding adventure. Thank to my daughter, Anaïs, who gave me the motivation during hard days. Together, they have skillfully succeeded to keep me in touch with real life.

With the hope that I will now be able to master the exploration/exploitation balance in my own life!

IMPROVING EVOLUTIONARY ALGORITHMS BY MEANS OF AN ADAPTIVE PARAMETER CONTROL APPROACH

Guillaume CORRIVEAU

RÉSUMÉ

Les algorithmes évolutionnaires (AE) représentent une classe de méthodes d'optimisation couramment employée afin de résoudre divers problèmes scientifiques complexes. Cependant, les AE convergent souvent prématurément sur un optimum local, le processus évolutif est coûteux et les paramètres internes des AE sont difficilement réglables. Afin de pallier à ces problématiques et améliorer la performance des AE, il nous apparaît nécessaire d'aborder en premier lieu la question du réglage des paramètres. Pour ce faire, l'objectif de la thèse est de développer une méthode adaptative apte à contrôler l'ensemble des paramètres des AE, et ce, en utilisant comme indicateur de rétroaction le concept d'exploration et d'exploitation permettant d'interpréter le chemin de recherche poursuivi par l'optimiseur. Ainsi, le système adaptatif permettra de guider l'optimiseur vers sa direction de recherche optimale, ce qui offrira la possibilité de réduire le phénomène de convergence prématurée et qui devrait influencer positivement les coûts de calcul du processus évolutif.

Dans un premier temps, afin de caractériser l'exploration de l'espace de recherche offert par l'optimiseur, les mesures de diversité génotypique disponibles sont revues et étudiées. Pour ce faire, un banc d'essai spécialement dédié à cette fin est introduit et trois critères de diversité sont proposés pour évaluer la pertinence de ces mesures en tant qu'indicateur de diversité d'une population. Les résultats de cette étude montrent qu'aucune formulation publiée à ce jour n'est en mesure d'estimer adéquatement la diversité. Ce faisant, une nouvelle formulation génotypique est développée et l'analyse de ses performances révèle qu'elle offre généralement une meilleure description de la diversité malgré le fait qu'elle possède certaines lacunes importantes.

Pour caractériser l'exploitation des zones prometteuses pourvues par l'optimiseur, une étude similaire est conduite. Néanmoins, elle est maintenant orientée sur les mesures de convergence phénotypique disponibles puisque l'exploitation est principalement représentée par la réponse des individus. Encore une fois, l'analyse détaillée révèle qu'aucun descripteur phénotypique publié à ce jour n'est susceptible de représenter adéquatement la distribution de solutions d'une population donnée. Par conséquent, une nouvelle formulation phénotypique est développée dans le cadre de cette thèse et une analyse approfondie montre que cet indicateur exprime parfaitement le comportement escompté, et ce, en suivant une variété de populations.

Inspirés par ces réalisations, nous introduisons un outil de diagnostic basé sur la nouvelle formulation génotypique et phénotypique. L'utilité de cet outil est illustré en comparant l'impact que peut engendrer divers paramètres des AE sur le processus de recherche. Bien

que le but principal de ce développement soit d'exposer la pertinence d'employer à la fois une mesure génotypique et une mesure phénotypique pour caractériser le processus de recherche, l'outil de diagnostic s'avère être l'un des rares outils à la disposition des utilisateurs des AE permettant d'améliorer le processus de recherche sur des problèmes d'optimisation réels.

Sur la base des connaissances acquises, l'objectif principal de cette thèse est finalement considéré par la mise en place d'une nouvelle approche adaptative permettant de contrôler les paramètres des AE. La proposition est fondée sur l'utilisation d'un réseau Bayesien permettant d'interagir simultanément sur l'ensemble des paramètres des AE. À notre connaissance, il s'agit du premier système adaptatif offrant de telles capacités. En ce qui concerne le schéma de récompense employé par le système pour juger de la productivité des paramètres, les mesures génotypique et phénotypique développées précédemment sont mises à contribution, et ce, en favorisant les paramètres offrant une maximisation de l'exploration et de l'exploitation. La performance de ce système est analysée à l'aide d'un algorithme génétique à état constant (SSGA), et ce, sur un banc d'essai reconnu dans le domaine (CEC'05). De plus, cette nouvelle approche est comparée à sept autres méthodes, dont FAUC-RMAB et G-CMA-ES représentant actuellement l'état de l'art en terme de système adaptatif. Dans l'ensemble, les résultats montrent statistiquement que l'approche proposée est comparable à G-CMA-ES et qu'elle surpasse la plupart des autres méthodes adaptatives. Toutefois, cette étude expose le fait qu'aucune des méthodes considérées n'est capable de localiser l'optimum global sur des problèmes hautement multimodaux. Ceci nous conduit à affirmer qu'il y a sans doute un manque de synergie et de complémentarité entre les différents paramètres impliqués.

Par conséquent, nous recommandons la poursuite des recherches sur ces thèmes avec la perspective d'obtenir de meilleurs algorithmes d'optimisation. Pour ce faire, diverses recommandations sont proposées à la fin de cette thèse.

Mots-clés : contrôle adaptatif des paramètres, mesures de diversité, algorithmes évolutionnaires, équilibre entre exploration/exploitation, convergence prématurée

IMPROVING EVOLUTIONARY ALGORITHMS BY MEANS OF AN ADAPTIVE PARAMETER CONTROL APPROACH

Guillaume CORRIVEAU

ABSTRACT

Evolutionary algorithms (EA) constitute a class of optimization methods that is widely used to solve complex scientific problems. However, EA often converge prematurely over suboptimal solutions, the evolution process is computational expensive, and setting the required EA parameters is quite difficult. We believe that the best way to address these problems is to begin by improving the parameter setting strategy, which will in turn improve the search path of the optimizer, and, we hope, ultimately help prevent premature convergence and relieve the computational burden. The strategy that will achieve this outcome, and the one we adopt in this research, is to ensure that the parameter setting approach takes into account the search path and attempts to drive it in the most advantageous direction. Our objective is therefore to develop an adaptive parameter setting approach capable of controlling all the EA parameters at once. To interpret the search path, we propose to incorporate the concept of exploration and exploitation into the feedback indicator.

The first step is to review and study the available genotypic diversity measurements used to characterize the exploration of the optimizer over the search space. We do this by implementing a specifically designed benchmark, and propose three diversity requirements for evaluating the meaningfulness of those measures as population diversity estimators. Results show that none of the published formulations is, in fact, a qualified diversity descriptor. To remedy this, we introduce a new genotypic formulation here, the performance analysis of which shows that it produces better results overall, notwithstanding some serious defects.

We initiate a similar study aimed at describing the role of exploitation in the search process, which is to indicate promising regions. However, since exploitation is mainly driven by the individuals' fitness, we turn our attention toward phenotypic convergence measures. Again, the in-depth analysis reveals that none of the published phenotypic descriptors is capable of portraying the fitness distribution of a population. Consequently, a new phenotypic formulation is developed here, which shows perfect agreement with the expected population behavior.

On the strength of these achievements, we devise an optimizer diagnostic tool based on the new genotypic and phenotypic formulations, and illustrate its value by comparing the impacts of various EA parameters. Although the main purpose of this development is to explore the relevance of using both a genotypic and a phenotypic measure to characterize the search process, our diagnostic tool proves to be one of the few tools available to practitioners for interpreting and customizing the way in which optimizers work over real-world problems.

With the knowledge gained in our research, the objective of this thesis is finally met, with the proposal of a new adaptive parameter control approach. The system is based on a Bayesian network that enables all the EA parameters to be considered at once. To the authors' knowledge, this is the first parameter setting proposal devised to do so. The genotypic and phenotypic measures developed are combined in the form of a credit assignment scheme for rewarding parameters by, among other things, promoting maximization of both exploration and exploitation. The proposed adaptive system is evaluated over a recognized benchmark (CEC'05) through the use of a steady-state genetic algorithm (SSGA), and then compared with seven other approaches, like FAUC-RMAB and G-CMA-ES, which are state-of-the-art adaptive methods. Overall, the results demonstrate statistically that the new proposal not only performs as well as G-CMA-ES, but outperforms almost all the other adaptive systems. Nonetheless, this investigation revealed that none of the methods tested is able to locate global optimum over complex multimodal problems. This led us to conclude that synergy and complementarity among the parameters involved is probably missing.

Consequently, more research on these topics is advised, with a view to devising enhanced optimizers. We provide numerous recommendations for such research at the end of this thesis.

Keywords: Adaptive parameter control, diversity measures, evolutionary algorithms, exploration/exploitation balance, premature convergence.

TABLE OF CONTENTS

Page

INTRODUCTION 1
0.1 Background 1
0.2 Research problem.....4
0.3 Objective 8
0.4 Organization.....9

CHAPTER 1 REVIEW AND STUDY OF GENOTYPIC DIVERSITY MEASURES FOR REAL-CODED REPRESENTATIONS 13
1.1 Introduction.....13
1.2 Genotypic diversity measure.....16
1.2.1 General concept 16
1.2.2 Normalization 17
1.2.3 Genotypic diversity measures 18
1.2.4 Prior observable flaws on certain GDMs..... 22
1.3 Review of comparative studies 23
1.4 Benchmark 25
1.5 Results.....28
1.5.1 Unimodal landscape experiment..... 29
1.5.2 Multimodal landscape experiment..... 32
1.5.3 Stability analysis 36
1.5.4 Sensitivity analysis..... 37
1.5.5 Effect of outliers 40
1.6 GDM comparison over the CEC'05 benchmark.....43
1.7 Discussion 47
1.8 Conclusion 50

CHAPTER 2 EVALUATION OF GENOTYPIC DIVERSITY MEASUREMENTS EXPLOITED IN REAL-CODED REPRESENTATION.....51
2.1 Introduction.....51
2.2 Problem statement.....55
2.3 Characterizing population diversity.....59
2.4 Validation of the representative GDMs 62
2.4.1 Reduced population arrangement 62
2.4.2 Controlled cases of population diversity..... 64
2.4.3 Discussion 67
2.5 Conclusion 68

CHAPTER 3 REVIEW OF PHENOTYPIC DIVERSITY FORMULATIONS FOR DIAGNOSTIC TOOL 71
3.1 Introduction.....72
3.2 EEB concept.....74

3.3 Review of phenotypic formulations 76
3.3.1 General concept 77
3.3.2 Normalization 77
3.3.3 PCM formulation 79
3.4 Validation of phenotypic formulations 82
3.4.1 Requirements for a suitable diversity measure 82
3.4.2 Validation framework for the requirements analysis 84
3.4.3 Relevance of the phenotypic formulations 85
3.5 New phenotypic formulation proposed 87
3.5.1 Analysis of the new phenotypic formulation over the diversity requirements 88
3.6 Analysis of PCMs over specifically designed landscapes 89
3.6.1 Linear function 90
3.6.1.1 Landscape definition 90
3.6.1.2 Behavioral results of the PCMs 91
3.6.2 Double-slope landscape 93
3.6.2.1 Landscape definition 93
3.6.2.2 Behavioral results of the PCMs 95
3.6.3 Saw tooth landscape 96
3.6.3.1 Landscape definition 96
3.6.3.2 Behavioral results of the PCMs 98
3.7 Assessment of desirable PCM qualities 100
3.7.1 Definition of desirable PCM qualities 100
3.7.2 PCM_{13} reliability analysis 100
3.7.3 PCM_{13} sensitivity analysis 101
3.7.4 PCM_{13} analysis with outliers 102
3.8 Application of the EEB diagnostic tool 105
3.9 Conclusion 110

CHAPTER 4 EVALUATION OF THE GENERALIZED PHENOTYPIC FORMULATION AS A GENOTYPIC DIVERSITY MEASURE 113
4.1 Introduction 113
4.2 Generalization of PCM_{13} as multivariate diversity measure 113
4.3 Performance evaluation of the proposed GDM 116
4.3.1 Validation framework 116
4.3.2 Unimodal landscape experiment 118
4.3.3 Multimodal landscape experiments 119
4.3.4 Desirable quality criteria 122
4.3.4.1 Repeatability 122
4.3.4.2 Robustness 123
4.3.4.3 Influence of outliers 130
4.4 CEC'05 benchmark GDM comparison 132
4.5 Conclusion 137

CHAPTER 5 BAYESIAN NETWORK AS AN ADAPTIVE PARAMETER SETTING APPROACH FOR GENETIC ALGORITHMS 141
5.1 Introduction.....142
5.2 Review of adaptive parameter control strategies 144
5.2.1 Parameters involved..... 145
5.2.2 Feedback indicators 146
5.2.3 Credit assignment scheme..... 147
5.2.4 Parameter selection rule..... 147
5.2.4.1 Heuristic rule..... 147
5.2.4.2 Fuzzy Logic Controller (FLC)..... 148
5.2.4.3 Probability Matching (PM)..... 148
5.2.4.4 Adaptive Pursuit (AP)..... 149
5.2.4.5 Multi-Armed Bandit (MAB)..... 149
5.2.4.6 Covariance Matrix Adaptation (CMA)..... 151
5.2.5 Discussion..... 151
5.3 Parameter adaptation through Bayesian network 152
5.3.1 Graphical model of BNGA 153
5.3.2 Credit assignment schemes 155
5.3.3 Conditional Probability Table (CPT)..... 158
5.3.4 BNGA process 159
5.4 Comparative study 160
5.4.1 Methodology..... 160
5.4.2 Parameter states involved 161
5.4.3 Hyperparameter sensitivity analysis 163
5.5 Results.....167
5.5.1 Results of the SSGA parameter setting approaches..... 168
5.5.2 Results of the EA parameter setting approaches..... 171
5.6 Concluding discussion 174

CONCLUSION 177

RECOMMENDATIONS.....185

APPENDIX I OVERVIEW OF THE CEC'05 TEST FUNCTIONS CHARACTERISTICS 187

APPENDIX II SELECTION OF THE CREDIT ASSIGMENT SCHEMES INVOLVED IN THE ADAPTIVE CONTROL SYSTEMS 189

APPENDIX III BNGA PROCESS.....193

APPENDIX IV SURVEY OF THE PARAMETER STATES USED BY THE VARIOUS ADAPTIVE APPROACHES OVER THE CEC'05 TEST FUNCTIONS.....197

APPENDIX V SUMMARY OF THE EEBs ACHIEVED OVER SOME CEC'05 TEST FUNCTIONS.....205

APPENDIX VI RELEVANCE OF THE KNOWLEDGE INTEGRATION MECHANISM WITHIN BNGA 213

LIST OF BIBLIOGRAPHICAL REFERENCES.....219

LIST OF TABLES

Page

Table 1.1 GDMs used for the comparative study 19

Table 1.2 Stability analysis – unimodal landscape, with $n = 2$ 37

Table 1.3 Sensitivity analysis – landscape dimensionality {2, 10, 30}, with $N = 100$.....39

Table 1.4 Sensitivity analysis – population size {50, 100, 300, 500}, with $n = 2$ 40

Table 1.5 End diversity ratio, with respect to D^N_{PW}, in the presence of outliers – unimodal landscape, with $n = 2$ and $N = 100$.....43

Table 1.6 Qualitative ranking of the descriptors (0→Unreliable, 1→Weak, 2→Good, 3→Excellent) 49

Table 2.1 Desirable GDM quality criteria 53

Table 2.2 Diversity properties defined by Weitzman (1992).....60

Table 2.3 Defined requirements for GDM trueness validation.....61

Table 2.4 Behavior of the representative GDMs over the seven frozen cases.....66

Table 2.5 Summary of the fulfillment of the diversity requirements by the representative GDMs (A – Violation identified through the reduced population arrangement framework, B – Violation identified through the controlled cases of population diversity framework).....68

Table 3.1 PCM formulations used for the comparative study 80

Table 3.2 Requirements of the phenotypic formulation 84

Table 3.3 Behavior of phenotypic formulations over the six frozen case framework 85

Table 3.4 Summary of the diversity requirement fulfillment by the phenotypic formulations 86

Table 3.5 Diversity level of the new proposal over the validation framework.....89

Table 3.6 Convergence schedules for Group 1 (G1) to Group 5 (G5) over the saw tooth landscape 97

Table 3.7 Average range among 96% of the repetition data for PCM_{13} over the three landscape designs 101

Table 4.1 Behavior of D_{LN} over the seven frozen cases 117

Table 4.2 Stability analysis – unimodal landscape, with $n = 2$ 122

Table 4.3 Sensitivity analysis – population size {50, 100, 300, 500}, with $n = 2$ 124

Table 4.4 Sensitivity analysis – landscape dimensionality {2, 10, 30}, with $N = 100$ 125

Table 4.5 Mean error of the best solutions with respect to the global optimum over the 25 repetitions 134

Table 5.1 Parameters involved and their potential states 161

Table 5.2 Hyperparameter values considered for the sensitivity analysis 163

Table 5.3 Results format 167

Table 5.4 Results of the SSGA parameter setting approaches over the CEC'05 benchmark 170

Table 5.5 Statistical comparison by function group 171

Table 5.6 Results of the EA parameter setting approaches over the CEC'05 benchmark 173

Table 5.7 Statistical comparison by function group 174

LIST OF FIGURES

Page

Figure 0.1 Parameter setting taxonomy 3

Figure 0.2 Survey of the published adaptive control approaches applied to EA parameter setting 6

Figure 0.3 Organization of the thesis 9

Figure 1.1 Population of uniformly distributed individuals (N = 100) on four optima positions in a 2-D landscape at a given iteration. a) Iteration 1, b) Iteration 15, c) Iteration 30, d) Iteration 45 28

Figure 1.2 Mean GDM values of D^{N2}_{DTAP}, D^{N}_{TD}, D^{N}_{MI}, D^{N}_{PW}, and D^{N}_{VAC} for the unimodal benchmark 30

Figure 1.3 Mean GDM values of GF^{N}_{S} and $GF^{N}_{HC}(\alpha, M)$ for the unimodal benchmark α = {0.1, 0.5, 1.1, 2.0} 31

Figure 1.4 Mean GDM values of $GF^{N}_{R}(\alpha, M)$, and GF^{N}_{PW} for the unimodal benchmark α = {0.1, 0.5, 2.0} 31

Figure 1.5 Simulation with two different populations (black and empty circles) 32

Figure 1.6 Mean GDM values of D^{N2}_{DTAP}, D^{N}_{TD}, D^{N}_{MI}, D^{N}_{PW}, and D^{N}_{VAC} on a two optima benchmark 34

Figure 1.7 Mean GDM values of D^{N2}_{DTAP}, D^{N}_{TD}, D^{N}_{MI}, D^{N}_{PW}, and D^{N}_{VAC} on a four optima benchmark 34

Figure 1.8 Effect of the ratio of individuals associated with each optimum on a four optima benchmark 35

Figure 1.9 Mean GDM values of GF^{N}_{S} on the unimodal, two optima, and four optima benchmarks 35

Figure 1.10 Effect of outliers on D^{N2}_{DTAP} for the default configuration in a unimodal landscape (outliers are introduced from the 10th iteration) 41

Figure 1.11 Effect of outliers on D^{N}_{TD} for the default configuration in a unimodal landscape (outliers are introduced from the 10th iteration) 41

Figure 1.12 Effect of outliers on D^{N}_{PW} for the default configuration in a unimodal landscape (outliers are introduced from the 10th iteration) 42

Figure 1.13 Genotypic diversity level of various GDMs for the median repetition of the 10-D F2 function: a) solved with G-CMA-ES, b) solved with SSGA 45

Figure 1.14 Genotypic diversity level of various GDMs for the median repetition of the 10-D F10 function: a) solved with G-CMA-ES, b) solved with SSGA 46

Figure 2.1 Genotypic diversity levels of D^N_{PW} and GF^N_S over the single-site and multi-site convergence processes (two and four optima) 56

Figure 2.2 Representation of a uniformly random population with 100 individuals bounded between $[0, 1]^2$, where diversity is evaluated by: a) D_L – union of the area associated with each individual, b) D_{MST} – total length of the *MST* 59

Figure 2.3 Diversity on P_5 (solid curves) and P_4 (dash curves) with respect to the normalized location of $\boldsymbol{x}_5$ evaluated from: a) D_{PW}, b) GF_S, c) D_L, d) D_{MST} 63

Figure 2.4 Position of the optima for Cases 2 to 5 on: a) two-optima landscape, b) four-optima landscape 65

Figure 3.1 Orthogonal EEB framework with differentiation into four search zones 75

Figure 3.2 Phenotypic diversity level of the new proposal registered over a population of three individuals: a) Multiplicative formulation, b) Logarithmic formulation, c) Phenotypic convergence level of PCM_{13} 87

Figure 3.3 Fitness functions of the generic benchmark: a) Three translating cases of the linear landscape, b) Double-slope landscape and the saw tooth landscape 90

Figure 3.4 PCM behavior observed over the linear landscape: a) PCM_1, b) $PCM_{4.4}$ 91

Figure 3.5 PCM behavior observed over the linear landscape: a) $PCM_{5.1}$, b) PCM_{10} 93

Figure 3.6 PCM behavior observed over the linear landscape: a) PCM_{11}, b) PCM_{13} 93

Figure 3.7 PCM behavior observed over the double-slope landscape: a) $PCM_{4.3}$ and $PCM_{4.5}$, b) $PCM_{5.4}$ 95

Figure 3.8 PCM behavior observed over the double-slope landscape: a) PCM_8, b) PCM_{13} 96

Figure 3.9 PCM behavior observed over the saw tooth landscape: a) PCM_1 to PCM_8, b) PCM_9 to PCM_{13} 99

Figure 3.10 Sensitivity of PCM_{13} with respect to the population size (N) observed over: a) Linear and double-slope landscapes, b) Saw tooth landscape....102

Figure 3.11 Impact of outliers on PCM_{13} observed over the: a) Linear and double-slope landscape, b) Saw tooth landscape....103

Figure 3.12 Impact on PCM_{13} of outliers that are far away, observed over the linear landscape: a) 1% outliers, b) 10% outliers 104

Figure 3.13 Impact on PCM_{13} of outliers that are far away, observed over the double-slope landscape: a) 1% outliers, b) 10% outliers....104

Figure 3.14 Impact on PCM_{13} of outliers that are far away, observed over the saw tooth landscape: a) 1% outliers, b) 10% outliers 105

Figure 3.15 Impact of various selection plans over the EEB: a) 10D-F2, b) 10D-F10....109

Figure 3.16 Impact of various replacement plans combined with FUSS selection over the EEB: a) 10D-F2, b) 10D-F10....109

Figure 3.17 Impact of various population sizes over the EEB: a) 10D-F2, b) 10D-F10....110

Figure 3.18 Impact of various crossover types over the EEB: a) 10D-F2, b) 10D-F21....110

Figure 4.1 Diversity on P_5 (solid curves) and P_4 (dashed curves) with respect to the normalized location of $\boldsymbol{x}_5$ evaluated from D_{LN}....117

Figure 4.2 Genotypic diversity levels of D^N_{LN} and D^N_{PW} for the unimodal landscape....119

Figure 4.3 Genotypic diversity level of D^N_{LN} (black curves) and D^N_{PW} (gray curves) for a uniformly distributed population: a) Average diversity values over the complete process for three different multimodal structures, b) Average end diversity values for multimodal structures ranging from 2 to 25 optima 120

Figure 4.4 Genotypic diversity level of D^N_{LN} (black curves) and D^N_{PW} (gray curves) for a population distributed according to a

monopolizing scheme: a) Average diversity values over the complete process for three different multimodal structures, b) Average end diversity values for multimodal structures ranging from 2 to 25 optima 121

Figure 4.5 a) Effect of population size (N) on D^N_{LN} for the unimodal landscape, b) Maximum difference among the four samplings over the unimodal landscape 125

Figure 4.6 Effect of population size (N) on D^N_{PW} for the unimodal landscape 126

Figure 4.7 Effect of landscape dimensionality (n) on D^N_{LN} for the unimodal landscape: a) with $1/n$ in the D^N_{LN} formulation, b) without $1/n$ in the D^N_{LN} formulation 127

Figure 4.8 Effect of landscape dimensionality (n) on D^N_{LN} for the multimodal landscape (two optima): a) with $1/n$ in the D^N_{LN} formulation, b) without $1/n$ in the D^N_{LN} formulation 128

Figure 4.9 Maximum GDM value difference among the three samplings over the unimodal and multimodal (two optimum structure) landscapes: a) with $1/n$ in the D^N_{LN} formulation, b) without $1/n$ in the D^N_{LN} formulation 129

Figure 4.10 Effect of landscape dimensionality (n) on D^N_{PW} for: a) the unimodal landscape, b) the multimodal landscape (two optima).....130

Figure 4.11 Effect of outliers on D^N_{LN} for the unimodal landscape 131

Figure 4.12 Effect of outliers on D^N_{PW} for the unimodal landscape 132

Figure 4.13 Genotypic diversity level of D^N_{LN} and D^N_{PW} for the 10-D F2 function: a) Median run with 1 and 10 subpopulations, b) Mean of the end diversity for different numbers of subpopulations 135

Figure 4.14 Genotypic diversity level of D^N_{LN} and D^N_{PW} for the 10-D F8 function: a) Median run with 1 and 10 subpopulations, b) Mean of the end diversity for different numbers of subpopulations 136

Figure 4.15 Genotypic diversity level of D^N_{LN} and D^N_{PW} for the 10-D F10 function: a) Median run with 1 and 10 subpopulations, b) Mean of the end diversity for different numbers of subpopulations 137

Figure 5.1 General framework describing the adaptive process (steps in bold type refer to fundamental components) 145

Figure 5.2 Graphical model of BNGA 154

Figure 5.3 Results of the hyperparameter sensitivity study: a) PM, b) AP 165

Figure 5.4 Results of the hyperparameter sensitivity study: a) FAUC-RMAB, b) BNGA 166

LIST OF ALGORITHMS

Page

Algorithm 1.1 GDMs benchmark 27

Algorithm 1.2 Statistical procedure for GDMs sensitivity analysis 39

Algorithm 4.1 Clustering RCGA search procedure 133

LIST OF ABREVIATIONS

AP	Adaptive Pursuit
AUC	Area Under the Curve
BBOB	Black-Box Optimization Benchmarking
BGA	Breeder Genetic Algorithm
BLX	Blended crossover
BN	Bayesian Network
BNGA	Bayesian Network as Genetic Algorithms parameter adaptation approach
CBR	Case Base Reasoning
CEC	Congress on Evolutionary Computation
CMA	Covariance Matrix Adaptation
CPT	Conditional Probability Table
DE	Differential Evolution
DMAB	Dynamic Multi-Armed Bandit
DOE	Design of Experiment
EA	Evolutionary Algorithms
EDA	Estimation of Distribution Algorithms
EEB	Exploration/Exploitation Balance
ES	Evolution Strategies
FAUC	Fitness-based Area Under the Curve
FLC	Fuzzy Logic Controller
FSR	Fitness-based Sum of the Ranks
FUSS	Fitness Uniform Selection Scheme

GA	Genetic Algorithms
GDM	Genotypic Diversity Measure
GP	Genetic Programming
KMP	Klee Measure Problem
MA	Memetic Algorithms
MAB	Multi-Armed Bandit
MOOP	Multi-Objective Optimization Problem
NFL	No Free Lunch theorem
PCM	Phenotypic Convergence Measure
PCX	Parent-Centric crossover
PDM	Phenotypic Diversity Measure
PM	Probability Matching
PNX	Parent-centric Normal crossover
PSO	Particle Swarm Optimization
RCGA	Real-coded Genetic Algorithms
RMAB	Rank-based Multi-Armed Bandit
RNG	Random Number Generator
SLMAB	Sliding Multi-Armed Bandit
SOM	Self-Organizing Map
SR	Sum of the Ranks
SSGA	Steady-State Genetic Algorithms
UNDX	Unimodal Normal Distribution crossover

UX	Uniform crossover

LIST OF SYMBOLS

b	Non uniform mutation parameter
BEL( )	Posterior probability
BFE	Best Fitness Error
C	Scaling factor used by MAB approaches
C	Covariance matrix of the mutation distribution
$d(\boldsymbol{x}_i,\boldsymbol{x}_j)$	Distance function between individuals i and j
D( )	Diversity
D	Distance-based diversity measure
D	Decay factor used by RMAB
E	Set of pairwise connections between individuals
F	Fitness distribution of a population
f	Fitness
$\bar{f}$	Average fitness
$G(X,E)$	Set of undirected Graph represented by X and E
GF	Gene Frequency diversity measure
H_0	Statistical test null hypothesis
H/C ratio	Number of hyperparameter required to control EA parameters
J	Total number of parameter states
l	Length of the side of an hypercube S
LB	Lower Bound
LD	Length of the landscape Diagonal
M	Total number of intervals
MBFE	Mean Best Fitness Error
MST	Minimum Spanning Tree
n	Landscape dimensionality and Number of time a particular parameter state was selected
N	Population size

NMDF	Normalization with Maximum Diversity so Far
p	Probability of a given SSGA parameter state
p	Fraction of N that belongs to an interval
p_c	Crossover probability
p_m	Mutation probability
P	Population
$P(r\|p_i)$	Likelihood of p_i given r
$\hat{q}$	Empirical quality estimate
r	Replacement plan
r	Reward
$r_{offspring}$	Reward provided to SSGA parameters involved in offspring creation
$r_{population}$	Reward provided to SSGA parameters involved in population governance
RWD	Reward type
$S(\boldsymbol{x}_i,l)$	Hypercube bounding the diversity contribution of individual i
SP	Success Performance
SR	Success Ratio
STD	MBFE Standard Deviation
TC	Total number of hyperparameters Combinations
u	Minimum between M and N used for gene frequency normalization
U( )	Uniform distribution
UB	Upper Bound
V	Volume of the landscape
VMD	Virtual Maximum Diversity
W	Window size used by the reward schedule
x	Location of the individual on the landscape
$\hat{x}$	Normalized location on the landscape
$\overline{x}_k$	Average gene values of the population

α	Adaptation rate, Normalization factor for Bayesian network inference, and Level of significance of a statistical test
β	Learning rate
η	PNX scattering parameter
λ	Number of created offspring
$\lambda_r(\ )$	Message pass from a child node to a parent node
μ	Number of selected parents and Total length of the *MST*
μ_L	Lebesgue measure
$\pi_r(\ )$	Message pass from a parent node to a child node
σ	Mutation step size
σ_ε	UNDX standard deviation parameter of a normal distribution
σ_η	PCX and UNDX standard deviation parameter of a normal distribution
σ_ζ	PCX standard deviation parameter of a normal distribution
$\hat{\sigma}_f$	Standard deviation of the fitness distribution

SUPERSCRIPT

N	Normalized measure
t	Generation number
*	Argument for which the function reach the maximum value and Location of the global optimum

SUBSCRIPT

avg	Average fitness of the individuals
best	Best individual in the population

i	Individual number $\in \{1, 2, ..., N\}$ and Identification number for SSGA parameter
j	Individual number $\in \{1, 2, ..., N\}$ and SSGA parameter state $\in \{1, 2, ..., J\}$
k	Gene locus $\in \{1, 2, ..., n\}$
m	Interval number $\in \{1, 2, ..., M\}$
min	Minimum probability for each parameter state
max	Maximum probability that may reach a particular parameter state
offspring	Individual created by the generation plan
parent	Individual picked from the selection plan
s	Selected state for a given parameter
worst	Worst individual in the population

INTRODUCTION

0.1 Background

Evolutionary algorithms (EA) are commonly used to optimize complex scientific problems. These methods attempt to emulate the theory of evolution through the *survival of the fittest* concept. Overall, the search process proceeds as follows: A population of *N* individuals is randomly generated, and then μ *parents* are selected from that population using strategies that promote fittest individuals. Subsequently, λ individuals, called *offspring*, are created by modifying the genotypic material of the parents. A number of operators can be used for this operation. They belong either to crossover or mutation operator, and, most of the time, they are activated following a probability p_c and p_m respectively. Then, some individuals from the population are temporarily removed to make room for the offspring, and the empty places are filled by the set formed with the offspring and the removed individuals, following a defined update scheme. The evolution process is repeated until a predefined stopping criterion is met. Frequently, elite individuals are retained in subsequent generations to ensure that the best solutions found so far are not lost in the process.

EA offer many benefits, among them the ability to deal with discontinuous, multimodal, noisy, and high dimensionality landscapes, since no particular knowledge about the search space is required. They also face three major challenges:

1. They often converge prematurely over suboptimal solutions;
2. The evolution process is computationally expensive;
3. Their internal parameters are difficult to set.

These challenges have a direct impact on the performance achieved, as they weaken the robustness and efficiency of the EA search process. In other words, they can diminish the overall quality of the solutions found over a wide range of problems and extend the time required to locate them. As a matter of fact, achieving both robustness and efficiency is

rather difficult. This is demonstrated by the *no free lunch* (NFL) theorem, formulated by Wolpert and Macready (1997), which stipulates that, globally, no search algorithm dominates across all possible problems. This means that the selection of the most suitable optimizer is problem-dependent. In spite of the glaring conclusion of this result, the major research effort expended on these challenges is justified by the fact that practitioners are often interested in a specific class of problems, or they want to design search algorithms that have greater scope.

In terms of the first challenge, the stochastic nature of EA may help to alleviate the problem of premature convergence, since multiple search repetitions are prerequisite[1]. However, numerous modifications to the basic EA process have also been proposed over the years to address this problem. For instance, mechanisms for maintaining or promoting diversity inside a population constitute a broad class of mitigation responses (Sareni and Krähenbühl, 1998; Lozano, Herrera, and Cano, 2008; Das *et al.*, 2011). Another important research theme related to premature convergence concerns the improvement of the EA search components. This approach includes the definition of smarter selection methods (Hutter and Legg, 2006), advanced variation operators (Kita, Ono, and Kobayashi, 1998; Deb, Anand, and Joshi, 2002; Beyer and Deb, 2001), and reformulation of the evolution model itself (Satoh, Yamamura, and Kobayashi, 1996). Other indirect alternatives might be beneficial, like assembling various optimizers into a portfolio, with the expectation of minimizing poor performance through the sharing of the computational budget (Peng *et al.*, 2010).

With respect to the computational cost of EA processes, it is possible to attenuate its implications by leveraging the intrinsic population-based characteristic of EA, which allow parallel fitness evaluation of the individuals. Numerous models have been developed to achieve this, such as the master-slave framework and the multiple-deme approaches (Nowostawski and Poli, 1999; Alba and Troya, 2002; Konfrst, 2004; Munawar *et al*., 2008). Nevertheless, EA are often applied in simulation-based optimization problems, where evaluation of each individual may require a huge amount of computational time. Various

[1] From 30 to 100 repetitions are recommended (Wineberg and Christensen, 2007, p.3787), but, due to the computational burden associated with simulation, 25 repetitions are often acknowledged as sufficient.

strategies have been devised to overcome this problem, like the design of surrogate models for estimating the fitness function (Jin, 2005; Jin, Chen, and Simpson, 2001), or simply a prohibition mechanism for obviating the need to evaluate the same individuals multiple times (Corriveau, Guilbault, and Tahan, 2010).

Regarding the third challenge, more than ten parameters could be required from practitioners to carry out an EA process, from population size to the underlying operators of the evolution model. We believe that this is one of the motivations for the development of simplified metaheuristic variants, like differential evolution (DE) (Price, Storn, and Lampinen, 2005) and particle swarm optimization (PSO) (Kennedy, Eberhart, and Shi, 2001). Nonetheless, parameter setting has been a common theme of EA research since the inception of the field (Lobo, Lima, and Michalewicz, 2007). Consequently, many approaches have been developed in recent decades, and all of them can be considered to belong to one or other of the classes depicted in Figure 0.1 (Eiben, Hinterding, and Michalewicz, 1999).

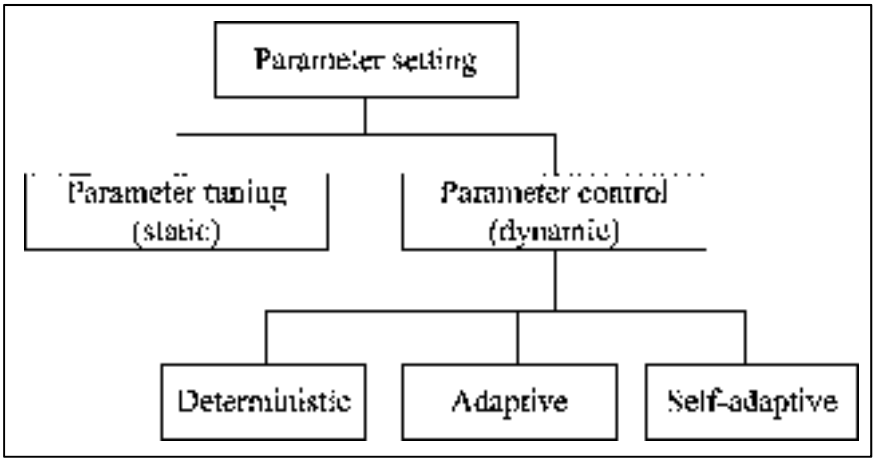


Figure 0.1 Parameter setting taxonomy

Parameter tuning involves fixing parameters before the optimization process takes place. Practitioners could approach the procedure by applying their experience with similar problems, or by using more rigorous approaches, like *design of experiments* (DOE). In contrast, parameter control involves modifying the parameters during the course of the

evolution process. Deterministic parameter control usually relies on iteration-based rules, as no feedback is gathered from the search process. Conversely, in adaptive parameter control methods, the parameters are modified based on feedback from the search process. Finally, in self-adaptive parameter control methods, the parameters are directly encoded in the chromosome of the individuals. Methods in this last category are often considered as "evolution of the evolution", since the EA process is responsible for modifying its own parameters.

Overall, this suggests that premature convergence, EA computational time, and parameter setting issues are intertwined. They are either a cause or an effect of the search path followed by the optimizer. In fact, parameter setting drives the direction of the search by regulating how resources are allocated. In contrast, premature convergence is a sign that the search path led to a dead end in terms of evolution, while excessive computational time may be evidence that the search path did not follow its optimal path. Consequently, we believe that addressing the parameter setting challenge might help prevent premature convergence, and it could also be the first rational step towards minimization of the search process computational time.

0.2 Research problem

In the previous section, we showed that resolving the parameter setting issue may help overcome the other EA challenges by guiding the search process towards its optimal path. However, this is a complex endeavor. In fact, EA parameters are problem-dependent, since particular parameter settings lead to specific algorithms to which the NFL theorem applies. Furthermore, parameter settings have been shown empirically to be evolution-dependent (Bäck, 1992; Stephens *et al.*, 1998; Eiben, Hinterding, and Michalewicz, 1999; Yoon and Moon, 2002), and to interact in a number of ways (Odetayo, 1997; Darwen, 2000; Rojas *et al.*, 2002; Nannen, Smit, and Eiben, 2008; Diaz-Gomez and Hougen, 2009).

The parameter setting categories identified in Figure 0.1 contain tuning approaches that can give us insight into the behavior of the parameters, and allow us, to some extent, to observe

the interactions among them. However, the results are valid only over a particular problem. Moreover, by itself, tuning is computationally intensive. In addition, it does not take into account the course of the evolution, because the parameters are set to be constant. In terms of deterministic control, generalization over different problems is troublesome, owing to the parameter setting schedule, which is user defined. Consequently, we believe that these approaches are not suitable responses to the challenges faced by EA.

In contrast, adaptive parameter control can be transferred from problem to problem. The course of its evolution is also taken into account through the information gathered from the search. Furthermore, nothing prevents adaptive control approaches from modeling the interaction among the parameters. However, in spite of these positive characteristics, there are major concerns with the adaptive system. For example, the relevance of the feedback indicator is crucial, and care must be taken to minimize its internal parameters, more commonly known as *hyperparameters*. As a matter of fact, it is useless to set EA parameters from a system requiring more hyperparameters, although exceptions would be possible if it could be demonstrated that the hyperparameters are far less sensitive than the EA parameters.

Self-adaptive control is attractive, as it is independent of the particular problem addressed. It also takes into account the evolutionary state, and the parameter interactions are implicitly handled. However, this comes at a price: increased dimensionality of the search space, and the process becoming a black box. Furthermore, parameters acting at the population level, like population size or selection method, require a special setting mechanism to aggregate the values proposed by the individuals (Eiben, Schut, de Wilde, 2006). Finally, in the words of Kramer (2010, p. 62), "*evolution rewards short term success.*" Therefore, parameter setting driven solely by the fitness of individuals, like self-adaptive control, can increase the occurrence of premature convergence.

Based on the above discussion, we believe that adaptive control is the most promising parameter setting scheme. However, as mentioned, an adaptation rule mechanism must be designed to enable all the parameters to be considered together and to involve as few

hyperparameters as possible. It must also have a proper feedback indicator. In order to investigate the applicability of the published adaptive approaches, we conducted a survey as part of this research, for which we collected and processed more than 100 papers on the topic. The outcomes of this survey are summarized in Figure 0.2.

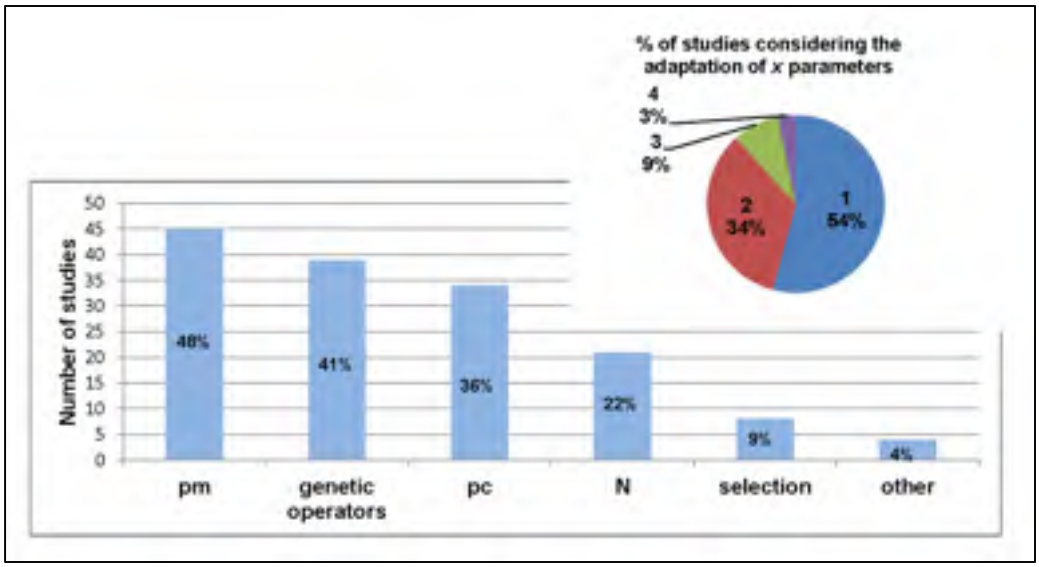


Figure 0.2 Survey of the published adaptive control approaches applied to EA parameter setting

The chart in Figure 0.2 illustrating the results of our survey shows that 88% of the adaptive parameter control proposals are restricted to two EA parameters or fewer, and no study was found in which more than four parameters were included. This suggests that no adaptive method can tackle all the EA parameters at once. In fact, most of the state-of-the-art adaptive systems are parameter independent (Goldberg, 1990; Thierens; 2005; Fialho, 2011). The lack of parameter interaction dependency makes difficult pursuing the optimal search path.

In terms of the feedback indicator, most of the studies proposing an adaptive mechanism endorse fitness improvement of the offspring. This is not surprising, given that the vast majority of adaptive control applications have been dedicated to setting the genetic operators and their related activation probabilities (see Figure 0.2). However, to expand the scope of

the adaptive parameter setting process, the search path followed must be accounted for in some fashion.

One way to represent this feature is to measure the emphasis placed on exploration of the search space and exploitation of promising regions by the optimizer. This is a key characteristic which provides a means to interpret the way an algorithm works (Eiben and Schippers, 1998) and is commonly known as the *exploration/exploitation balance* (EEB). Too much exploration leads to prohibitive computational cost, while too much exploitation leads to premature convergence. Handling the EEB can be a great way to manage the other two EA challenges discussed in section 0.1 with respect to the parameter setting framework. However, Beyer (1998, p. 334) concluded his investigation of the subject by stating that there is no model in existence for conducting a deep evaluation of the relationship between EEB and EA performance. This observation was generalized by Eiben and Schippers (1998, p. 13), who maintain that there is no accepted reading of the EEB, and obviously intensive research is needed to better understand the essence of evolutionary search processes. In fact, exploration and exploitation are either regarded as opposing forces or orthogonal forces (Gupta, Smith, and Shalley, 2006). In the opposing forces framework, diminishing one aspect results in a proportional increase in the other. Conversely, in the orthogonal forces framework, it is possible to maximize both exploration and exploitation, as they are treated independently. Beyond this representational question, a means to assess exploration and exploitation is required.

Population diversity monitoring is the technique predominantly used for this assessment, although there is no consensus in the community about the best way to apply it. Either a genotypic formulation or a phenotypic formulation, or both, can be used (Olorunda and Engelbrecht 2008; Tirronen and Neri, 2009; Herrera and Lozano, 1996). The former is based on the location of the individuals over the search space, and the latter on the fitness distribution of the individuals. On top of that, numerous genotypic and phenotypic expressions have been proposed over the years; however, to the best of the authors' knowledge, none was ever validated as a true population diversity metric.

Accordingly, we believe that developing an adaptive control system which supports all EA parameters while at the same time managing the search path through an appropriate feedback indicator remains an open challenge. Fialho emphasized the need for an advanced feedback indicator in the conclusion to his research on adaptive control parameter setting by declaring that (2011, p. 170) "[...] *in order to efficiently tackle multi-modal problems, the maintenance of some level of diversity in the population should also be accounted somehow for* [sic] *in the rewarding of operator applications* [...]."

0.3 Objective

That said, the objective of this thesis is to develop an adaptive parameter setting approach for controlling all the EA parameters at once. The work starts with the hypothesis that the search EEB, which serves as the feedback indicator, is most likely handled best by means of the orthogonal framework, where genotypic measurement is used to express the exploration axis, and phenotypic measurement is adopted to characterize the exploitation axis. To achieve the aim of this thesis, the following questions are investigated:

1. What is the best genotypic formulation for estimating the exploration provided by the search process?
2. What is the best phenotypic formulation for outlining the exploitation supplied by the search process?
3. How can the exploration and exploitation knowledge for adapting the EA parameters be converted into a reward?
4. Which adaptive system can best handle the parameter setting dependencies?
5. How can the performance of the adaptive parameter control proposal be assessed?

The core of this thesis is applied to real-coded genetic algorithms (RCGA) for continuous optimization problems. This does not limit the generality of the outcomes, however, as the concepts developed are directly applicable to any real-coded, population-based search process. It is to note also that the aim of this research is to bring flexible EA strategies, like

RCGA, to the performance level achieved by state-of-the-art EAs, like the covariance matrix adaptation evolution strategy (CMA-ES) (Hansen and Ostermeier, 1996).

0.4 Organization

To achieve the objective stated above, the thesis is divided into five chapters. The first four chapters relate to the assessment of the EEB concept through diversity measurements, while chapter 5 capitalizes on this knowledge with the development of an adaptive EA parameter control system. An overview of the topics covered and the relationships among them is illustrated in Figure 0.3.

In chapter 1, we review the genotypic diversity formulations that have been proposed over the years. From there, we develop a benchmark with various modalities for assessing the

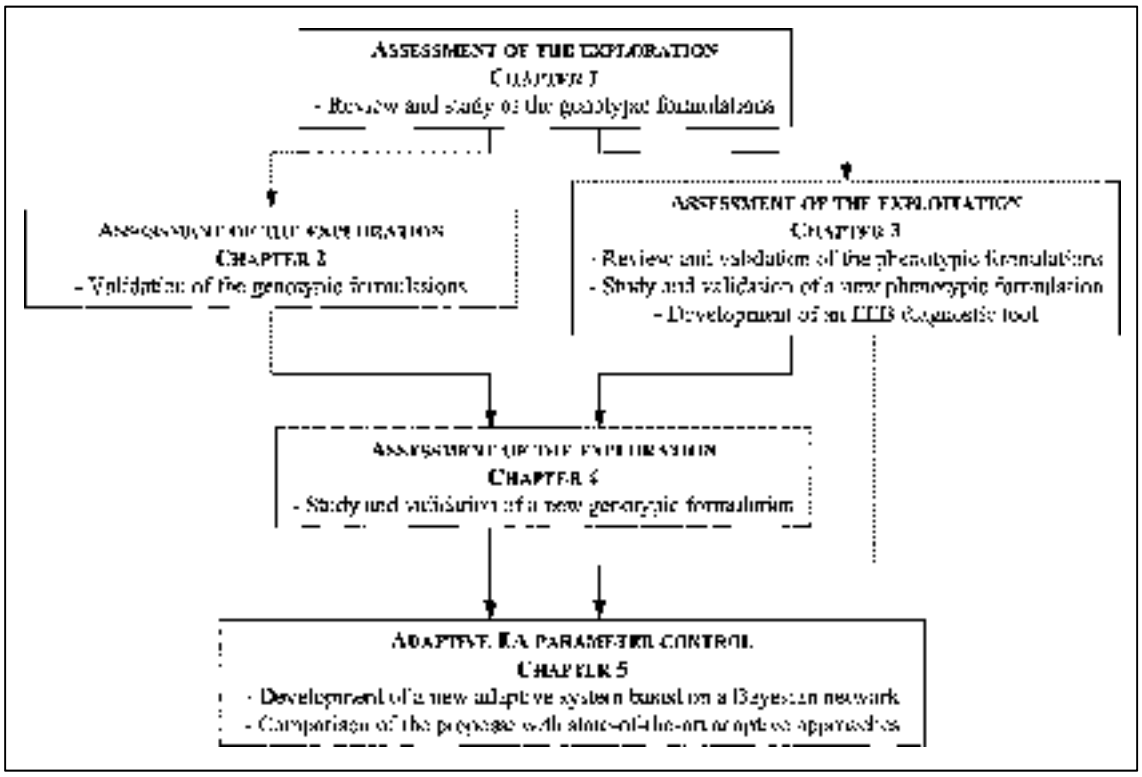


Figure 0.3 Organization of the thesis

behavior of these formulations. This study is performed using a controlled evolution process, in order to avoid any bias arising from a particular EA parameter setting system. The outcomes of the analysis enable us to identify similarities and differences in the genotypic formulations, as well as their limitations. This chapter provides the basis for answering question #1. The work described there was published in *IEEE Transactions on Evolutionary Computation* (Corriveau *et al.*, 2012a).

In chapter 2, a more in-depth investigation of the genotypic diversity formulations is carried out by defining three requirements that must be met by a genotypic measurement in order to accurately estimate population diversity. Validation of these requirements is handled by a reduced population arrangement that we propose and seven cases of controlled population diversity. Four genotypic measures are specifically selected for this analysis. Elements of the response to question #1 are provided in this chapter, and the work described in it has been submitted to *Applied Soft Computing* for a forthcoming publication (Corriveau *et al.*, 2012b).

In chapter 3, the study of chapters 1 and 2 is repeated, but this time from a phenotypic formulation standpoint. A review of the formulations considered is provided, along with their validation over the three diversity requirements reformulated within the phenotypic context. In addition, the results of this validation analysis are confirmed over a specifically designed benchmark, making it possible to observe the phenotypic formulation behavior over different fitness functions. A new phenotypic formulation is also proposed, and its performance is studied following the same approach. It is worth noting that the hypothesis supporting the usefulness of the phenotypic descriptor was justified in chapter 1 in a discussion about the most suitable EEB representation framework. Chapter 3 provides the answer to question #2. Furthermore, the outcomes of this investigation enable the introduction of an optimizer diagnostic tool dedicated to evaluating the impact of each parameter choice in terms of its EEB footprint. The value of such a monitoring tool lies in its ability to compare for instance, numerous common EA selection schemes, crossover operators, deletion schemes, and population sizes. This work has been published in the *Applied Soft Computing* journal (Corriveau *et al.*, 2013).

Chapter 4 is devoted to evaluating a new genotypic diversity formulation introduced within the context of the EEB diagnostic tool presented in chapter 3. The analysis follows the same benchmark and validation framework that was defined in the first two chapters of this thesis, and completes our response to question #1.

In chapter 5, we fulfill the main objective of this thesis, which is the development of an adaptive parameter setting approach for controlling all the EA parameters at once. In order to achieve this, questions #3 to #5 are addressed. The proposed new system, which supports EEB management as the feedback indicator, takes advantage of the findings presented in the previous chapters. A flexible way to translate this knowledge into parameter rewards is also suggested. The key feature of our adaptive system is that all the EA parameters are handled and parameter interactions are supported. This is made possible by the use of a Bayesian network as the foundation for the adaptive control approach. The performance of this proposal is evaluated over a recognized benchmark (CEC'05) and compared with various state-of-the-art adaptive techniques, like G-CMA-ES. This work has been submitted for publication to the *Applied Soft Computing* journal (Corriveau *et al.*, 2012c).

Finally, a summary of the results achieved is provided as conclusion. This is followed by recommendations for a future research effort towards the development of enhanced optimizers. Overall, the outcomes of this research are intended to provide practitioners with better optimization tools, although we do not claim, under any circumstances, that our approach is the best way to solve complex scientific problems.

CHAPTER 1

REVIEW AND STUDY OF GENOTYPIC DIVERSITY MEASURES FOR REAL-CODED REPRESENTATIONS

The exploration/exploitation balance (EEB) is a major concern in the control of evolutionary algorithms (EA) performance. Exploration is associated with the distribution of individuals on a landscape, and can be estimated by a genotypic diversity measure (GDM). In contrast, exploitation is related to individual responses, which can be described with a phenotypic diversity measure (PDM). Many diversity measures have been proposed in the literature without a comprehensive study of their differences. This chapter looks at surveys of GDMs published over the years for real-coded representations, and compares them based on a new benchmark, one that allows a better description of their behavior. The results demonstrate that none of the available GDMs is able to reflect the true diversity of all search processes. Nonetheless, the normalized pairwise diversity measurement (D^N_{PW}) proves to be the best genotypic diversity measurement for standard EAs, as it shows non-dominated behavior with respect to the desired GDM requirements.

This chapter was published in *IEEE Transactions on Evolutionary Computation* (Corriveau *et al*., 2012a).

1.1 Introduction

One of the major problems with evolutionary algorithms (EA) is premature convergence (De Jong, 1975; Mauldin, 1984; Goldberg, 1989; Eshelman and Schaffer, 1991). However, no method exists that offers adequate control of this phenomenon. The origin of premature convergence is the exploration/exploitation balance (EEB) (Eiben and Schippers, 1998). Too much exploration leads to random searching and a waste of computational resources, while too much exploitation leads to local searching and premature convergence. This balance could be controlled by setting the EA parameters (Eiben, Hinterding, and Michalewicz,

1999). Here, we consider parameter-setting in the broad sense of the term. For example, the population number, the type of evolution model, and restart strategies are all possible options for controlling the EEB. It is worth noting that the EEB dilemma is not unique to EAs, as it is essentially a resource allocation problem that any adaptive system must face (Gupta, Smith, and Shalley, 2006; Ishii, Yoshida, and Yoshimoto, 2002; Lee and Ryu, 2002).

The EEB can be viewed conceptually following one of two paradigms. In one of these, exploration and exploitation can be regarded as opposing forces, which means that increasing one reduces the other, while in the other, they can be regarded as orthogonal forces (Gupta, Smith, and Shalley, 2006). This second paradigm offers the possibility of increasing both exploration and exploitation simultaneously.

In fact, the opposing forces paradigm is a special case of the orthogonal forces paradigm, in that, for a unimodal landscape, reducing exploration increases exploitation proportionally. However, the situation is different for a multimodal landscape, where exploration and exploitation can be intensified simultaneously. For instance, to locate each peak of a landscape having uniformly distributed peaks of the same amplitude, and a population size equal to the number of peaks, exploration and exploitation have to be maximized concurrently. With very rugged landscapes, exploration could be in a maximal state, but with poor exploitation. In contrast, if the population converges over a very rugged, restricted region, exploration and exploitation would be in a minimal state at the same time. Finally, a converged population over a small plateau would be characterized by poor exploration and extensive exploitation. From this we can conclude that the orthogonal EEB concept is more suitable than the opposing forces concept to represent any landscape type. It also demonstrates that it could be useful to consider both genotypic and phenotypic diversity to characterize a given landscape effectively.

Exploration is adequately monitored by genotypic diversity evaluation (diversity of solutions), whereas exploitation is better described by phenotypic diversity (diversity of solution responses). These two diversity measurements also refer to the quantity and quality

of the population diversity respectively (Herrera and Lozano, 1996). In fact, genotypic diversity is built from the spread of the individuals over the search space, and phenotypic diversity is defined by the fitness distribution of the population. This means that phenotypic diversity is synonymous with fitness diversity. With normalized evaluation, unitary genotypic and phenotypic diversity values relate to maximum exploration and exploitation respectively. It is important to note that, unlike genotypic diversity, phenotypic diversity defines maximum exploitation when it is in a state of convergence. Properly evaluating genotypic and phenotypic diversity is difficult, however. Multiple diversity measures have been proposed in the literature without a clearly elucidated study of their differences being performed. This chapter focuses on a comprehensive study of genotypic diversity measures (GDM).

Depending on the problem and the representation used, the number of diversity measures could be infinite (Burke, Gustafson, and Kendall, 2004). It is important, therefore, to clearly define the scope of applicability. First, the diversity measures considered here are those that can be related to the search space location of the individuals in the population. The diversity measures related to the tree structure representation, used in the genetic programming (GP) field, are not covered (Burke, Gustafson, and Kendall, 2004; McPhee and Hopper, 1999; Monsieurs and Flerackers, 2003). Also, the analysis is restricted to real-coded representations.

Diversity assessment is critical to monitoring and/or controlling the EEB. The aim of this chapter is to provide an initial stepping stone toward EEB management, and it does so by studying the similarities and differences among GDMs. Of course, a good GDM should be capable of representing the real genotypic diversity of a population. However, it should also:

1. Demonstrate repeatability with respect to a similarly scattered population;
2. Be robust with respect to the simulation parameters, like population size and landscape dimensionality;
3. Adequately describe the presence of outliers inside the population.

To perform this comparison task, our complete analysis makes use of a new, simple benchmark that allows clear definition of the tested indicator behavior. Furthermore, we restricted this study to the available GDM formulations provided in the literature. The chapter is organized as follows: the next section, section 1.2, describes the various genotypic diversity measures studied; section 1.3 presents the published comparative studies; section 1.4 discusses the proposed benchmark; section 1.5 presents the results; section 1.6 validates the use of the proposed benchmark; and section 1.7 provides our concluding discussion.

1.2 Genotypic diversity measure

Even though no consensus has emerged on the definition of diversity (Lieberson, 1969; Patil and Taillie, 1982), the concept can be defined as the degree of heterogeneity or homogeneity between individuals in a studied population (Gouvêa Jr. and Araújo, 2008).

1.2.1 General concept

Genotypic diversity can be evaluated using one of two approaches. The first is based on a measurement of the distance between individuals. This distance may be evaluated from the mean spatial position of the population (Ursem, 2002; Abbass and Deb, 2003; Morrison and De Jong, 2002), from the position of the fittest individual (Herrera and Lozano, 1996), or the position of each of the individuals, which in this case would range from the pairwise measure (Olorunda and Engelbrecht, 2008; Barker and Martin, 2000) to the maximum distance between two individuals (Olorunda and Engelbrecht, 2008). The Euclidian distance is more common for distance estimation with real-coded genes, since the landscape is defined in a Euclidian space $\mathbf{R}^n$, where n represents the landscape dimensionality.

The second approach scans gene frequency. This concept is generalized from binary representations, where the probability of the alleles at each locus is calculated within the complete population (Wineberg and Oppacher, 2003). In a real-coded framework, all genes are continuous. Consequently, gene scanning requires gene partitioning. The predefined

intervals ($m \in \{1, 2, ..., M\}$) are then considered as possible alleles. Ichikawa and Ishii (1993) applied this procedure to integer representations, and the technique was later generalized to any symbolic alphabet by Wineberg and Oppacher (2003). Nevertheless, the number of intervals (M) involved in the discretization constitutes a severe limitation; they directly influence diversity estimation, which could make it difficult to achieve meaningful usage for a small population size or high dimensionality. Moreover, the gene frequency combination among all landscape variables must be defined. For example, Gouvêa Jr. and Araújo (2008) proposed using a representative gene to characterize the population diversity. In other words, the diversity measure is reduced to the consideration of only one gene or landscape variable characterizing the individuals. As they mentioned, the selected gene has to be a significant one. Therefore, to avoid a misleading diversity estimation, an average evaluation obtained from the diversity measure of each gene may be preferred (Wineberg and Oppacher, 2003). Collins and Jefferson (1991) also used the average gene frequency to determine the population diversity. However, this study was limited to binary representations.

1.2.2 Normalization

Normalization of the various GDMs is preferable for comparison purposes, as the descriptors can then be evaluated on the same basis.

When defined, the maximum value can be used as a normalization factor. In the case of distance measurement, the landscape diagonal (LD), that is, the maximum distance between opposite corners of the landscape, can also be used for normalization. Otherwise, the following simple normalization approach is proposed: the maximum value obtained so far during the evolution process of a given problem could serve as a normalization factor. The first iteration then becomes the reference, until a more diverse population is found. Since the initial EA population is generally created from a random uniform distribution, it is supposed to be the most diverse population. However, as information continues to arrive during the process, the indicator is updated if required. This normalization method is referred to here as

NMDF (Normalized with Maximum Diversity so Far). NMDF is similar to the normalization used by Herrera *et al.* (1994).

1.2.3 Genotypic diversity measures

The GDMs based on distance measurements (D) and gene frequency (GF) considered in this chapter are listed in the following table. They are presented in their normalized form. The asterisk following specific equation indicates that the corresponding measure uses a normalization method not defined in its original form.

The 1st GDM in this table corresponds to the diameter of the population (D^N_{DP}), which is a pairwise measure considering only the distance between the two most widely separated individuals in the population.

The 2nd GDM (Olorunda and Engelbrecht, 2008) represents the radius of the population (D^N_{RP}), and determines the distance between the individual farthest away and the mean position of the population. It is possible to generalize D_{RP} to account for only a certain fraction (f) of the individuals around the mean position. This leads to the 3rd GDM in Table 1.1, $D^N_{RP}(f)$, where the population size is sorted in ascending order with respect to the mean position. Therefore, extreme individuals can be set aside.

The 4th GDM, proposed by Ursem (2002), is the distance-to-average-point measure (D^N_{DTAP}), and it represents the mean radius of the population. In this study, a modified normalization version of this GDM is also considered, that is, D^{N2}_{DTAP}, which is presented as the 5th GDM in Table 1.1. With this form, the LD normalization factor is replaced by NMDF. This expression can also be considered as the normalization alternative to the D_{DTAP} measure proposed by Abbass and Deb (2003). No justification was provided in (Ursem, 2002; Abbass and Deb, 2003) to justify the usefulness of D_{DTAP}, except its intuitive formulation meaning.

Table 1.1 GDMs used for the comparative study

No.	GDM Formulation	No.	GDM Formulation
1.	$D_{CH} = \frac{1}{LD \cdot N} \max_{i,j=1,\dots,N} \left\{ \sqrt{\sum_{k=1}^{L} (x_{i,k} - x_{j,k})^2} \right\}$		$\bar{x}_k = \frac{1}{N} \sum_{i=1}^{N} x_{i,k}$
2.	$D_{DB} = \frac{1}{LD \cdot N} \max_{i,j=1,\dots,N} \left\{ \sqrt{\sum_{k=1}^{L} (\bar{x}_{i,k} - \bar{x}_{j,k})^2} \right\}$		$\bar{x} = \frac{1}{n \cdot N} \sum_{i=1}^{N} \sum_{k=1}^{L} x_{i,k}$
3.	$D_{DB}(f) = \frac{1}{LD \cdot N} \max_{i,j=1,\dots,N} \left\{ \sqrt{\sum_{k=1}^{L} (x_{i,k} - \bar{x}_{j,k})^2} \right\}$	11.	$D_{CD} = \frac{\bar{d}}{d_{min}} \cdot \frac{d_{max}}{\bar{d}}$,
4.	$D_{Gamma} = \frac{1}{LD \cdot N} \sum_{i=1}^{N} \sqrt{\sum_{k=1}^{L} (x_{i,k} - x_k)^2}$		where,
5.	$D_{S-D} = \frac{\frac{1}{N} \sum_{i=1}^{N} \sqrt{\sum_{k=1}^{L} (x_{i,k} - \bar{x}_k)^2}}{MIDP}$ "		$\bar{d} = \frac{1}{N} \sum_{i=1}^{N} \sqrt{\sum_{k=1}^{L} (x_{i,k} - x_{k,min})^2}$
6.	$D_{SD} = \frac{\frac{1}{N} \sum_{i=1}^{N} \frac{1}{N} \sum_{j=1}^{N} \sqrt{\sum_{k=1}^{L} (x_{i,k} - x_{j,k})^2}}{MIDP}$ "		$d_{max} = \max_{i,j=1,\dots,N} \left\{ \sqrt{\sum_{k=1}^{L} (x_{i,k} - x_{j,k})^2} \right\}$,
7.	$D_{S} = \frac{\frac{1}{n} \sqrt{\sum_{k=1}^{L} (\overline{x_k^2} - (\bar{x}_k)^2)}}{MIDP}$ "		$d_{min} = \min_{i,j=1,\dots,N, i \neq j} \left\{ \sqrt{\sum_{k=1}^{L} (x_{i,k} - x_{j,k})^2} \right\}$,
	where,	12.	$GE_2^{\beta} = \frac{1}{n \cdot \log(n)} \sum_{i=1}^{n} \sum_{j=1}^{N} p_{i,j} \log(p_{i,j})$ "
	$\overline{x^2} = \frac{1}{N} \sum_{i=1}^{N} x_i^2$	13.	$GE_{\alpha}^{\beta}(\alpha) = \frac{1}{n(1-\alpha)} \sum_{i=1}^{n} \left[1 - \sum_{j=1}^{N} p_{ij}^{\alpha} \right]$ "
8.	$D_{W} = \frac{\sum_{i=1}^{L} \sum_{j=1}^{N} (x_{i,j} - \bar{x}_i)^2}{MIDP}$ "	14.	$GE_3^{\beta}(\alpha) = \frac{1}{n \cdot \log(n)} \sum_{i=1}^{n} \frac{\log \left(\sum_{j=1}^{N} p_{ij}^{\alpha} \right)}{1-\alpha}$ "
9.	$D_{SW} = \frac{\frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \sqrt{\sum_{k=1}^{L} (x_{i,k} - x_{j,k})^2}}{MIDP}$ "	15.	$GE_{PW}^{\beta} = \beta \cdot \sum_{i=1}^{n} \sum_{j=1}^{N} p_{ij} (1 - p_{ij})$
10.	$D_{SW} = \frac{\frac{1}{N} \sum_{i=1}^{N} (\bar{x}_i - \bar{x})^2}{MIDP}$ "		where,
	where,		$\beta = \begin{cases} \frac{\Delta d}{n \cdot (\Delta d - 1) - \frac{n(N-1)}{N^2}}, & \text{if } \Delta d < N \\ \frac{N}{n(N-1)}, & \text{otherwise} \end{cases}$

The 6th GDM, proposed by Olorunda and Engelbrecht (2008), defined a measure considering the average of the average distance around the individuals of the population (D_{ALL}). In this formulation, the center is represented by individuals *i*. D_{ALL} was defined to give an indication of the dispersion of the individuals with respect to each other. In fact, with normalization, D^N_{ALL} becomes identical to D^N_{PW} (9th GDM in Table 1.1), but its formulation is more computationally intensive than the latter. Therefore, D^N_{ALL} is not considered further in this study.

In order to reduce the calculation time associated with pairwise measurements, which is $O(n \cdot N^2)$, to a linear relation $O(n \cdot N)$, where n and N represent respectively the landscape dimensionality and the population size, Wineberg and Oppacher (2003) propose a measure, named "true diversity" (D_{TD}), which represents the average standard deviation of each gene in the population. The "true diversity" normalized with NMDF is given by the expression D^N_{TD} , which corresponds to the 7th GDM in Table 1.1.

Following the computational improvement idea, Morrison and De Jong (2002) proposed the moment of inertia measure (D_{MI}), which leads to D^N_{MI} (the 8th GDM in Table 1.1) when normalized with NMDF. As with the physical concept, the remote points (outliers) should have greater influence on this measurement. The development of this GDM was justified by the goal of having a unique diversity measure, whatever binary or real-coded representation is used.

The mean of the pairwise distance among individuals in the population (D_{PW}) is an intuitive GDM (Barker and Martin 2000). This corresponds to the 9th entry in Table 1.1. Even though this measure may be more time-consuming, it could be quite effective for describing population diversity. Moreover, it is worth making the point that it is better to use a slower, but effective measure than an indicator that is fast, but prone to be inaccurate. For this study, the NMDF normalization factor is used for D^N_{PW}.

Herrera *et al.* (1994) proposed two GDMs as input to their fuzzy logic system: the variance average chromosomes (D_{VAC}) and the average variance alleles (D_{AVA}), both of which are defined for real-coded representations. The latter is not presented in Table 1.1, since it is equivalent to D_{MI} divided by ($n \cdot N$) and the term $n \cdot N$ remains constant in the evolution process considered. D^N_{VAC} is normalized by NMDF and it is the 10th GDM in Table 1.1. No justification was provided for characterizing the usefulness of these GDMs, except the fact that they are indifferent to the mutual exchange of individuals in a population and they take a low value when the population moves toward a genotypic convergence state.

The last GDM, based on the distance measure described in this study, is represented by the 11th entry in Table 1.1. It is D^N_{ED}, proposed by Herrera and Lozano (1996) without any justification. This diversity measure requires the pre-identification of the fittest individual (f_{best}) in the population, since it uses this individual as a reference to measure the distance from the other individuals. Other variants of this GDM are possible. Nevertheless, as will be explained in the next subsection, a major flaw can be seen in this kind of measurement.

In terms of gene frequency (GF) measures, the Shannon entropy (GF^N_S) (Shannon, 1948) is the best-known method employed as a GDM. It is intuitive, since entropy defines the level of disorder in a population (Rosca, 1995). The normalization of GF_S requires its maximum value. This is obtained when the gene frequencies are similar, which means that $p_{m,k} = 1/M$. However, it is important to note that this is true only if $M \leq N$. Otherwise, the maximum value is obtained when $p_{m,k} = 1/N$. In these cases, the most uniformly spread out distribution is $1/N$. Thus, replacing $p_{m,k}$ in the GF equation by one of these two upper bounds leads to that maximum value. This observation is valid for all GF measures, and the expressions 12 to 15 in Table 1.1 present the normalized version of the GF, where $u = \min\{M, N\}$. The Havrda and Charvát entropy (GF_{HC}) (Havrda and Charvát, 1967) is another important GF measure. This descriptor has been well analyzed by Nayak (1985). The following conditions are required for this family: $\alpha>0$ and $\alpha\neq1$. It is interesting to note that, when $\alpha=2$, GF_{HC} reduces to the Gini-Simpson index (Gini, 1921; Simpson, 1949). Good (1982) offers an excellent historical perspective on this index, and Rényi (1961) has proposed another entropy family

(GF_R). It is worth noting that, as $\alpha \rightarrow 1$, GF_{HC} and GF_R tend towards GF_S (Nayak, 1985). Finally, Wineberg and Oppacher (2003) published a GF that was developed for the same reasons as D_{TD}. This GF is designed to work with a finite-sized alphabet, which means that it can be used in the present context, where the total number of intervals on a gene (M) depicts the alphabet. This GDM is designated GF_{PW}. These authors have shown that GF_{PW} is correlated to GF_S (Wineberg and Oppacher, 2003). In fact, by means of a Taylor expansion of the second term of GF_S, ($\log(p_{m,k})$), they demonstrated that the last term of GF_{PW}, ($1-p_{m,k}$) constitutes the first term of this series, and dominates all the other terms. The normalization process for GF_{PW} is identical to that of the other GF measures. However, Wineberg and Oppacher added a correction term ($r = N \bmod M$) to account for the cases where M is not a common divisor of N, and is therefore applied when $M<N$.

To the authors' knowledge, all the published GDMs for real-coded representations available in the literature have been presented here.

1.2.4 Prior observable flaws on certain GDMs

Before moving further in the comparative study of GDMs, it is useful to eliminate those that present observable flaws in their formulations. This applies to D^N_{DP}, D^N_{RP}, $D^N_{RP}(f)$, D^N_{DTAP}, and D^N_{ED}.

Actually, D^N_{DP} is not an appropriate GDM for two reasons. First, the diversity estimate of the population is led by only the two most distant individuals, and this is the case whatever the scattering of the remaining individuals. Furthermore, the maximum value obtained by D^N_{DP} is when these two individuals are located on the extreme corners of the landscape, which is not, in any case, a sign that the population is fully diverse.

The formulation of D^N_{RP} shows similar flaws, as the diversity is based on the location of the individual farthest from the center of mass of the population. Therefore, a fully diverse population will be described by this indicator with a value near 0.5. The true diversity state of

the population is misleading, as the value goes toward 1. In fact, this indicates that the population converges near a landscape corner, whereas an outlier exists near the opposite corner of the landscape.

$D^N_{RP}(f)$ was introduced to reduce the potential impact of outliers on the preceding GDM. However, the factor f has to be properly defined, and, even though it increases robustness, it inevitably generates information leakage. Moreover, this indicator faces the same issue as D^N_{RP} with respect to coverage of the diversity range.

D^N_{DTAP} copes with the same issue as the three preceding GDMs, in terms of the diversity range coverage. This aspect is related to the LD used as the normalization factor. Furthermore, it is worth noting that the LD makes the diversity evaluation very sensitive to the landscape dimensionality, as the distance between the extreme corners of the landscape increases with the number of dimensions.

In contrast, D^N_{ED} is unable to describe the population diversity, since its normalization term decreases with its numerator, when the population moves toward convergence. Therefore, over a linearly convergent process, this indicator will remain constant, even if the population shows a linear reduction in its diversity.

In the next section, we present and discuss the comparative studies available in the literature.

1.3 Review of comparative studies

Gouvêa Jr. and Araújo (2008) presented five GDMs that can be used with real-coded representations: D^N_{DTAP}, GF_S, GF_{PW}, and $GF_{HC}(2.0)$ (Gini-Simpson index). The GDM not listed is a GF measure developed in (Mei-Yi, Zi-Xing, and Guo-Yun, 2004) for binary representations, and adaptable to real-coded representations. In fact, it uses D^N_{DTAP} for the intervals in a formulation similar to the Shannon entropy. Preliminary tests conducted in this study show that this descriptor is not adequate for the diversity evaluation of real-coded

representations, and so it is not considered here. Gouvêa Jr. and Araújo promoted the use of $GF_{HC}(2.0)$ with M=10, and consider only one representative gene. However, they did not provide any clear justification for doing so. They developed their EA adaptive control with this measure, and compared the resulting performance with Ursem's approach (Ursem, 2002) and a standard genetic algorithm (GA) on three dynamic environment problems. They concluded that their method outperformed the other two.

Olorunda and Engelbrecht (2008) compared six GDMs (D_{DP}, D_{RP}, D_{DTAP}, D^{N*}_{DTAP}, D_{ALL}, and swarm coherence) on four synthetic test functions treated with a particle swarm optimization (PSO) approach. D^{N*}_{DTAP} is a normalized version of D_{DTAP} which is different from D^{N}_{DTAP} and D^{N2}_{DTAP}. It considers the population diameter instead of the diagonal of the search space. Olorunda and Engelbrecht referred to this measure as the one used by Riget and Vesterstrom (2002). However, Riget and Vesterstrom clearly state that the normalization of their measure was achieved with the LD. In contrast, the swarm coherence measure requires the velocity of the swarm, which makes it PSO-specific. Olorunda and Engelbrecht also showed that it can produce ambiguous results. Consequently, swarm coherence was not included in section 1.2.3. Finally, the authors only include the D^{N*}_{DTAP} results in their study, which makes the analysis close to an intuitive comparison. Nevertheless, they rank the measures according to their sensitivity to outliers. From the most sensitive to the most robust, the classification is as follows: D^{N*}_{DTAP}, D_{DP}, D_{RP}, D_{DTAP}, D_{ALL}. They recommend D_{DTAP} based on this ranking and on the computation time.

As mentioned above, Wineberg and Oppacher (2003) showed that GF_{PW} is actually an approximation of GF_S, and, since D_{TD} corresponds to the average standard deviation of each gene, they all seem to be the same measure. These authors claim that, as a result, experiments were not required to choose the best GDM. However, in this chapter we will demonstrate that this belief appears to be a mistaken one, at least for real-coded representations.

1.4 Benchmark

The EA domain offers recognized benchmarks, such as CEC'05 (Suganthan *et al.*, 2005) and BBOB'09 (Hansen *et al.*, 2009), for single objective environment test cases. Nevertheless, for GDMs comparison purpose their usefulness can be problematic owing to two major reasons. First, since the use of a particular EA dictated the EEB over the optimization process, the diversity level of the population is biased by the underlying choice of EA parameters. Therefore, no information about the real diversity state of a population is available, except the one from the GDM comparison. This leads to an ill-defined problem, as we get different estimations from the GDMs without being able to say which one best reflects the true diversity value. The second aspect is related to the benchmark definition. Indeed, genotypic diversity is only concerned with the location of the individuals over a landscape, and not with its associated fitness function. Therefore, the sole requirement is to provide an environment for the GDMs where the population moves from a fully scattered state to a fully converged one. The number of optima over the landscape should also have an impact on the GDMs. A well-defined benchmark has to be able to simulate the modality influence.

In contrast, it could be interesting to link the GDM analysis to EA convergence tools as the takeover time concept, which is the time required by the best individual to populate the entire population (Goldberg and Deb, 1990). Within this framework, we will obtain a reference boundary between a fully scattered population (first generation) and a fully converged population (takeover time generation) for any landscape. However, as will be clearly seen in section 1.5, the most important zone where the behavior of the GDMs can be discriminated is between those boundaries where any convergence tools remain silent about EA behavior, and this is because of the stochastic nature of EAs.

We believe that an appropriate benchmark problem should present a population diversity that is known quantitatively, or at least qualitatively, throughout the evolution process. This section presents such benchmark problems for both uni- and multimodal landscapes with two

and four optima. These modality choices are made with the aim of visualizing the effect of GDM behavior on different landscape structures. The main steps of this benchmark are presented in Algorithm 1.1.

For quantitative comparison purposes, the simplest benchmark decreases the diversity linearly from a fully scattered population to a fully converged one. This is achieved by creating a uniformly distributed random population over the search space (line 29 of Algorithm 1.1) and reducing the available hyperspace towards a given location at a constant rate (line 9 of Algorithm 1.1). This simulates convergence toward an optimum position. It is important to mention that this involves no evolution operators, since a new population is generated within the converging population bounds of the genotypic space for each iteration. The reduction rate chosen per iteration is 2% of the distance between the landscape frontiers and the optimum position. The process then requires 51 iterations to converge, and ensures a clearly observable GDM behavior. To avoid the introduction of any bias, the optimum position is randomly generated on the landscape at each repetition (line 4 of Algorithm 1.1). For all experiments presented in this study, the genes ($x_{i,k}$) range from -50 to +50.

The multimodal landscape is similar to the unimodal one. However, since many optimum positions are fixed randomly at each repetition, the population is distributed uniformly or with a predefined ratio inside the respective bounds (line 28 of Algorithm 1.1). For example, Figure 1.1 shows four optima on a two-dimensional landscape. The population is uniformly attributed to each optimum position. In this example, the square boxes represent the space boundaries for each optimum at a given iteration.

Algorithm 1.1 GDMs benchmark

Algorithm 1: (*number_of_repetitions*, *modality_variation*, *pop_size*, *Number_of_outliers*)

Input: number of repetitions, modality of the landscape, maximum number of iterations for the search process, population size, number of simulated outliers, and scenario where outliers appear.

Output: optimum positions and populations generated for the complete process of all repetitions

```
for r = 1, ..., number_of_repetitions do
    // initialization of the initial populations //
    for peak = 1, ..., modality do
        // randomly select an initial position in the landscape limits //
    end for
    for t = 1, ..., iterations do
        // calculate the fitness of the individuals in the current population //
        for peak = 1, ..., modality do
            // evaluate the inner boundaries the upper bound for each dimension with respect to t and peak //
        end for
        // calculate the fitness of the individuals in the current population //
        if Scenario = 1 and t > t_outlier then
            N = pop_size - Number_of_outliers // evaluate the new population size //
            if t = t_outlier then
                // initialize outlier bounds //
                Set the outside lower bound and the outside upper bound of the outliers with the landscape limits
                for peak = 1, ..., modality do
                    Set the inside lower bound and the inside upper bound of the outliers by the lower and upper bounds of peak for the current t
                end for
            end if
        else
            N = pop_size
        end if
        // calculation of the modification vector //
        for ind = 1, ..., N do
            for peak = 1, ..., modality do
                // calculate the distance between the individual ind and the current peak //
                If dist(ind, peak) < (peak - 1) * 1/modality and dist(ind, peak) < peak * 1/modality then
                    Generate another random number that lies between the boundaries of peak for the current t
                end if
            end for
        end for
        // calculate the new fitness of the individuals //
        If Scenario = 2 and t ≥ t_outlier then
            for ind = 1, ..., Number_of_outliers do
                Generate random number that is located between the inside and outside outlier bounds of all peaks
            end for
        end if
    end for
end for
Return optimum positions for each repetition, and the population genotype of each iteration for each repetition
```

As mentioned by Olorunda and Engelbrecht (2008), different GDMs may have different sensitivity to outlier individuals, which means that the proposed benchmark must be adapted to reflect this aspect. For outlier influence simulation purposes, the initial benchmark remains

unchanged up to the 10th iteration. Then, a fraction of the population ($N_{outlier}$) is generated randomly (line 36 of Algorithm 1.1) within the hyperspace comprising the 1st to 10th iterations (lines 13 to 20 of Algorithm 1.1). Consequently, the outliers are free to move in a restrictive zone, while remaining at a significant distance from the optimum. Outliers bring exploration capabilities to the population. Nevertheless, their influence on GDMs has to be related to their number. In other words, the outliers should increase the diversity evaluation, but never dominate the measure.

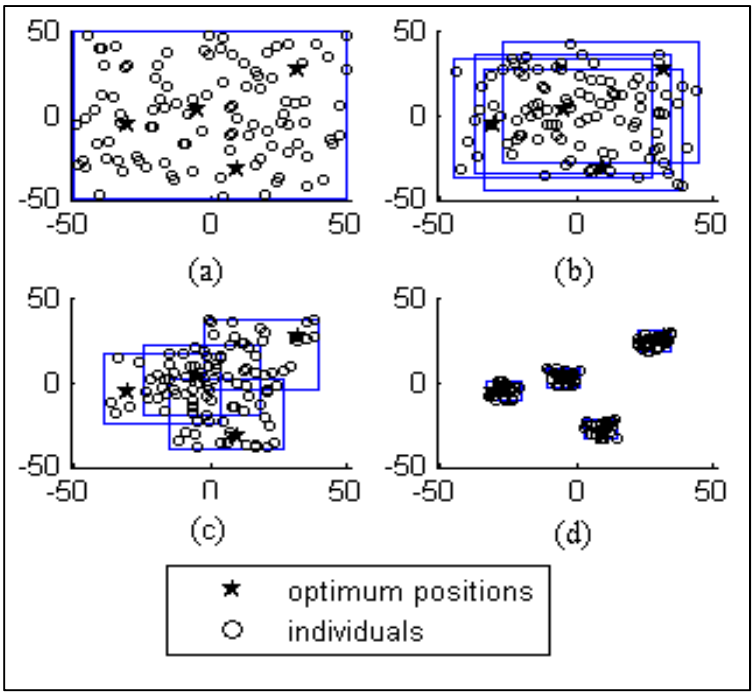


Figure 1.1 Population of uniformly distributed individuals (N = 100) on four optima positions in a 2-D landscape at a given iteration. a) Iteration 1, b) Iteration 15, c) Iteration 30, d) Iteration 45

1.5 Results

A default configuration allowing analysis of all GDMs on a similar basis is employed: the population size (N) is 100, and the number of intervals (M) for GF measures is fixed to 100

for each gene. The benchmark is defined on two-dimensional landscapes. Finally, the results are averaged over 50 repetitions.

The first subsection presents the behavior of all the GDMs on a unimodal landscape. Thereafter, the GDMs are studied on multimodal landscapes.

1.5.1 Unimodal landscape experiment

Figures 1.2-1.4 show the GDMs response on the unimodal landscape. Figure 1.2 indicates that D^{N2}_{DTAP}, D^{N}_{TD} and D^{N}_{PW}, with overlaid curves, precisely describe the linear relation intended by this benchmark. D^{N}_{MI}, even though showing a quadratic shape, still offers good discrimination of the diversity state. D^{N}_{VAC} acts similarly to D^{N}_{MI}. The behavior of these measures is expected to be quadratic, since they are based on genotypic variance. A linear trend could be achieved by taking their square root. However, this is not considered here, as this study is limited to GDMs that have already been suggested.

Figures 1.3-1.4 present the GF diversity measures. Given that all these measures share common properties, they are combined in the following discussion. First, the parameter α has a greater impact on GF^{N}_{HC} than on GF^{N}_{R}, making this latter GDM more reliable. In fact, α has an inverse influence on the two measure families. Also, the Gini-Simpson index (GF^{N}_{HC} (2.0)) appears to be similar to GF^{N}_{PW}. These measures were found to have a major drawback, however, which is that they remain very close to their maximum values for a long period during the process. In other words, they provide the worst discriminating diversity evaluations. Their formulations place the emphasis on crowded species or intervals (Lyons and Hutcheson, 1978). Therefore, diversity changes begin to be measured only when all the individuals pile up in a small number of intervals, which happens close to when the convergence state is reached.

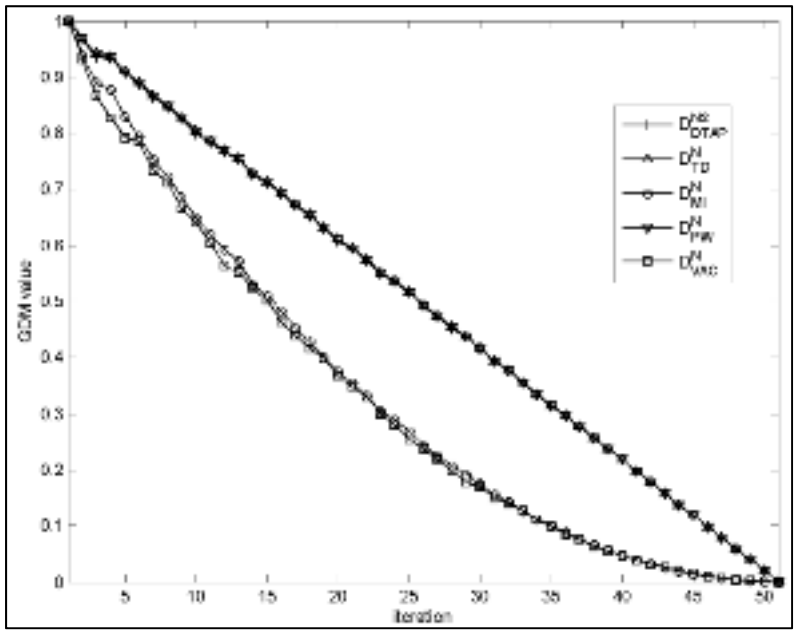


Figure 1.2 Mean GDM values of D^{N2}_{DTAP}, D^{N}_{TD}, D^{N}_{MI}, D^{N}_{PW}, and D^{N}_{VAC} for the unimodal benchmark

Complementary information about this GF drawback is presented in Figure 1.5. In this figure, the black and empty circles represent two different populations. Each contains 10 individuals, and a 10 by 10 grid is used for interval control. The black circle population is obviously more scattered than the empty circle one. However, the diversity evaluations for all the GF measurements indicate that these two populations are equally distributed. In contrast, distance-based measurements demonstrate the difference between them. For instance, D_{PW} indicates that population 2 (empty circles) is about 68% less diversified than population 1 (black circles). The non discrimination phenomenon observed for all GF measures can be explained by the fact that all GF measures are based on the proportion of individuals resident in the various intervals for each gene, and there is no consideration at all of the location of these intervals over the gene axis. This is a major weakness, which, as illustrated, could rapidly result in a misleading diversity analysis.

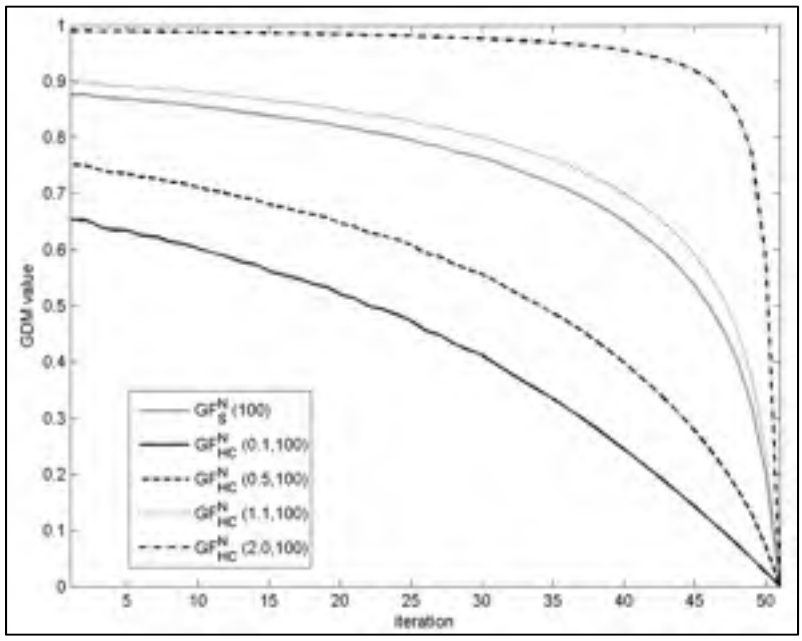


Figure 1.3 Mean GDM values of GF^N_S and $GF^N_{HC}(\alpha, M)$ for the unimodal benchmark $\alpha = \{0.1, 0.5, 1.1, 2.0\}$

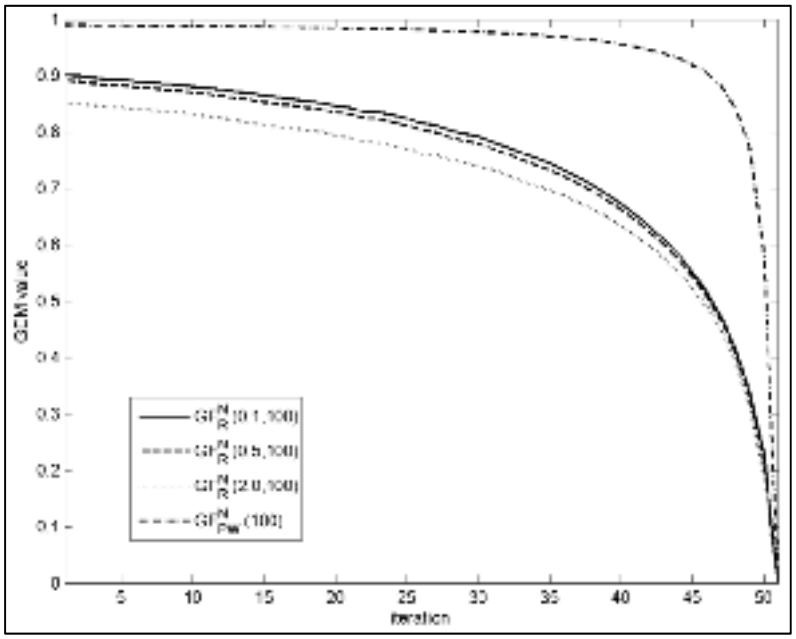


Figure 1.4 Mean GDM values of $GF^N_R(\alpha, M)$, and GF^N_{PW} for the unimodal benchmark $\alpha = \{0.1, 0.5, 2.0\}$

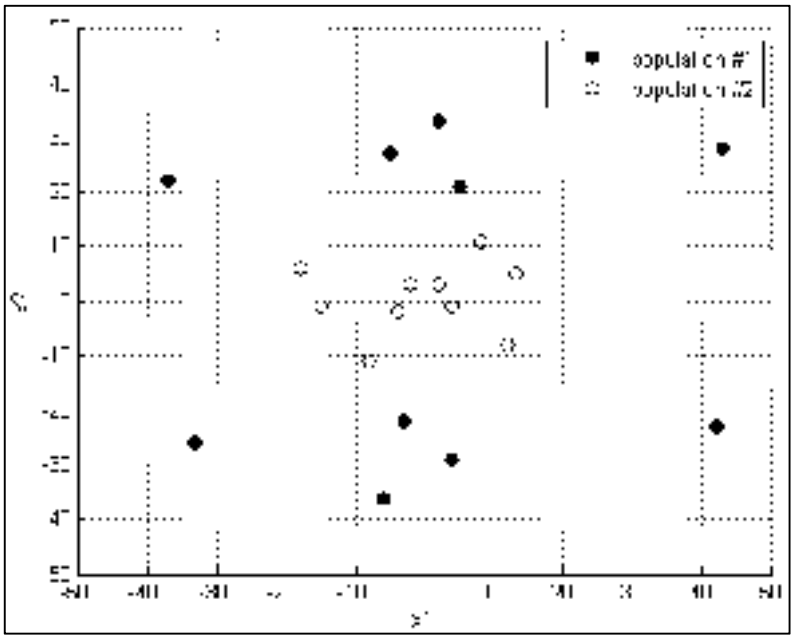


Figure 1.5 Simulation with two different populations (black and empty circles)

1.5.2 Multimodal landscape experiment

This subsection presents the response of selected GDMs to the multimodal benchmarks. Figure 1.6 shows the evaluation of the five GDMs normalized with NMDF on the multimodal benchmark with two optima. Figure 1.7 displays the same GDMs on a four-optimum landscape.

A general quadratic shape with a minimum somewhere in the process appears with these GDMs. The trend is accentuated as the modality increases. This phenomenon is explained as follows: at the beginning of the process, all the attracting pool boundaries share the entire landscape. As the process goes on, every bounded space shrinks around its respective optimum. As long as the boundaries overlap, diversity decreases, but then starts to increase with the separation of the bounded hyperspaces (Figure 1.1). Moreover, the rises in measured diversity depend on the ratio of the number of individuals converging to each optimum and the distance between the optima. Figure 1.8 illustrates the ratio effect with two different

GDMs on the four-optimum landscape. The comparison is performed for a uniform ratio (25% of N attached to each optimum) and a monopolizing optimum (70% of N to the dominant point, and the remaining 30% equally distributed among the other three optima). In light of Figure 1.8, the influence of the ratio becomes obvious: the non uniform case behaves as a unimodal landscape with the less attractive points acting as outlier clusters.

Figure 1.9 presents the characteristic GF pattern. In reality, the figure is restricted to the GF^N_S response for the uni- and multimodal landscapes with two or four optima. The curves clearly indicate that, as modality increases, the discriminating GF power deteriorates. This is because the convergence is less concentrated in a few intervals.

Since only two (four) optimum locations are represented by the population at the end of the multimodal process, gene frequency-based measurement seems to provide a better estimate of the real diversity than distance-based measurement. The difficulty of the latter is probably due to the non special treatment afforded to duplicated individuals. Nevertheless, this phenomenon is not studied further in this chapter, as no better discrimination capability could be found among the indicators compared. At the same time, it is difficult for GF measures to adequately describe the diversity of the population throughout the majority of the process, as no consideration is given to the location intervals. Consequently, this experiment has demonstrated that none of the GDMs is capable of reflecting the true diversity over a multi-site convergence process.

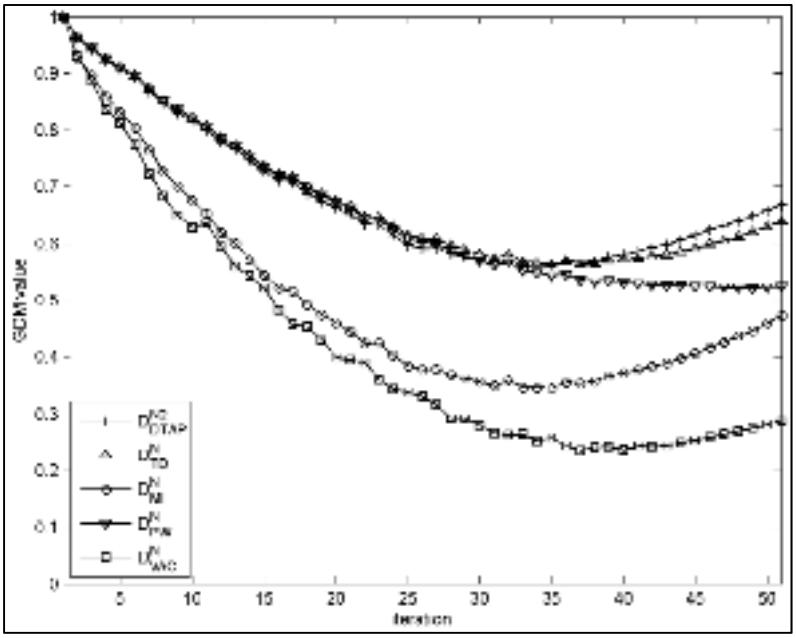


Figure 1.6 Mean GDM values of D^{N2}_{DTAP}, D^{N}_{TD}, D^{N}_{MI}, D^{N}_{PW}, and D^{N}_{VAC} on a two optima benchmark

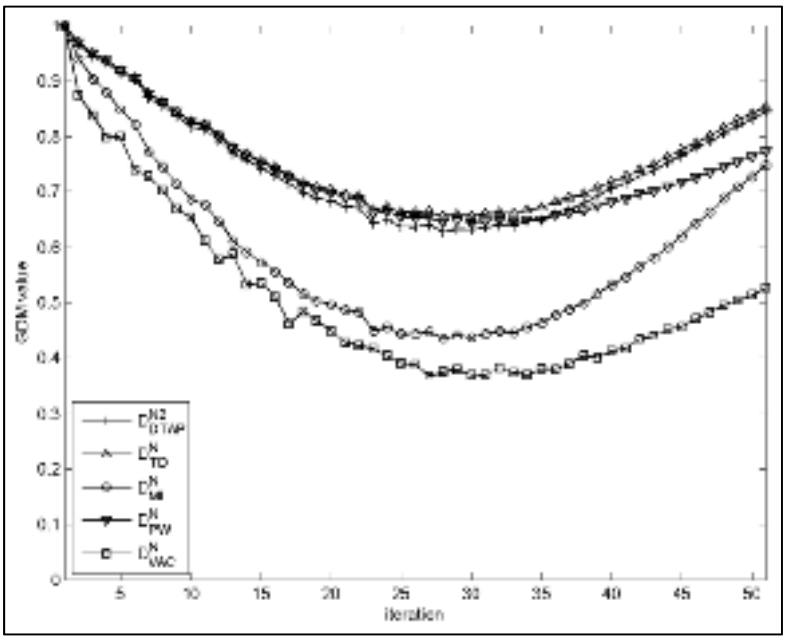


Figure 1.7 Mean GDM values of D^{N2}_{DTAP}, D^{N}_{TD}, D^{N}_{MI}, D^{N}_{PW}, and D^{N}_{VAC} on a four optima benchmark

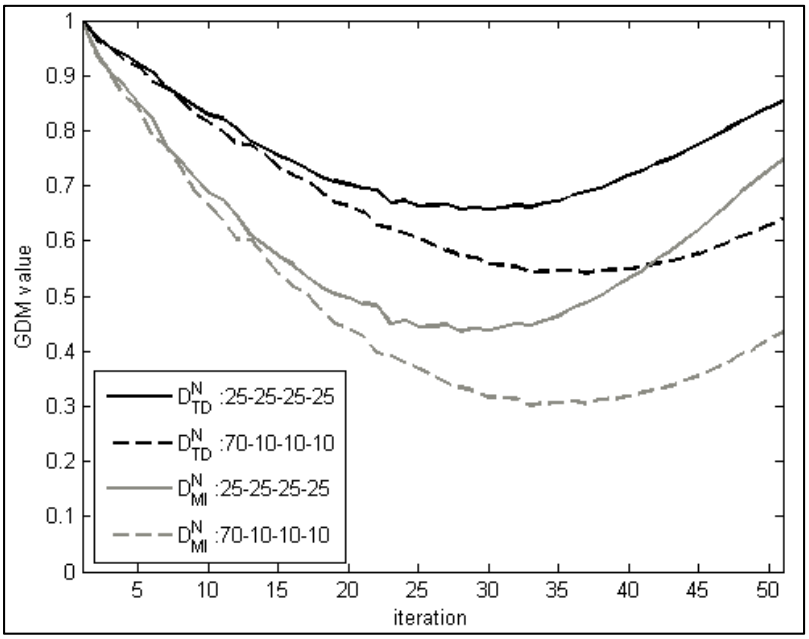


Figure 1.8 Effect of the ratio of individuals associated with each optimum on a four optima benchmark

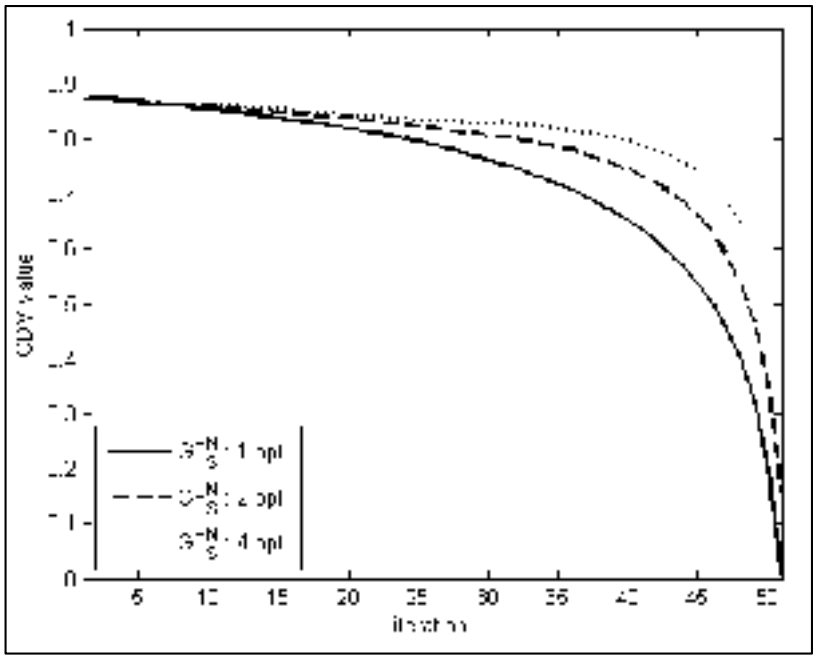


Figure 1.9 Mean GDM values of GF^N_S on the unimodal, two optima, and four optima benchmarks

1.5.3 Stability analysis

To further discriminate the power of the various GDMs, a stability analysis is produced in this section, followed by a sensitivity analysis and an outlier study.

GDMs should be stable in their measurement of the diversity value over similarly scattered populations. This property could be analyzed by looking at the dispersion of the 50 repetitions for a given iteration. Because the samples do not follow a normal distribution, the standard deviation is not a suitable indicator. Indeed, the normality assumption associated with the samples was tested and invalidated in this study using the Kolmogorov-Smirnov test (0.05 significance level). Stability is therefore evaluated by considering the dispersion range among 96% of the repetition data, which provides the same stability basis for all GDMs. That means that the difference between the second highest diversity value and the second lowest diversity value of the repetition at each iteration is computed. To present this analysis in a comprehensible manner, the dispersion values are averaged over the whole process.

Table 1.2 presents the stability computed for the five GDMs normalized with NMDF. Only the unimodal landscape is processed, since random positioning among optima on multimodal landscapes makes the stability analysis unreliable. However, the analysis is presented over four commonly used population sizes in EAs: $N \in \{50, 100, 300, 500\}$. This allows the sampling error to be visualized, since stability improves as the population size increases. By considering the largest population size, the sampling error is minimized. Thus, for this configuration (N = 500), four GDMs (D^{N2}_{DTAP}, D^{N}_{TD}, D^{N}_{MI}, and D^{N}_{PW}) have an average dispersion value under 0.05, which can be qualified as stable. The remaining GDM (D^{N}_{VAC}) could be considered less stable. The classification of the five GDMs, presented, in increasing order of stability, is: D^{N}_{VAC}, D^{N}_{MI}, D^{N2}_{DTAP}, D^{N}_{PW}, and D^{N}_{TD}.

The high degree of stability of most GDMs justifies our presentation of the above experiments, which shows that the mean curves of the GDM behavior are representative. It is also interesting to note that, even if most GDMs are stable and some have demonstrated a

similar trend in terms of their mean curves ($\{D^{N2}_{DTAP}, D^{N}_{TD}, D^{N}_{PW}\}$, $\{D^{N}_{MI}, D^{N}_{VAC}\}$, $\{GF^{N}_{HC}(2.0), GF^{N}_{PW}\}$), preliminary statistical tests based on the Wilcoxon signed rank indicated that none of them is built from the same diversity distribution over the 50 repetitions.

Table 1.2 Stability analysis – unimodal landscape, with $n = 2$

GDM	population size (N)			
	50	60	500	502
D^{N}_{DTAP}	0.110	0.086	0.019	0.040
D^{N}_{TD}	0.093	0.074	0.042	0.036
D^{N}_{MI}	0.119	0.096	0.055	0.045
D^{N}_{PW}	0.096	0.079	0.043	0.035
GF^{N}_{HC}	0.200	0.162	0.091	0.076

1.5.4 Sensitivity analysis

The robustness of the various GDMs with respect to the underlying parameters of the analyses (n and N) is also a concern. A one-at-a-time sensitivity analysis based on the Friedman statistical test allows a good definition of robustness. This is a non parametric statistical test with the implicit assumption that the samples are related. It could be viewed as a non parametric version of the repeated-measures ANOVA. The null hypothesis is that the sample distributions are the same, while the alternative is that their medians are different, at least for one sample (Sprent and Smeeton, 2000). The application of this test is justified for two reasons. First, as previously mentioned (see section 1.5.3), the sampling considered does not follow a normal distribution. Second, the same GDM is compared for different repeated simulations (sensitivity with respect to n or N), and they are thus related. More details on this statistical test in the EA context are provided in an excellent description by Garcia *et al.* (2009).

Before the results of the statistical test are presented, one question remains to be answered. It is related to the composition of the sampling used for comparison, since 50 repetitions were

conducted during a 51-iteration process. Should the sampling be formed with the mean of the 50 repetitions at each iteration (51 points in each sample and 1 p-value), or should a test be conducted for each iteration with the 50 repetition values (50 points in each sample and 51 p-values)? The second option appears to be the more relevant one, as comparing the mean of the repetitions at each iteration would cloud the analysis, and the null hypothesis would be rejected if the median of the mean values were statistically different for the samples compared. For example, if two simulations were to monotonically decrease over the convergence process, the statistical test would be based only on the difference in their mean values calculated exactly at the central iteration of the whole process. In contrast, the use of the 50 repetition values raises another question: how should we treat the 51 p-values (each related to a different iteration) to accept or reject the null hypothesis? In this study, we decided to rely on the percentage of p-values that fall below the predefined level of significance (α), which is fixed here at 0.05. Thus, the percentage value reflects the number of rejections of the null hypothesis over the 51-iteration process. A low percentage would indicate that most of the p-values were over the significance level, in which case the null hypothesis would not be rejected. A rejection then means that the GDM tested is sensitive to the scrutinized parameter. The default configuration described at the beginning of section 1.5 serves as a reference for the fixed parameters. No potential cross-influences between factors are included in this analysis. First, the impact of landscape dimensionality (n) is studied, followed by the effect of population size (N). Algorithm 1.2 presents the general procedure for the statistical comparison.

Table 1.3 presents the statistical test results for three landscape dimensions: 2, 10, and 30. The robustness of a GDM with respect to the dimensionality of the landscape is synonymous with scalability, which is important in the EA context. In other words, it means that the GDM offers similar diversity estimation, whatever the dimensionality of the landscape. As this analysis indicates, all NMDF-normalized GDMs show a relatively high degree of robustness, since fewer than one-third of the iterations reject the similarity among the samplings. Furthermore, in general, the sensitivity decreases as the modality of the landscape increases.

Based on this study, the classification, in terms of increasing order of robustness with respect to the dimensionality, is as follows: D^{N2}_{DTAP}, D^{N}_{PW}, D^{N}_{TD}, D^{N}_{VAC}, and D^{N}_{MI}.

Algorithm 1.2 Statistical procedure for GDMs sensitivity analysis

Algorithm 2: statistical_comparison(*iterations, samples, α*)

Inputs: maximum number of iterations for the entire process, number of samples for the statistical comparison and level of significance.

Output: percentage of rejection of the null hypothesis H_0 over the entire process.

1: *number_reject_H0* = 0

2: for *it* = 1, …, *iterations* **do**

3: **for** *s*=1, …, *samples* **do**

4: Define the sample set for the GDM repetition values for *it* and the benchmark parameter analyzed

5: **end for**

6: *p_value* = Evaluate H_0 with the statistical test

7: **if** *p_value* < *α* **then**

8: *number_reject_H0* = *number_reject_H0* + 1

9: **end if**

10: **end for**

11: **Return** (*number_reject_H0* / *iterations* *100)

Table 1.3 Sensitivity analysis – landscape dimensionality {2, 10, 30}, with $N = 100$

GDM	% p values < α		
	1 optimum	2 optima	optima
D^{N2}_{DTAP}	31.57	59.22	27.45
D^{N}_{PW}	25.53	11.76	0
D^{N}_{MI}	19.51	11.76	0
D^{N}_{TD}	21.57	15.69	1.96
D^{N}_{VAC}	11.76	21.57	0

Table 1.4 presents the sensitivity analysis results for the population size. The range was chosen to reflect common EA population sizes: $N \in \{50, 100, 300, 500\}$. No clear trend stands out from this analysis. However, we can see that D^{N}_{VAC} is very sensitive to population size, as is D^{N2}_{DTAP} for a low modality structure. Based on this study, the classification, in

terms of increasing order of robustness with respect to the population size, is as follows: D^N_{VAC}, D^{N2}_{DTAP}, D^N_{TD}, D^N_{MI}, and D^N_{PW}.

Table 1.4 Sensitivity analysis – population size {50, 100, 300, 500}, with $n = 2$

GDM	% p-values < α		
	optimum	2 optima	4 optima
D^{N2}_{DTAP}	50.98	37.25	11.76
D^N_{TD}	29.41	29.41	13.73
D^N_{MI}	27.12	29.41	13.73
D^N_{PW}	21.57	35.29	11.76
D^N_{VAC}	78.43	19.61	66.67

1.5.5 Effect of outliers

The following experiments illustrate the effect of outliers on the GDMs. Intuitively, the presence of outliers should increase diversity. Nevertheless, even though their number must be correctly reflected, outliers should never dominate the diversity evaluation, since, by definition, they correspond to a small portion of the population. The simulations were conducted with 1%, 2%, 5%, and 10% of outliers in the population. The following discussion uses the configuration presented at the beginning of section 1.5 as a reference. In addition, other experiments were performed with different population sizes: $N \in \{300, 500\}$. The results indicate that the effect of outliers on diversity evaluation is the same for all these population sizes. Also, it could be shown that outliers have a similar influence in both unimodal and multimodal cases. Consequently, to abbreviate the description, the multimodal landscape results are not incorporated. Moreover, for the sake of conciseness, even though the discussion includes the five GDMs based on NMDF, the following figures (Figures 1.10-1.12) present only the three GDMs that show perfect identification of the diversity level over the unimodal benchmark (D^{N2}_{DTAP}, D^N_{TD}, and D^N_{PW}).

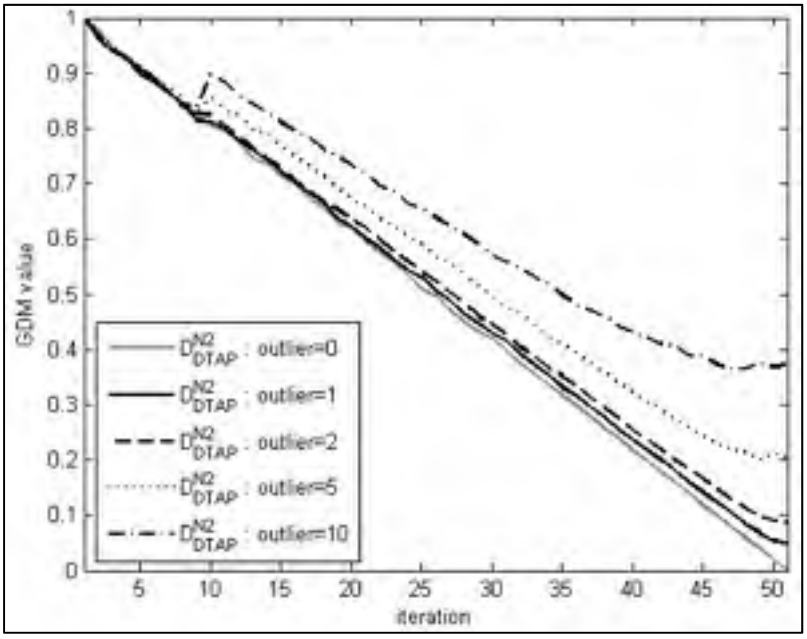


Figure 1.10 Effect of outliers on D^{N2}_{DTAP} for the default configuration in a unimodal landscape (outliers are introduced from the 10th iteration)

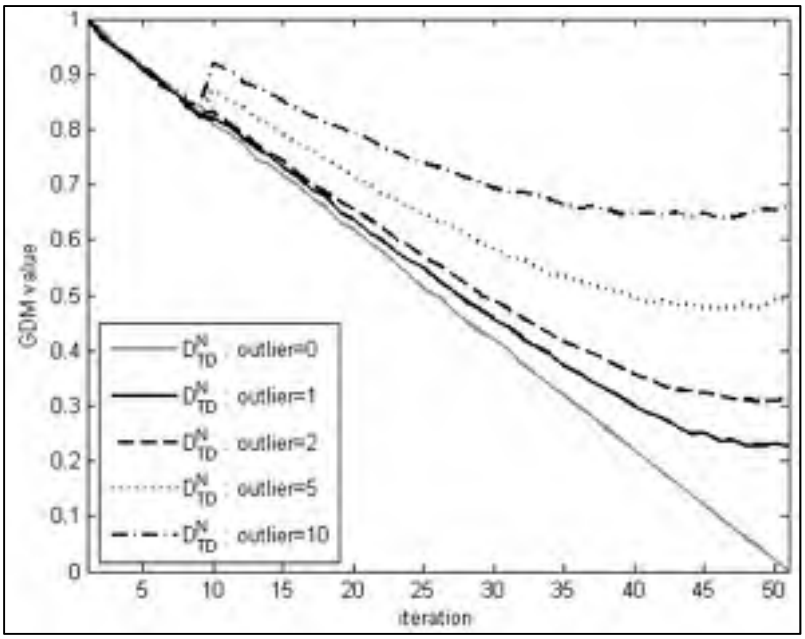


Figure 1.11 Effect of outliers on D^{N}_{TD} for the default configuration in a unimodal landscape (outliers are introduced from the 10th iteration)

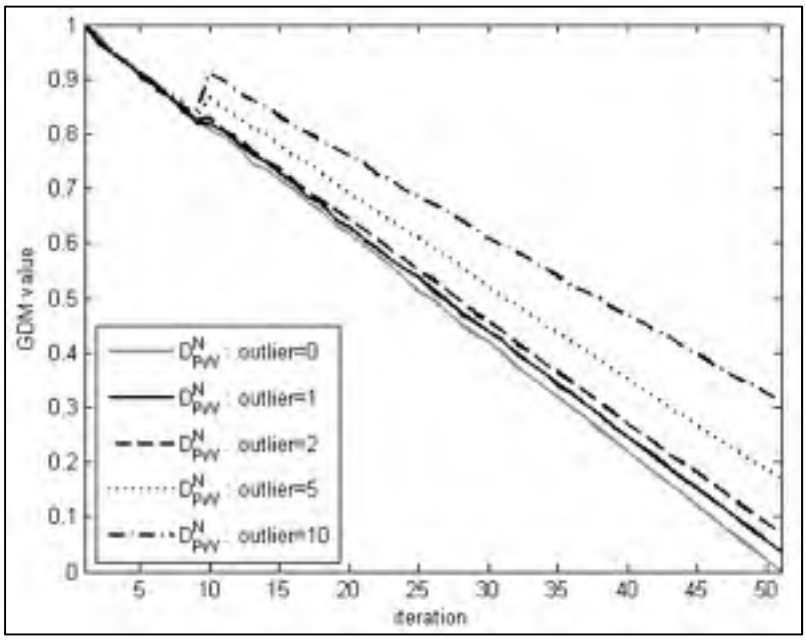


Figure 1.12 Effect of outliers on D^N_{PW} for the default configuration in a unimodal landscape (outliers are introduced from the 10th iteration)

These three GDMs show adequate patterns. They present a translating trend with respect to their no outlier mean curve. This translation is proportional to the percentage of outliers. However, D^{N2}_{DTAP} (Figure 1.10) and D^N_{TD} (Figure 1.11) reveal the distinct influence of the number of outliers at the end of the process. This phenomenon is explained as follows: in most repetitions, the outliers are far from the population mean. As the process evolves, the difference between each individual and the center of the population becomes dominated by the outliers and culminates at the last iteration.

Table 1.5 presents a comparison, with respect to D^N_{PW}, of the diversity value at the end of the process for each GDM based on NMDF. D^N_{PW} served as a reference because this GDM showed the most stable outlier evaluation (Figure 1.12). The comparison is summarized with a robustness classification (in increasing order of robustness): D^N_{TD}, D^N_{MI}, D^N_{VAC}, D^{N2}_{DTAP}, and D^N_{PW}.

Table 1.5 End diversity ratio, with respect to D^N_{PW}, in the presence of outliers – unimodal landscape, with n = 2 and N = 100

GDM	Outliers [illegible]				mean value
	[illegible]	[illegible]	[illegible]	[illegible]	
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

1.6 GDM comparison over the CEC'05 benchmark

To strengthen the usefulness of this study, all the GDMs presented were compared over the CEC'05 benchmark (Suganthan *et al.*, 2005) (see appendix I for an overview). To accomplish this task, a state-of-the-art EA was used, which is G-CMA-ES[2] (Auger and Hansen, 2005) and a particular EA specifically designed to promote diversity (Molina *et al.*, 2010).

The former was considered the best algorithm of the eleven EAs over the CEC'05 benchmark (Hansen, 2006; Garcia *et al.*, 2009). G-CMA-ES is an evolution strategy (ES) based on the covariance matrix adaptation (CMA) and a restart feature implemented to increase the exploration capability, as the population size is doubled at each restart. This feature is triggered by five independent convergence criteria related to CMA-ES parameters (Auger and Hansen, 2005). The parameters of G-CMA-ES used were the same as for CEC'05, except for the population size. Indeed, to make the observable behavior of the various GDMs clearer, and to have the same comparative basis as the study presented in the previous sections, an initial population size of 100 was used, instead of $4+\lfloor 3\cdot\ln(n)\rfloor$. For the gene frequency measures, $M = N$.

[2] CMA-ES version 3.51.beta was used to conduct this analysis. It can be accessed via http://www.lri.fr/~hansen/cmaes.m

The latter EA is based on a real-coded steady-state genetic algorithm (SSGA), where the selection plan and genetic operators are specifically chosen to promote diversity. In fact, a negative assortative mating strategy is used, as well as BLX-0.5 and a BGA mutation operator. This combination was selected in a memetic algorithm (MA) context, where the main assumption is that an EA is responsible for focusing on exploration, and exploitation is driven by local search algorithms (Molina *et al.*, 2010). Nevertheless, the true behavior of the explorative search method is often only implicitly addressed. Therefore, the following experiment attempts to explicitly characterize the explorative capability of the chosen strategy by means of GDMs. The parameters used within this SSGA framework are the same as those defined by (Molina *et al.*, 2010), except that the population size is fixed at 100 instead of 60, for the same reason as for G-CMA-ES.

A similar comparison was performed by Mattiussi, Waibel, and Floreano (2004) for binary GDMs over the two-dimensional sine envelope sine wave function, and they did this using a GA. They reported the average genotypic diversity over ten repetitions to demonstrate the similar behavior among different GDMs. However, due to the restart strategy of G-CMA-ES, and the fact that each repetition does not show the same convergence history, it is not helpful to compare the GDMs based on the average diversity obtained over the repetitions. Therefore, we have provided an analysis here for the median repetition of different CEC'05 benchmark functions.

To be concise, only the results of 10-dimensional F2 (the shifted Schwefel problem 1.2) and 10-dimensional F10 (the shifted rotated Rastrigin function) are presented, which are a unimodal and a multimodal landscape respectively. For the median repetition, G-CMA-ES found the optimum within a 1e-6 tolerance in 8 900 evaluations for F2, whereas the F10 optimum was achieved within a 1e-2 tolerance in 38 500 evaluations. In contrast, the SSGA implemented with diversity promoting features did not find the optimum within the CEC'05-prescribed tolerance, even after 100 000 evaluations.

Figure 1.13 exposes the genotypic diversity history of F2, and Figure 1.14 presents this history for F10. The restart strategy of G-CMA-ES is clearly observable over F10, where one restart was required, owing to the loss of all diversity without the global optimum being reached. To be comprehensive, only six GDMs are provided over these median runs; three of the most efficient measures (D^{N2}_{DTAP}, D^{N}_{MI}, and D^{N}_{PW}) and three of the worst descriptors (D^{N}_{DP}, D^{N}_{ED}, and GF^{N}_{PW}).

The discrimination problem of GF^{N}_{PW}, discussed in section 1.5.1, is clearly observable. In fact, this drawback, which characterizes all gene frequency measures, can dramatically distort the conclusion drawn on the search algorithm behavior, as demonstrated for the SSGA simulation over F10. The normalization problem raised in section 1.2.4 for LD-based measurements is noticeable with D^{N}_{DP}, and the inability of D^{N}_{ED} to describe genotypic

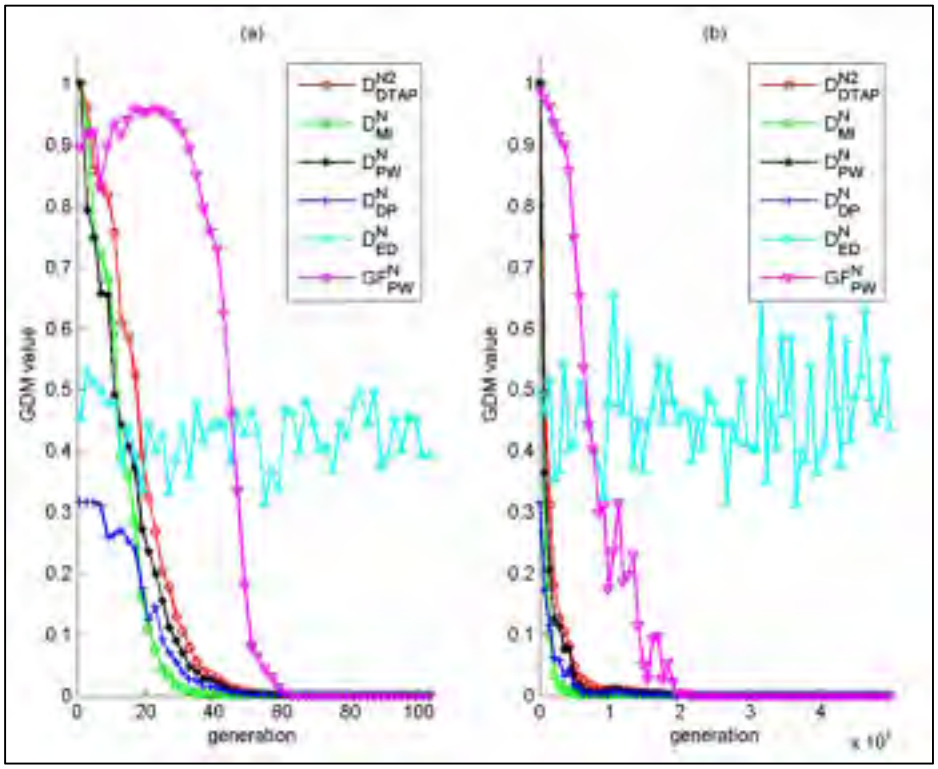


Figure 1.13 Genotypic diversity level of various GDMs for the median repetition of the 10-D F2 function: a) solved with G-CMA-ES, b) solved with SSGA

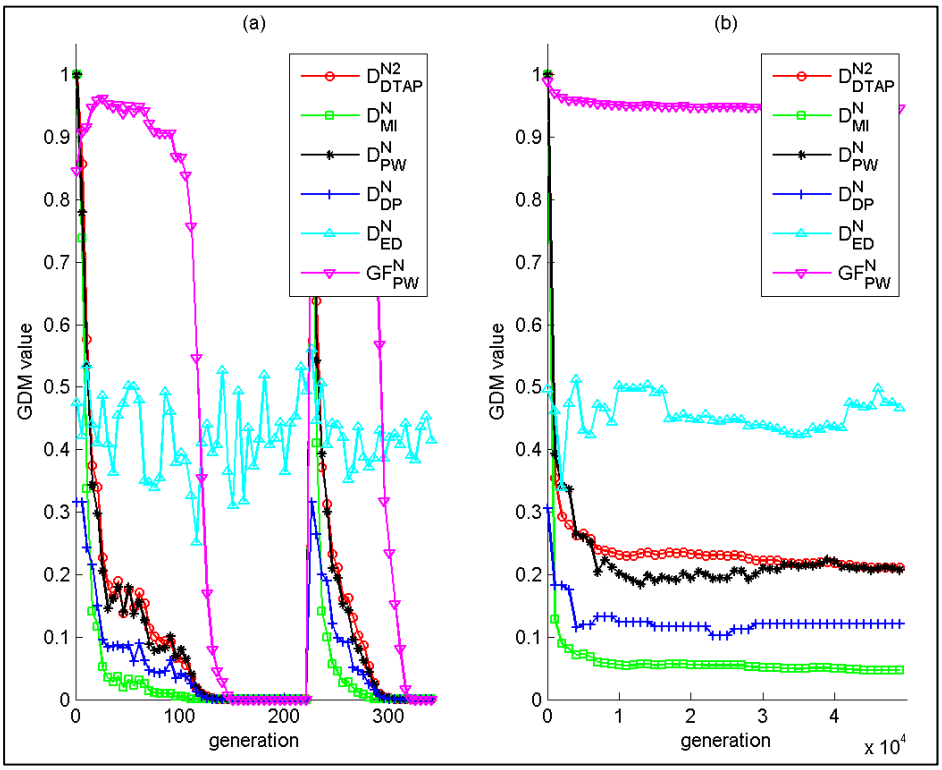


Figure 1.14 Genotypic diversity level of various GDMs for the median repetition of the 10-D F10 function: a) solved with G-CMA-ES, b) solved with SSGA

diversity is demonstrated by its relatively constant value over the process. As a result, neither of these GDMs is helpful for estimating the genotypic diversity of a population.

In contrast, D^{N2}_{DTAP}, D^{N}_{MI}, and D^{N}_{PW} show comparable genotypic diversity monitoring without conclusive difference. For all the functions analyzed, they present a comparable monitoring trend with different diversity levels. This becomes quite obvious from observing SSGA history over F10 (Figure 1.14 b)). For the same function solved by G-CMA-ES, the maximum difference between D^{N2}_{DTAP} and D^{N}_{PW} is achieved at the 3rd generation with a diversity gap of 0.34. Therefore, without any knowledge of the real diversity within the population, it is impossible to endorse the selection of any of these three GDMs. Furthermore, D^{N2}_{DTAP}, D^{N}_{MI}, and D^{N}_{PW} achieved a convergence state at the same evolutionary stage. For the F2 function, this behavior is expected, as it is characterized by only one

convergence site. However, the multi-site convergence phenomenon described in section 1.5.2 is hidden from the multimodal F10 function because of the EA search bias. Indeed, G-CMA-ES converges toward a single location, which, by the way, proves the usefulness of a restart strategy. In contrast, at the end of the SSGA process, 90% of the individuals remain unique, and they do so with a radius threshold of 0.1 unit, or 1% of the distance between the F10 landscape frontiers. This is worth noting, considering the relatively low diversity of these three GDMs (< 0.21) at the end of the process. In fact, we demonstrate that, even if no convergence status is monitored, all the individuals are neighbors. Furthermore, this happens quite rapidly during the process, as more than 75% of the generation stabilizes around this state. Therefore, it is possible that this particular SSGA strategy does not react as intended in the MA context. As a matter of fact, if the explorative strategy does not provide enough diversity, the occurrence of premature convergence could be exacerbated within an MA framework.

In summary, this analysis validated some of the GDM observations described in the previous sections. Nonetheless, the methodology has several limitations. The mere fact that each repetition has a different convergence history makes it impossible to use the mean GDM response that is necessary to reduce noise and produce sensitivity analyses that help to discriminate among GDMs. Also, the bias introduced by the EA does not allow multi-site convergence search pattern to be visualized, which is of interest for GDM comparison purposes. By themselves, these shortcomings validate the formulation of a specific GDM comparative benchmark, as proposed in section 1.4.

1.7 Discussion

This chapter has presented a detailed comparative study of more than 15 genotypic diversity measures common in the EA domain. We define these measures as exploration descriptors, since they are related to the spatial location of individuals in a given population. In this investigation, the evolution process had to be controlled to ensure a population diversity that is known throughout the progression. This fact was reinforced by the analysis presented over

the CEC'05 benchmark. We demonstrated that it is difficult to capture the fundamental properties of the various GDMs using an EA. This led to the development of a simple benchmark, which ensured the convergence of an initially fully scattered population in a chosen number of iterations. All the diversity measures were normalized to make it possible to compare them on the same basis. The results are summarized below.

Based on their formulation, five GDMs were eliminated prior to the comparative study: D^N_{DP}, D^N_{RP}, D^N_{RP} (f), D^N_{DTAP}, and D^N_{ED}. It was demonstrated that their underlying idea and/or their normalization method could be misleading in the genotypic diversity analysis. Therefore, these indicators are no longer recommended. Furthermore, D^N_{ALL} was not included in the comparative study, since its normalized version leads to D^N_{PW}.

Based on the GDM behavior requirements established in section 1.1, the five remaining distance-based GDMs (D^{N2}_{DTAP}, D^N_{TD}, D^N_{MI}, D^N_{PW}, and D^N_{VAC}) are capable of describing the intended diversity of the unimodal benchmark or single-site convergence problem, although some, because they are variance-based (D^N_{MI} and D^N_{VAC}), do so with more difficulty. In contrast, all the gene frequency measures (GF^N_S, GF^N_{HC}, GF^N_R, and GF^N_{PW}) have the same shortcoming with respect to this benchmark, which is an inability to discriminate the diversity level until a nearly converged population state has been reached.

For the multi-site convergence pattern, none of the available GDMs is capable of representing the diversity history. In fact, the multimodal experiments reveal that all distance-based GDMs (D^{N2}_{DTAP}, D^N_{TD}, D^N_{MI}, D^N_{PW}, and D^N_{VAC}) overestimate the end diversity, as no special treatment is afforded to duplicated individuals. Now, the GF measures have the same non discrimination issue throughout the scattered history of the population as in the case of the single-site convergence problem, even if they reach the intended convergence status level. It is worth noting that multi-site convergence does not usually occur in conventional EAs, as the population is frequently steered one way or another toward only one convergence location. Incidentally, that is one of the root causes of premature convergence. Therefore, we shall account for multi-site convergence with a GDM, in order to

validate and appreciate new developments based on diversity promotion methods (such as niching methods (Das *et al.*, 2011)), or any other strategy aimed at improving EA performance.

That said, the available distance-based GDMs are at least potentially usable within standard EA frameworks. For a better depiction of the performance of the GDMs, the stability, sensitivity with simulation parameters, and consideration of outliers were also analyzed. From this stage onward, GF measurements were set aside in our presentation, owing to their poor power to discriminate diversity. All distance-based GDMs demonstrate stability characteristics that are good to excellent, like their insensitivity with respect to landscape dimensionality. In contrast, none of these GDMs provides excellent insensitivity with respect to population size. In fact, D^{N2}_{DTAP} and D^{N}_{VAC} could be considered very sensitive to this parameter. Finally, D^{N}_{PW} is the best GDM for adequately taking into account the presence of outliers.

The behaviors of GDMs are ranked qualitatively in Table 1.6, based on the comparative study results. GF^{N}_{S} is inserted as the representative GF measurement, with the aim of providing a global picture of the potential GDMs. This table clearly shows the multi-objective aspect of choosing the most interesting of them. Therefore, based on the dominance concept widely used to solve multi-objective optimization problems, we could assert that

Table 1.6 Qualitative ranking of the descriptors (0→Unreliable, 1→Weak, 2→Good, 3→Excellent)

GDM	1 - Single-site convergence	2 - Multi-sites convergence	3 - Stability	4 - Insensitivity with respect to:			5 - Outliers
				dimension	pop. size	interval	
[illegible]	2	0	3	3	1	-	3
[illegible]	2	0	3	2	2	-	1
[illegible]		0	3	3	2		2
[illegible]	2	0	3	3	2	-	3
[illegible]		0	3	3	0	-	3
[illegible]	1	1	3	2	0	0	1

D^N_{PW} is the sole non-dominated genotypic diversity indicator, which would make it the best available GDM. Nevertheless, as previously discussed, this GDM is not suitable for describing multi-site convergence processes. As a result, a GDM formulation that is appropriate for dealing with any kind of search process remains an open question.

1.8 Conclusion

All things considered, this chapter has demonstrated that no measurement is capable of reflecting the diversity of a population for any search process. Nonetheless, the development of this kind of measure may support the establishment of, for instance, the foundation for a feedback mechanism used in adaptive methods. In fact, these mechanisms are probably the most interesting application for diversity measures, as the GDM could be used to assess, in part, the quality of the EEB driving the optimization process.

In the next chapter, we continue our study of genotypic diversity measurements by evaluating the meaningfulness of the formulations as population diversity estimates.

CHAPTER 2

EVALUATION OF GENOTYPIC DIVERSITY MEASUREMENTS EXPLOITED IN REAL-CODED REPRESENTATION

Numerous genotypic diversity measures (GDM) are available in the literature to assess the convergence status of an evolutionary algorithm (EA) or describe its search behavior. In the previous chapter, the authors drew attention to the need for a GDM validation framework. In response, this chapter proposes three requirements (monotonicity in individual varieties, twinning, and monotonicity in distance) that can clearly portray any GDM. These diversity requirements are analysed by means of controlled population arrangements. In this chapter four GDMs are evaluated with the proposed validation framework. The results confirm that properly evaluating population diversity is a rather difficult task, as none of the analysed GDMs complies with all the diversity requirements.

This chapter was submitted to *Applied Soft Computing* in October 2012 (Corriveau *et al.*, 2012b).

2.1 Introduction

One of the major problems with evolutionary algorithm (EA) is premature convergence towards a suboptimal solution (De Jong, 1975; Mauldin, 1984; Goldberg, 1989; Eshelman and Schaffer, 1991). This is due to a lack of diversity within the population. Single-site convergence schemes often lead to diversity losses, while strategies favoring multi-site convergence are considered to ensure better diversity. Among the most popular ways to solve this problem are the promotion of diversity approaches (Matsui, 1999; Hutter and Legg, 2006), the application of niching methods (Mahfoud, 1995; Das *et al.*, 2011), and the use of subpopulations (Ursem, 1999; Dezinger and Kidney, 2003). They are all designed to prevent population being trapped in one location. In contrast, other search methods, such as memetic algorithm (MA), are built on the assumption that EA provide significant diversity. This

implies that landscape exploration is driven by the EA, and the exploitation of promising regions is left to local search methods (Molina *et al.*, 2010). In reality, the explorative ability of MA is often only implicitly addressed. As a result, the performance of MA and other previously presented strategies is commonly evaluated indirectly by comparing their results (best fitness or average fitness) with those of other algorithms that do not implement the proposed features (Ursem, 2002). For instance, the performance of niching methods is frequently measured based on the number of peaks identified (Sareni and Krähenbühl, 1998). Of course, this technique is limited to problems having known optima locations. A more appropriate way to evaluate the performance of these strategies would be direct assessment. For this, the use of a diversity measure is preferable, since it allows for better characterization of the search behavior, and so provides a framework for algorithm comparison. Furthermore, tracking the diversity history throughout the process would make it possible to manage the exploration/exploitation balance (EEB) often sought by EA parameter control strategies (Lobo, Lima, and Michalewicz, 2007).

Two types of measurement are convenient for diversity monitoring: the genotypic diversity measure (GDM), which characterizes the distribution of a population over a landscape, and the phenotypic diversity measure (PDM), which describes the fitness distribution (Herrera and Lozano, 1996). GDM is more reliable than PDM for tracing premature convergence issues and for comparing the performance of multi-site convergence search processes, since the latter is influenced by the landscape relief. However, it is more difficult to assess diversity with GDM than it is with PDM, given that GDM is built on a multivariate distribution instead of a univariate distribution, as is the case for PDM (Tirronen and Neri, 2009).

In spite of the inherent complexity of GDM, numerous formulations have been proposed in the literature for the real-coded representation context. They can be classified into the following two families: distance-based measures, and gene frequency measures. The distance-based measurements consider the distance between individuals, which can be evaluated from the mean spatial position of the population (Ursem, 2002; Abbass and Deb,

2003; Morrison and De Jong, 2002) or from the position of the fittest individual (Herrera and Lozano, 1996). The position of each individual could also be used. This evaluation ranges from the pairwise measure (Olorunda and Engelbrecht, 2008; Barker and Martin, 2000) to the maximum distance between two individuals (Olorunda and Engelbrecht, 2008). The second family scans the gene frequency. This concept is generalized from binary representations, where the probability of the alleles at each locus is calculated within the entire population (Wineberg and Oppacher, 2003). In contrast, for a real-coded framework, all genes are continuous. Consequently, the gene scanning operation requires gene partitioning, where predefined intervals are considered as possible alleles. The number of intervals (*M*) involved in the discretization constitutes a severe limitation, as this number directly influences diversity estimation, especially for small populations or high dimensionality problems. Moreover, the gene frequency combination among all the landscape variables must be defined. For instance, in (Gouvêa Jr. and Araújo, 2008), a representative gene was preferred over averaging the diversity contribution of each gene (Wineberg and Oppacher, 2003).

Having many definitions of the same measure raises the question, what are the qualities of a good GDM? Table 2.1 lists three recognized quality criteria (Olorunda and Engelbrecht, 2008) that are desirable for a diversity indicator (section 1.5.3-1.5.5). Assessment frameworks are also proposed in the table. It is difficult to rank these criteria in terms of desirability, and so we consider them all to be equally important.

Table 2.1 Desirable GDM quality criteria

QUALITY CRITERION	ASSESSMENT FRAMEWORK
Repeatability	Measuring the variance or the dispersion range of GDM values within the repeated evaluation process
Robustness	Applying a statistical test a) with different population size (N) as sample b) with different landscape dimensions or variables (n) as sample
Outlier handling capability	Measuring the diversity differences between a population without any outliers and a population with a fraction of the individuals acting as outliers

Olorunda and Engelbrecht (2008) compare six GDMs on four test functions treated with a particle swarm optimization (PSO) approach. They rank the diversity measures according to their sensitivity to outliers. In contrast, Wineberg and Oppacher (2003) show that variance-based diversity measures, as well as the gene frequency family, are variants of the same basic concept: the sum of the distance between all possible pairs of elements considered. They conclude that experiments are not required for selecting the best GDM. However, the previous chapter presented very different conclusions. From the comparison over 15 GDMs, results show that the mean pairwise distance between the individuals in the population (D_{PW}) yields better diversity descriptors than other GDMs. Nevertheless, the response of D_{PW} is inadequate when convergence appears over multiple locations.

This leads to the question of whether or not D_{PW} and the other distance-based measures are capable of describing population diversity efficiently. If they are not, then the quality criteria in Table 2.1 would seem to be insufficient for appropriate diversity measure selection, and the following question has to be answered: *Do any available GDMs truly reflect population diversity?*

To the best of the authors' knowledge, no framework is available in the literature to validate the capability of a GDM as a diversity monitoring indicator. Not only must such a framework be provided, but a reliable GDM formulation must be identified to ensure accurate description of search behavior. This chapter addresses these challenging issues.

The chapter is organized as follows: the next section provides the background of our GDM validation study; section 2.3 introduces diversity requirements for GDM validation purposes; section 2.4 describes the behavior of typical GDM with respect to the proposed validation framework; and section 2.5 presents our concluding discussion.

2.2 Problem statement

The following simulations illustrate the response of two GDMs (D^N_{PW} and GF^N_S). These results were obtained from the generic benchmark developed in chapter 1 that simulates the convergence process of a population over single-site and multi-site locations. This process is depicted in Figure 1.1, where the rectangles represent the hyperspace allowed to the individuals associated with a given optimum. This generic benchmark does not account for the fitness distribution. Instead, the optima are randomly defined over the landscape at the beginning of the process. The hyperspaces shrink over a 51 iteration schedule, until all the individuals pile up on their respective optimum. The proposed convergence is simulated without any genetic operator, and the individuals are randomly generated at each iteration within their hyperspace boundaries. This generic benchmark eliminates any search bias coming from the operator. The simulations presented were conducted with a population of 100 individuals over a two-dimensional landscape.

Figure 2.1 presents the diversity mean value history for 50 repetitions with the normalized version of D_{PW} (D^N_{PW}). The normalization is based on the maximum diversity achieved so far in the optimization process (NMDF) as defined in section 1.2.2. Figure 2.1 also includes the normalized Shannon entropy (GF^N_S) (Shannon, 1948), which is a recognized gene frequency measurement. The normalization of GF_S is realized with its maximum value. This is achieved when the gene frequencies are similar over the gene intervals (M), which means that the fraction of the population ($p_{m,k}$) belonging to interval $m \in \{1,\ldots, M\}$ on gene $k \in \{1,\ldots,n\}$ must equal $1/M$, where n stands for the landscape dimensionality. In this experiment, M was set to 100. However, it is important to note that this is true only if $M \leq N$, where N represents the population size. Otherwise, the maximum value is obtained when $p_{m,k} = 1/N$. The formulation of D_{PW} and GF_S is given by equations 2.1 and 2.2 respectively.

$$D_{PW} = \frac{2}{N(N-1)} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \sqrt{\sum_{k=1}^{n} \left(x_{i,k} - x_{j,k} \right)^2} \tag{2.1}$$

$$GF_S(M) = -\frac{1}{n}\sum_{k=1}^{n}\sum_{m=1}^{M} p_{m,k}\log\left(p_{m,k}\right) \tag{2.2}$$

In equation 2.1, $x_{i,k}$ and $x_{j,k}$ represent the location of gene $k \in \{1,\ldots, n\}$ of the individual i and $j \in \{1,\ldots, N\}$ respectively.

As indicated in Figure 2.1, the D^N_{PW} end diversity estimations are 48% and 75% for the two- and four-optima landscapes respectively. This obviously represents an overestimation of the true population diversity, since the final population (iteration 51) is concentrated at two/four sites. This overestimation behavior results from the deficient treatment of duplicate individuals in D^N_{PW} (Ulrich, Bader, and Thiele, 2010), as in other distance-based GDMs (Lacevic, Konjicita, and Avdagic, 2007), (section 1.5.2). In contrast, GF^N_S seems to better describe the end diversity at convergence for multi-site processes. Nevertheless, as can be seen in Figure 2.1, GF^N_S does not offer representative diversity discrimination during the

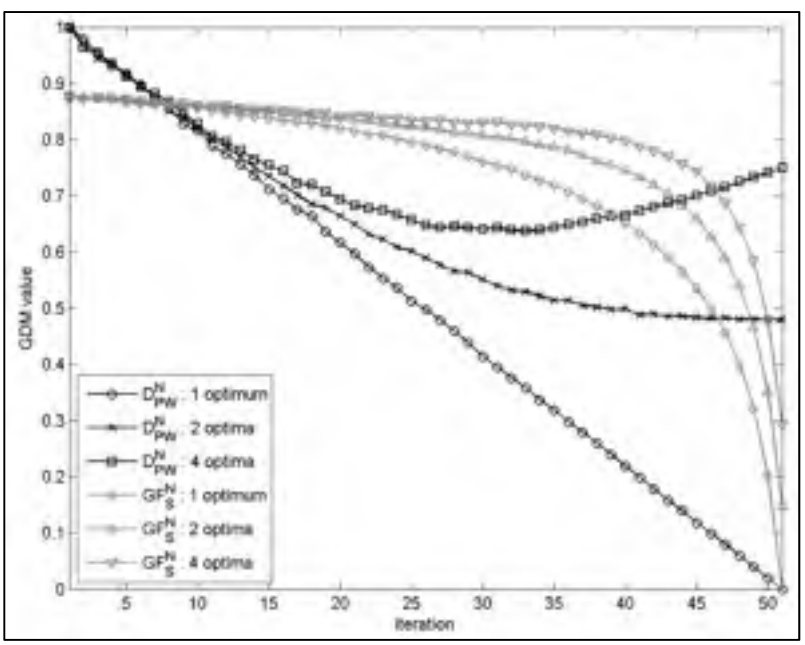


Figure 2.1 Genotypic diversity levels of D^N_{PW} and GF^N_S over the single-site and multi-site convergence processes (two and four optima)

process, even for single-site convergence. This is explained by the fact that all GF measures are based on the proportion of individuals resident in the various intervals for each gene, and the location of these intervals over the gene axis is not considered (section 1.5.1).. In other words, the diversity variations become obvious only when most of the individuals pile up in the same interval.

The previous observations indicate that none of these GDMs seems to be valuable over the multi-site convergence process. This makes assessing the underlying performance of any diversity promoting strategy troublesome. Moreover, even though standard EA do not commonly follow a multi-site convergence scheme, duplicate individuals are always a possibility throughout the evolution process. Consequently, any population-based search process may suffer from diversity distortion and so mislead the search behavior analysis.

In response to the weakness of the previous indicators, Lacevic, Konjicija, and Avdagic (2007) proposed the volume-based measure (L-diversity) as the GDM. They argued that it is probably the most intuitive and accurate way to evaluate diversity of a population. This measure is designed to compute the volume of the union of n axis-aligned hyper-rectangles. In computational geometry, this is known as the Klee measure problem (KMP) (Klee, 1977), and it represents a generalization of the dominated hypervolume measure used in multi-objective optimization problems (MOOP) for assessing the approximation quality of the Pareto front (Beume and Rudolph, 2006). The L-diversity, referred to here as D_L, is given by:

$$D_L = \mu_L\left(\bigcup_{i=1}^{N} S(\boldsymbol{x}_i, l)\right) \tag{2.3}$$

where μ_L(A) represents the Lebesgue measure of a set A. The parameter l corresponds to the length of the side of a hypercube $S(\boldsymbol{x}_i, l)$ bounding the diversity contribution of the individual $\boldsymbol{x}_i$. Setting $l = \sqrt[n]{V/N}$ promotes full coverage of the search space volume (V) when the individuals are uniformly distributed. D_L suffers from its computational complexity exponentially growing with respect to the dimensionality of the landscape, leading to

$O\left(N^{\lfloor n/2 \rfloor}\right)$ when all the hypercubes have the same size (Boissonnat *et al*., 1995). This condition makes D_L practically intractable as GDM.

This problem led Lacevic, Konjicija, and Avdagic (2007) to searching which measure best approximates D_L. They based their investigation on a correlation analysis over various controlled population arrangements. This study was later extended to include more GDMs in (Lacevic and Amaldi, 2011). As a result, the Euclidean minimum spanning tree measure (D_{MST}) turns out to be the best alternative to D_L. Its formulation is defined by:

$$D_{MST} = \mu\left(MST\left(G\left(X,E\right)\right)\right) \tag{2.4}$$

where $MST(G(X,E))$ represents the minimum spanning tree subgraph of the complete undirected graph $G(X,E)$, which is defined by the set X representing the location of the individuals of the population and the set of edges E denoting all the pairwise connections between individuals. The summation of the total length of the *MST* subgraph is symbolized by μ. The rationale behind the D_{MST} proposal is to extract only the "principal" distances, in order to alleviate the issue of duplicated individuals (Lacevic and Amaldi, 2011). Figure 2.2 illustrates a 2D example of the diversity evaluation mechanism of D_L and D_{MST}.

In addition, Lacevic and Amaldi (2011) developed the theoretical concept of ectropy for evaluating to what extent an indicator penalizes duplicate individuals. The ectropy concept helped justify the use of D_L as a reference in the correlation study; the maximum evaluation of D_L is never obtained in presence of duplicate individuals. Ectropy was also used for illustrating the weakness of D_{PW} and other distance-based measurements. However, ectropy analysis was restricted to a limited set of GDM, due to the difficulty of analytically deriving the maximal state of any formulation. This analysis illustrates the limited capacities of theoretical development in assessing the relevance of GDMs.

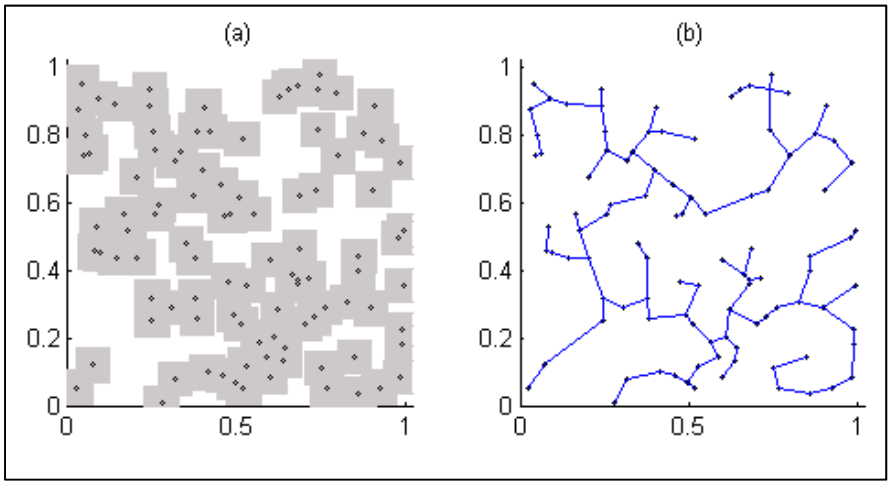


Figure 2.2 Representation of a uniformly random population with 100 individuals bounded between $[0, 1]^2$, where diversity is evaluated by: a) D_L – union of the area associated with each individual, b) D_{MST} – total length of the *MST*

2.3 Characterizing population diversity

As mentioned, numerous GDM formulations are available as well as different analysis frameworks for their comparison. However, the lack of precise feature characterizing population diversity makes the choice of the best measure problematic. Defining such requirements may provide common ground for validating which GDM accurately describes population diversities.

In pioneering research, Weitzman (1992) listed fourteen salient characteristics of reliable measures. Among them, six are considered to be mathematical characteristics, two are categorized as taxonomic aspects, one is an ecological consideration, and five are economic considerations. The Weitzman properties are summarized in Table 2.2. Weitzman acknowledged that these properties are not equally important. Later, Solow and Polasky (1994) identified three of them as fundamental requirements:

1. Monotonicity in species: adding a species (an individual, in the current context) should not decrease the diversity D, or $D(P') \leq D(P)$, if P' is a subset of population P.
2. Twinning: the addition of an individual or a species already in the population should not increase the diversity D, or $D(P \cup x_i) = D(P)$, if the distance between individuals x_i and x_j, $d(x_i, x_j) = 0$, where $x_j \in P$ and $x_i \notin P$.
3. Monotonicity in distance: an unambiguous increase in distance between individuals should be reflected in the diversity measurement D, or $D(P') \leq D(P)$. This requirement reflects the following situation, all the elements in population P equal those in population P', except individuals x_i and x_j from population P, and x_i' and x_j' from population P', while their distances respect the following inequality: $d(x_i', x_j') \leq d(x_i, x_j)$.

Table 2.2 Diversity properties defined by Weitzman (1992)

#	PROPERTY	CATEGORY
1	Monotonicity in species	Mathematical
2	Twin property	Mathematical
3	Twin property	Mathematical
4	Continuity in distance	Mathematical
5	Monotonicity in distance	Mathematical
6	Maximum diversity can be added by new species	Mathematical
7	Disaggregation	Taxonomic
8	Taxonomic distances reduce diversity theory in distribution theory	Taxonomic
9	Removal of false dissimilarity identifying individuals from the same species	Ecological
10	Preservation of biologically related species	Economic
11	Endeavour of equally distant relatives	Economic
12	Rule of the snake	Economic
13	Additivity properties of functional utility functions	Economic
14	Markov extinction	Economic

Even though the diversity measures studied by Weitzman (1992) and Solow and Polasky (1994) were not formulated for the present context, the proposed fundamental requirements are still suitable for evaluating GDM trueness in reflecting a diversity measure. In reality,

diversity measurement should be understood as a coverage space indicator. This concept is completely and rigorously expressed by those diversity requirements. Therefore, the three requirements are adapted to EA real-coded GDMs in Table 2.3.

Table 2.3 Defined requirements for GDM trueness validation

#	REQUIREMENT	BRIEF DESCRIPTION
1	Monotonicity in individual varieties	Adding a non-duplicate individual should decrease diversity - A uniformly distributed population provides upper bound diversity
2	Twinning	Duplicate individuals should reduce diversity as the population moves away from a uniformly distributed population
3	Monotonicity in distances	Diversity should decrease as individuals move closer together

Monotonicity in species will be referred to here as *monotonicity in individual varieties*. This is a more general expression, and is applicable in the EA context, since maximal diversity is achieved with a uniformly distributed population ($U(P)$). Such a population is constructed by ensuring that, on each gene, individuals are separated by the same distance. This distance is defined by $(UB_k\text{-}LB_k)/(N^{1/n}\text{-}1)$, where LB_k and UB_k represent the lower and upper bounds of the landscape k axis ($k \in \{1,..., n\}$) respectively. This requirement establishes the upper bound of the possible diversity of a population. The mathematical formulation becomes $D(P') \leq D(P) \leq D(U(P))$, where P' is a subset of population P.

The initial definition of the *twinning* requirement is directly transferrable to the present context. However, for fixed population sizes, the existence of duplicate individuals inevitably reduces the diversity of a population. The mathematical form becomes $D((P \setminus x_q) \cup x_i) \leq D(P)$ if $d(x_i, x_j) = 0$, where $x_j \in P, x_i \notin P$, and x_q is an individual removed from the population P to make room for x_i. As a matter of fact, the twinning requirement has the same meaning as the ectropy concept described before.

Finally, the requirement of *monotonicity in distance* is reformulated to highlight the fact that genotypic diversity should be based on the location of the various individuals. For example,

considering two uniformly distributed populations ($U(P_A)$ and $U(P_B)$) over region A and B of the same landscape, the corresponding diversities should present the following relation: $D\left(U\left(P_A\right)\right) < D\left(U\left(P_B\right)\right)$, if $\prod_{k=1}^{n}\left(UB_k - LB_k\right)_A < \prod_{k=1}^{n}\left(UB_k - LB_k\right)_B$.

2.4 Validation of the representative GDMs

In this section, only representative GDMs are considered. Therefore D_{PW} and GF_S, with M = 10, are selected to characterize common distance-based and gene frequency measurements respectively, while D_L and D_{MST} are included as potential GDM candidates following the recommendation in (Lacevic and Amaldi, 2011). The validation analyses the response of the GDMs to three diversity requirements on two frameworks: a reduced population arrangement, and various controlled cases of population diversity, as explained below.

2.4.1 Reduced population arrangement

The first framework intends to validate the general behavior of the GDMs in a simple an intuitive manner. A population of 5 individuals (P_5) is promoted on a 2D landscape bounded between $[0, 1]^2$, four of these individuals are fixed at the landscape corner ($\boldsymbol{x}_1$ = (0, 0), $\boldsymbol{x}_2$ = (1, 0), $\boldsymbol{x}_3$ = (1, 1), $\boldsymbol{x}_4$ = (0, 1)), and the remaining individual ($\boldsymbol{x}_5$) is moved on the diagonal connecting $\boldsymbol{x}_1$ and $\boldsymbol{x}_3$. This framework makes it possible to break down the multivariate aspect of GDM into a univariate problem by tracking the diversity variation of the normalized location of $\boldsymbol{x}_5$ ($\hat{x}_5$). For comparative purposes, diversity of a static population with 4 individuals (P_4) located at $\boldsymbol{x}_1$ to $\boldsymbol{x}_4$ is also included.

To respect the diversity requirements established in Table 2.3, the following conditions must be satisfied:

1. Monotonicity in individual varieties: $D\left(P_5 \middle| \hat{x}_5 = 0.5\right) \geq D\left(P_4\right)$;

2. Twinning: $\min D(P_5) = D\left(P_5 \middle| \hat{x}_5 = 0 \vee \hat{x}_5 = 1\right)$;

3. Monotonicity in distance: $\max D(P_5) = D\left(P_5 \middle| \hat{x}_5 = 0.5\right)$.

The results of this framework are presented in Figure 2.3. The charts indicate that D_{PW} is unable to respect any of the diversity requirements, as $D(P_5)$ is always lower than $D(P_4)$ and the maximum diversity state of P_5 is achieved in the presence of duplicate individuals $\left(\hat{x}_5 = 0 \wedge \hat{x}_5 = 1\right)$. A similar conclusion may be drawn for D_{MST}. It is interesting to note, however, that D_{MST} gives the same diversity for P_5 with $\boldsymbol{x}_5$ at boundaries (duplicate individuals) than for P_4. This is obvious from the *MST* computation standpoint, but it demonstrates that D_{MST} has a problem penalizing duplicate individuals. In fact, this issue stems from the disagreement between the summation of the "principal" distances and the monotonicity in distance. In other words, the diversity level of D_{MST} with P_5 is neither

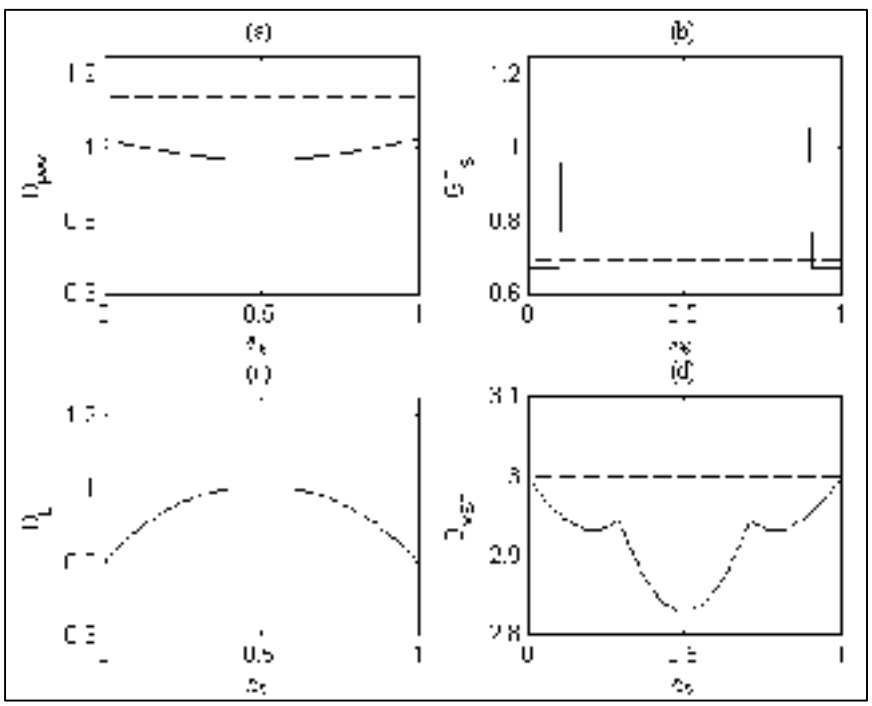


Figure 2.3 Diversity on P_5 (solid curves) and P_4 (dash curves) with respect to the normalized location of $\boldsymbol{x}_5$ evaluated from: a) D_{PW}, b) GF_S, c) D_L, d) D_{MST}

monotonic nor decreasing, as $\boldsymbol{x_5}$ comes closer to $\boldsymbol{x_1}$ or $\boldsymbol{x_3}$. As a matter of fact, the local peaks obtained by D_{MST} around $\hat{x}_5 = 0.3$ and $\hat{x}_5 = 0.7$ are due to changes in the *MST* connections.

On this reduced population framework, GF_S and D_L show good respect of all three diversity requirements. Clearly, diversity level discrimination is better for D_L than for GF_S. This is in accordance with the issue described in section 2.2 on gene frequency measurements. Furthermore, the evaluation of D_L, $D\left(P_5 \middle| \hat{x}_5 = 0.5\right) = D\left(P_4\right)$, is more likely to conform with the population size robustness criterion (Table 2.1). Having the upper bound of the diversity included in the D_L formulation, through the definition of l, makes the measurement independent of the population size parameter.

2.4.2 Controlled cases of population diversity

The second GDM trueness validation framework involves the examination of seven frozen cases of population diversity. Besides the difference in the population arrangements, the benefit of this framework is a better representation of common EA population sizes. A population size (N) of 100 is used for all cases on a 2D landscape bounded between $[-1, 1]^2$. These simple deterministic cases allow us to illustrate the three requirements listed in Table 2.3, while at the same time avoiding costly simulations. Of the seven cases, which are defined below, four are directly related to the modality of the landscape (individuals attached to predefined optima (Cases 2 to 5)).

Case 1: The population is fixed at one point on the landscape.

Case 2: The population is distributed with a uniform ratio on the optima located at a mid-point between the landscape center and corners.

Case 3: The population is distributed with a non-uniform ratio on the optima located at a mid-point between the landscape center and corners.

Case 4: The population is distributed with a uniform ratio on the optima located at the corners of the landscape.

Case 5: The population is distributed with a non-uniform ratio on the optima located at the corners of the landscape.

Case 6: The population is distributed uniformly over the landscape diagonal.

Case 7: The population is distributed uniformly over the landscape.

Case 1 and 7 simulate the complete convergence and full diversity conditions of a genotypic population respectively. Cases 2 and 3 and Cases 4 and 5 offer an identical geographical position. However, in Cases 3 and 5, one optimum monopolizes 70% of the individuals, with the rest equally distributed over the remaining optima. Figure 2.4 presents the geographical map of the population for these cases. To validate the coherence of GDMs response over multi-site locations, a two- and four- optima landscape are considered for these four cases. Therefore, for Cases 3 and 5 with two-optima, the individuals match the 70/30 arrangement, while for the four-optima landscape, the individuals follow a 70/10/10/10 distribution. Case 6 corresponds to a situation where an individual would only have identical gene values, with those values evenly spaced among individuals. This is described by $x_{i,k} = LB_k+(i-1)*(UB_k-LB_k)/(N-1)$, $\forall\ k \in \{1,\ldots, n\}$, where $i \in \{1,\ldots, N\}$. In such a situation, the individuals would be distributed along a landscape diagonal.

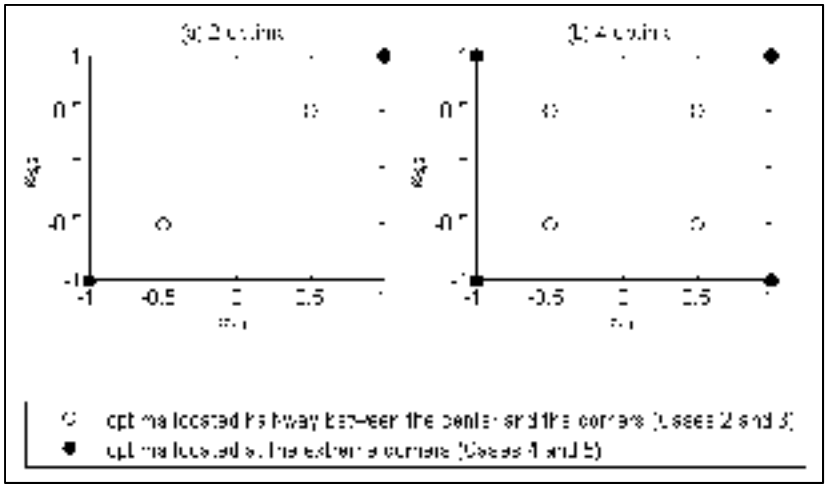


Figure 2.4 Position of the optima for Cases 2 to 5 on: a) two-optima landscape, b) four-optima landscape

The frozen case list is presented in increasing order of diversity. Consequently, to respect the 1st requirement in Table 2.3, an adequate GDM will rank the cases in the same order. Moreover, considering the twinning, Case 2 will be equal to Case 3, and Case 4 will be equal to Case 5. Finally, the monotonicity in distance is accounted for if Cases 4 and 5 present higher diversities than Cases 2 and 3. Table 2.4 presents the results obtained for all cases with D_{PW}, GF_S, D_L, and D_{MST}. Within this framework, the interval number (M) used by GF_S is set to 100.

Since the diversity levels obtained are higher for the cases where the individuals are located at the landscape corners (Cases 4 and 5) than for Case 7, Table 2.4 indicates that D_{PW} does not respect the 1st requirement. In addition, the diversity estimations for Cases 2 and 4 are higher than for Cases 3 and 5 respectively. This reveals the additional contribution of the duplicate individuals within D_{PW}, which indicates that the 2nd requirement is not respected either. In addition, since Case 6 exhibits a lower diversity than Cases 4 and 5, Table 2.4 also reveals that D_{PW} does not fulfill the requirement of monotonicity in distance. Based on these observations, the frozen case experiment accurately reflects the observed shortcoming of distance-based measures over the multi-site convergence process (section 2.2).

Table 2.4 also indicates that GF_S violates all three requirements. Diversity assessment by the aggregation of each gene leads to violation of the 1st requirement, since considering each

Table 2.4 Behavior of the representative GDMs over the seven frozen cases

GDM	LANDSCAPE	GENOTYPIC DISTRIBUTION CASES						
		1	2	3	4	5	6	7
D_{PW}	2 epitopes	0	0.71	0.00	1.41	1.00	0.96	1.16
	1 epitope		0.86	0.00	1.72	1.00		
GF_S	2 epitopes	0	0.69	0.01	0.69	0.01	0.61	3.40
	1 epitope		0.69	0.00	0.69	0.00		
D_L	2 epitopes	0.04	0.08	0.08	0.08	0.08	0.80	1.00
	1 epitope		0.16	0.16	0.16	0.16		
D_{MST}	2 epitopes	0	1.41	1.41	2.83	2.83	2.83	32.00
	1 epitope		3.00	3.00	6.00	6.00		

gene independently increases the diagonal distribution (Case 6) diversity estimation. This observation demonstrates that the generalization by aggregation of a univariate diversity indicator into a multivariate framework can be problematic. The 2nd requirement is not respected, since the distribution of the duplicate individuals impacts the diversity level (Case 2 ≠ Case 3, and Case 4 ≠ Case 5). Finally, the diversity level does not decrease as the optima move closer to one another (Case 2 $\nless$ Case 4, and Case 3 $\nless$ Case 5), and so the 3rd requirement is not respected either.

The results of Table 2.4 indicate that D_L cannot respect the requirement of monotonicity in distance; because D_L aggregates the volume covered by each individual regardless of their locations, the descriptor makes no difference between the optima location (Case 2 = Case 4, and Case 3 = Case 5).

Finally, Table 2.4 indicates that D_{MST} violates the requirement of monotonicity in distance, and to some extent that of monotonicity in individual varieties. In fact, no distinction appears between cases with all individuals fixed at the corner (Case 4 and Case 5) and cases with individuals set on the landscape diagonal (Case 6). These cases share the same *MST*, although the diversity state of Case 6 is higher than that of Cases 4 and 5.

2.4.3 Discussion

Table 2.5 summarizes the results obtained from the two frameworks. The superscripts indicate the framework revealing the deficient response. The aggregation of these results demonstrates that the two frameworks, taken individually, are insufficient for a complete validation of GDMs. On the other hand, associated, they offer efficient validation of GDM performances. In addition, Table 2.5 particularly reveals that none of the studied GDM guaranties accurate description of the population diversity. We are therefore forced to conclude that all evaluated measurements could represent a misleading factor in monitoring diversity.

Table 2.5 Summary of the fulfillment of the diversity requirements by the representative GDMs (A – Violation identified through the reduced population arrangement framework, B – Violation identified through the controlled cases of population diversity framework)

GDM	REQUIREMENTS		
	1 Monotonicity in individual varieties	2 Twinning	3 Monotonicity in distance
D_{Lev}	$NC^{A,B}$	$NC^{A,B}$	$NC^{A,B}$
GIT_S	NC^{A}	NC^{A}	NC^{A}
$D_{\cdot}$	YES	YES	NC^{A}
D_{HSC}	$NC^{A,B}$	NC^{A}	$NC^{A,B}$

2.5 Conclusion

GDM is a useful concept for monitoring and/or managing the exploration of an optimization process. Premature convergence towards a suboptimal solution can be minimized through strategies using the information gathered by a GDM. Multiple GDMs have been proposed in the literature over the years. However, to the best of the authors' knowledge, their ability to describe population diversity has never been exhaustively investigated. In GDM-related applications as well as in GDM comparison study, the assumption that a particular GDM truly reflects population diversity is often adopted. However, the issues observed with some of these formulations, such as poor handling of duplicate individuals, lead us to question the trustworthiness of this premise. Consequently, using a GDM not fulfilling this assumption can potentially disrupt the analysis of the search process.

The aim of this chapter is to look at the development of a framework that allows GDM to be assessed as population diversity descriptors. To achieve this, we extracted three diversity requirements from the literature to form the basis for our investigation. The requirements are: monotonicity in individual varieties, twinning, and monotonicity in distance. These diversity

requirements are intuitive properties that GDM must have, in order to offer an accurate coverage space description. Our study here is restricted to real-coded representation, although the established diversity requirements are not limited to this context. We identified and evaluated four GDMs from previous studies: the mean pairwise measure (D_{PW}), the Shannon entropy (GF_S), the L-diversity or volume-based measure (D_L), and the minimum spanning tree measure (D_{MST}).

The response of the selected GDMs to the requirements was evaluated by means of two validation frameworks involving a reduced population arrangement of 4 and 5 individuals, and seven test cases with controlled population diversity. These simple frameworks showed that the three diversity requirements are sufficient for proper evaluation of the GDM response. The frameworks also served to identify and characterize the limitations of the available GDMs.

In summary, D_{PW}, GF_S, and D_{MST} showed improper response to all three diversity requirements. Mostly because they do not consider a uniformly distributed population as the most diverse state. They also present some difficulties in managing duplicate individuals and cannot efficiently account for relative locations of the individuals within the population. On the other hand, D_L was revealed to be the sole formulation able to meet two of the three requirements. Nevertheless, besides its prohibitive computational cost, it offers no reliable mechanism to account for the requirement of monotonicity in distance. As illustrated by the controlled cases of population diversity framework, its failure to meet the third requirement could impact the diversity analysis when the population is configured in non intersecting clusters.

Globally, the present investigation demonstrated that the definition of an adequate genotypic diversity formulation for real-coded representation remains an open question. Moreover, the proposed GDM validation frameworks will facilitate the evaluation of any new proposals, by relating simple cases of controlled diversity to the fundamental requirements that the diversity descriptor must exhibit. It important to mention that even if the proposed GDM

validation framework combination was sufficient for detection of inadequate response of the tested GDMs, the reciprocal should not be assumed: the framework combination alone remains insufficient to guaranty the validity of a given GDM. The proposed evaluation tool should only be considered as a first gate, since the GDM must be tested, thereafter, within higher dimensionality landscapes.

We caution, as a general recommendation, that care should be exercised regarding the generalization of a univariate diversity indicator by aggregation into a multivariate context for GDM purposes. In addition, special attention should be paid to monotonicity in distance during the development of new formulations, since no GDM tested was able to completely meet this diversity requirement.

In the next chapter, we provide a similar study, oriented towards phenotypic measurements.

CHAPTER 3

REVIEW OF PHENOTYPIC DIVERSITY FORMULATIONS FOR DIAGNOSTIC TOOL

Practitioners often rely on search results to learn about the performance of a particular optimizer as applied to a real-world problem. However, even the best fitness measure is often not precise enough to reveal the behavior of the optimizer's added features or the nature of the interactions among its parameters. This makes customization of an efficient search method a rather difficult task.

The aim of this chapter is to propose a diagnostic tool to help determine the impact of parameter setting by monitoring the exploration/exploitation balance (EEB) of the search process, as this constitutes a key characteristic of any population-based optimizer. It is common practice to evaluate the EEB through a diversity measure. For any diagnostic tool developed to perform this function, it will be critical to be able to certify its reliability. To achieve this, the performance of the selected measure needs to be assessed, and the EEB framework must be able to accommodate any landscape structure. We show that to devise a diagnostic tool, the EEB must be viewed from an orthogonal perspective, which means that two diversity measures need to be involved: one for the exploration axis, and one for the exploitation axis. Exploration is best described by a genotypic diversity measure (GDM), while exploitation is better represented by a phenotypic convergence measure (PCM). This study includes a complete review of PCM formulations, and compares nearly all the published PCMs over a validation framework involving six test cases that offer controlled fitness distribution. This simple framework makes it possible to portray the underlying behavior of phenotypic formulations based on three established requirements: monotonicity in fitness varieties, twinning, and monotonicity in distance. We prove that these requirements are sufficient to identify phenotypic formulation weaknesses, and, from this conclusion, we propose a new PCM, which, once validated, is shown to comply with all the above-mentioned requirements. We then compare these phenotypic formulations over three

specially designed fitness landscapes, and, finally, the new phenotypic formulation is combined with a genotypic formulation to form the foundation of the EEB diagnostic tool. The value of such a tool is substantiated through a comparison of the behaviors of various genetic operators and parameters.

This chapter was published in *Applied Soft Computing* (Corriveau *et al.*, 2013).

3.1 Introduction

To estimate the performance of a particular optimizer, practitioners commonly rely on search results, such as the best fitness. However, this information alone may not reveal the underlying behavior of a customized search strategy. Furthermore, theories in the field of metaheuristics are generally difficult to translate into the realities of real-world problems. In fact, these theories are usually either restricted to specific landscape problems or derived for an isolated component of the search process (Eiben and Smith, 2003). For these reasons, and considering the No Free Lunch (NFL) theorem, which stipulates that no one optimizer can dominate in all situations (Wolpert and Macready, 1997), designing an efficient search strategy may be difficult. In this chapter, efficiency refers to the ability to find a valuable solution, or solutions, in the shortest possible time.

Diagnostic tools for optimizers may help practitioners determine the impact of different strategies implemented during the search process. More importantly, the information gathered can serve to devise a better search strategy, customized for the problem at hand.

Monitoring the search exploration/exploitation balance (EEB) offers a valuable description of the *working of an algorithm* (Eiben and Schippers, 1998). In other words, as it is responsible for the specific search path pursued, the EEB may be regarded as a basic efficiency characteristic for any population-based optimizer. The EEB summarizes the way in which resources are allocated. Samples directed toward exploration help in the gathering of knowledge on infrequently visited landscape areas, while exploitation relates to resources

dedicated to digging in promising regions. Clearly, excessive exploration can lead to random searching and a waste of computational resources. At the same time, excessive exploitation can lead to local searching and convergence to a suboptimal solution. In fact, what is needed for conducting a search over unknown landscapes with limited resources is a precise EEB, and tools that capitalize on EEB information can be a powerful means for diagnosing the impact of a search strategy and for selecting the best combinations of search parameters.

In a similar line of thought, Bassett and De Jong (2011) have provided an evolutionary algorithm (EA) customization tool for monitoring the EEB, with the aim of diagnosing customized reproductive operators. They use multivariate quantitative genetics theory to develop two indicators, perturbation and heritability. The former describes exploration capacity, and the latter estimates exploitation capability. However, this customization tool does not support search component interaction. Turkey and Poli (2012) considered a different approach to describe the emergent collective behavior of population-based search process. They used a self-organizing map (SOM), which is a kind of artificial neural network, for tracking the population dynamics. With this system, they extracted multiple properties for characterizing the EEB. Nevertheless, the impact of the SOM parameters, such as grid size and training approach, on the quality of the retrieved EEB features remains unclear.

Our objective here is to develop a diagnostic tool based on population diversity formulations for indicating the optimizer EEB. Two kinds of diversity descriptors can be used to define this framework: genotypic diversity measures (GDMs), and phenotypic diversity measures (PDMs). GDMs characterize the spatial distribution of the population, whereas PDMs depict its fitness distribution, and so refer to the quantity and the quality of the population diversity respectively (Herrera and Lozano, 1996).

To develop the diagnostic tool, two underlying objectives must be achieved. First, the role of both diversity measures must be established. Moreover, since numerous diversity measures have been proposed in the literature over the years, the efficiency and reliability of these formulations must be established. Some studies, as the one proposed in chapter 1, compare

the similarities and differences of GDMs (Wineberg and Oppacher, 2003; Olorunda and Engelbrecht, 2008). However, to the best of the authors' knowledge, no such study involving PDMs has been conducted. This leads us to state the second objective, which is to review and assess the performance of phenotypic formulations.

The chapter is organized as follows: in the next section, we show how the EEB can be represented through diversity measures; in section 3.3, we review the phenotypic formulations proposed in the literature; in section 3.4, we propose a validation framework and analyze some phenotypic formulations; in section 3.5, we develop and validate a new phenotypic formulation; in section 3.6, we compare all these formulations over specially designed landscapes; in section 3.7, we establish and assess the desirable qualities of a formulation; in section 3.8, we present the proposed diagnostic tool and describe it through an application in a genetic algorithm (GA) parameter setting context; finally, in section 3.9, we conclude the chapter.

3.2 EEB concept

The EEB can be viewed in terms of one of two paradigms (Gupta, Smith, and Shalley, 2006): 1) exploration and exploitation act as opposing forces, where increasing one reduces the other; or 2) they can be considered as orthogonal forces. This second perspective offers the possibility of increasing both exploration and exploitation simultaneously. In fact, it has been shown in section 1.1 that the opposing forces paradigm is a special case of the orthogonal forces paradigm.

Consequently, monitoring the EEB must involve two metrics: one for the exploration axis, and one for the exploitation axis. Exploration is best described by the genotypic formulation, as it summarizes the distribution of the individuals over the search space, while exploitation is best characterized by phenotypic formulations, as promising regions are targeted based on fitness information. This orthogonal EEB framework is illustrated in Figure 3.1. With

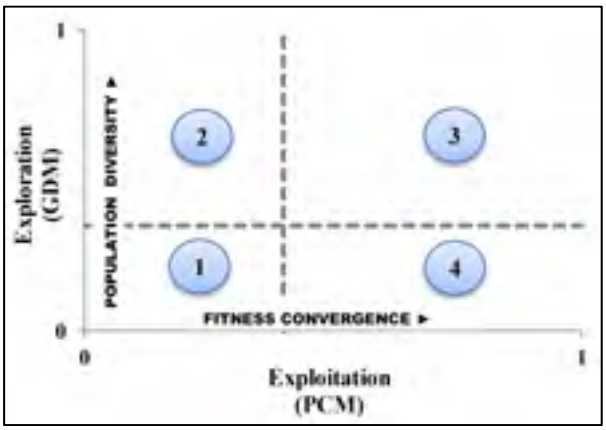


Figure 3.1 Orthogonal EEB framework with differentiation into four search zones

normalized evaluation, unitary genotypic and phenotypic values relate to maximum exploration and exploitation respectively.

According to this framework, exploration increases with a rise in genotypic diversity. In contrast, exploitation corresponds to the intensification of phenotypic convergence. To avoid confusion, we will refer to the phenotypic convergence measure (PCM) instead of the PDM when dealing with the EEB framework.

Since a mode can be generated from neighboring fitness values belonging to highly scattered individuals, phenotypic convergence should not be employed to terminate a search process. Generally, for population-based optimizers, advancing toward convergence indicates that individuals are becoming increasingly similar. Therefore, phenotypic convergence without genotypic convergence indicates that multiple solutions perform equally well. However, such a condition does not necessarily correspond to a multimodal landscape. In reality, it could refer to a “ring” formation produced by the individuals around a particular optimum. Despite this condition, representation of the orthogonal EEB framework through a GDM-PCM combination provides a way to depict the concept of *useful diversity* introduced by Goldberg and Richardson (1987). As they point out, preserving diversity by itself is not the ultimate goal; it is maintaining diversity that can lead to the identification of good individuals.

Finally, to further illustrate the value of the orthogonal EEB framework, four zones are proposed in Figure 3.1. Zone 1 characterizes a population with spatially similar individuals (low genotypic diversity) and heterogeneous fitness (low phenotypic convergence). Zone 2 is characterized by high exploration capability and low exploitation strength. A random search, for instance, would be located in this zone. Zone 3 is the useful diversity area discussed previously, in which exploration and exploitation are maximized simultaneously. Finally, in Zone 4, a searching process is directed toward converges to a single solution. The orthogonal EEB framework appears, therefore, to be more descriptive of the search process than the concept of opposing forces, where, to some extent, only the second and the fourth zones are distinguished.

3.3 Review of phenotypic formulations

For any landscape structure, the orthogonal EEB framework portrays the way resources are allocated, and, consequently, optimizer performance. In fact, the use of a phenotypic formulation is only justifiable from this perspective. To reduce computational effort, some researchers only consider phenotypic diversity (the EEB concept of opposing forces), on the assumption that fitness differences reflect genotypic space diversity (Motoki, 2002). This is a limitation, however, and few researchers using this approach take it into account (Neri, Toivanen, and Mäkinen, 2007; Caponio *et al.*, 2007; Tirronen and Nerri, 2009; Friedrich, Hebbinghaus, and Neumann, 2009). The following scenario illustrates the problem: A population of N individuals located on N different peaks of the same magnitude would be considered to be in a state of convergence from the phenotypic point of view, whereas from a genotypic perspective, the diversity would be clearly visible. Therefore, in the presence of an unknown landscape structure, relying solely on phenotypic measurement could be misleading in the search performance analysis.

Phenotypic formulations have frequently been involved in the heuristic formulations used to adapt EA parameters to control the EEB. However, modifying the EEB during a search considerably increases process complexity. Burke, Gustafson, and Kendall (2004, p. 48-49)

summarized the problem as follows: "*The type and amount of diversity required at different evolutionary times remains rather unclear*."

The first objective of this chapter having been met with the above detailed discussion on phenotypic measures, we present below nine formulations retrieved from the literature, as well as some variants adapted to the present context. This results in a total of 19 different PCMs.

3.3.1 General concept

In this study, minimization problems are considered, which requires the adaptation of some PCM formulations.

Phenotypic formulations can be evaluated from two perspectives: 1) based on distance-based measurement, where the measurements are estimated by the best fitness (f_{best}), the average fitness (f_{avg}), the worst fitness (f_{worst}), or the standard deviation of the fitness distribution ($\hat{\sigma}_f$) (evaluations based on the distances between solution responses are also possible, and may be evaluated according to a particular descriptive fitness (f_{best}, f_{avg}), or between each individual fitness measure); or 2) by scanning the fitness frequency of a population. However, two points have restricted the latter perspective from being adopted for phenotypic formulation evaluation. First, since the fitness distribution is generally continuous, the fitness space has to be partitioned. Also, the maximum range of the fitness values is unknown, unless the search space is completely enumerated, and so the partitioning process needs to be adaptive, to account for the extension of the fitness range.

3.3.2 Normalization

In this study, all PCMs are limited to unitary ranges. Full exploitation in the EEB framework is associated with a PCM value of 1, as the phenotypic convergence state is achieved, while a

0 value represents the maximum phenotypic diversity state. Aside from the advantage of normalized measures for comparison purposes, this makes the PCM equal to 1–PDM.

Some existing PCMs are normalized in their original formulation, while others rely on the normalization with maximum diversity so far (NMDF). However, this normalization is not suitable in a phenotypic context, as it could distort the measurement. In fact, NMDF assumes that the starting population is drawn from a uniform distribution. Since the fitness distribution is a function of the landscape relief, this assumption cannot be made. Consequently, this approach would then consider the initial phenotypic distribution as the most diversified state, regardless of its real level. We therefore propose the virtual maximum diversity (VMD) as a normalization alternative. For a given population size, this approach considers that the most diverse population state is achieved when the fitness distribution is uniformly distributed between the worst and the best values obtained up to that point in the optimization process. This means that the diversity is computed for a virtual population in which the fitness of the individuals is uniformly distributed over the absolute fitness range respecting a predefined distance $\left(\left|f_{worst} - f_{best}\right| / \left(N-1\right)\right)$. Since *VMD* is established for *N* individuals, it must be recalculated when the population size and/or the absolute fitness range are modified.

For fitness frequency measurements, the maximum diversity value is obtained when the fitness distribution is partitioned uniformly over the total number of intervals (*M*). The maximum value is calculated by setting $p_m = 1/M$ in the formulation, where p_m represents the fraction of the population size (*N*) that belongs to the interval *m*. However, for $N < M$, the maximum value is achieved when $p_m = 1/N$.

Finally, the presence of phenotypic outliers could lead to an overestimation of the convergence state, due to the widening of the absolute fitness range. However, in real-world problems, identifying phenotypic outliers is difficult, since they can represent unvisited regions, instead of a single extreme value.

3.3.3 PCM formulation

Table 3.1 presents the PCMs considered in our comparison, some of which were developed specifically for phenotypic distribution, while others were proposed in multivariate distribution contexts and so are reformulated here. The latter are marked with an asterisk in the table.

PCM_1 and PCM_2 are simple ratio indicators, whereas PCM_3 could be considered as an extreme ratio. Lee and Takagi (1993) used PCM_1, PCM_2, and the change in best fitness as inputs of fuzzy logic controllers for adapting GA parameters. Subbu, Sanderson, and Bonissone (1998) later proposed a similar adaptation scheme, in which they promote PCM_1 and a GDM based on the Hamming distance as inputs. Herrera and Lozano (1996) also used PCM_1 and a GDM based on Euclidian distance as inputs to their fuzzy logic controller. Finally, Vasconcelos *et al.* (2001) and Pellerin, Pigeon, and Delisle (2004) promoted the use of a PCM with the same meaning as PCM_1 to adapt GA parameters following heuristic rules.

PCM_4 represents a family of PCMs based on the difference between the average and the best fitness. This difference could serve as a phenotypic convergence detector (Srinivas and Patnaik, 1994). $PCM_{4.1}$ is normalized by the fitness range (Arnone, Dell'Orto, and Tettamanzi, 1994), while $PCM_{4.2}$ is the absolute version of $PCM_{4.1}$ proposed by (Neri *et al.*, 2007; Nerri, Kotilainen, and Vapa, 2008) to adapt parameters and activate local searchers with heuristic rules in a memetic algorithm (MA) context. Caponio, Neri, and Tirronen (2009) proposed $PCM_{4.3}$, which is an NMDF normalized version. They use this indicator with a hybrid algorithm to detect super-fit individuals, and thus activate different local searchers following heuristic rules. $PCM_{4.4}$ was proposed by Caponio *et al.* (2007). Again, it is used to adapt EA parameters and activate local search procedures following heuristic rules. $PCM_{4.4}$ was later used for other applications with similar adaptation rules (Neri, Toivanen, and Mäkinen, 2007; Neri *et al.*, 2006; Neri and Mäkinen, 2007). $PCM_{4.5}$ is the VMD normalized version proposed here.

Table 3.1 PCM formulations used for the comparative study

No.	PCM Formulation	No	PCM Formulation
1.	$PCM_1 = \frac{f_{best}}{f_{avg}}$	13.	$PCM_{13} = 1 - \frac{\frac{1}{N}\sum_{i=1}^{N}\lvert f_i - f_{avg}\rvert}{PCM_{12}}$
2.	$PCM_2 = \frac{f_{avg}}{f_{worst}}$	14.	$PCM_{14} = 1 - \frac{\sum_{i=1}^{N}(f_i - f_{avg})^2}{PCM_{12}}$
3.	$PCM_3 = \frac{f_{best}}{f_{worst}}$	15.	$PCM_{15} = 1 - \frac{\frac{2}{N(N-1)}\sum_{i=1}^{N}\sum_{j=i+1}^{N}\lvert f_i - f_j\rvert}{PCM_{12}}$
4.	$PCM_4 = 1 - \frac{f_{avg} - f_{best}}{f_{worst} - f_{best}}$	16.	$PCM_{16} = 1 - \frac{1}{\log(N)}\sum_{i=1}^{N} p_i \log(p_i)$
5.	$PCM_5 = 1 - \frac{f_{avg} - f_{best}}{\lvert f_{worst} - f_{best}\rvert}$	17.	$PCM_{17} = 1 - \frac{\sum_{i=1}^{N} p_i^2}{1 - N^{-2}}$
6.	$PCM_6 = 1 - \frac{f_{avg} - f_{best}}{PCM_{12}}$	18.	$PCM_{18} = 1 - \frac{\log\left(\sum_{i=1}^{N} p_i^2\right)}{(1-\alpha)\log(N)}$
7.	$PCM_7 = 1 - \min\left(\frac{f_{avg} - f_{best}}{f_{best}}, 1\right)$	19.	$PCM_{19} = 1 - \beta\sum_{i=1}^{N} p_i(1 - p_i)$ where, $\beta = \begin{cases} \frac{M}{(2M-1) - \frac{2(M-1)}{N^2}}, & \text{if } M \le N \\ \frac{N}{N-1}, & \text{otherwise} \end{cases}$
8.	$PCM_8 = 1 - \frac{f_{avg} - f_{best}}{PCM_{12}}$		
9.	$PCM_9 = 1 - \frac{\sigma_f}{\lvert f_{avg} - f_{best}\rvert}$		
10.	$PCM_{10} = 1 - \frac{\sigma_f}{\sqrt{\frac{(f_{avg} - f_{best})^2 + (f_{avg} - f_{worst})^2}{2}}}$		
11.	$PCM_{11} = 1 - \min\left(\frac{\sigma_f}{\lvert f_{avg}\rvert}, 1\right)$		
12.	$PCM_{12} = 1 - \frac{\sigma_f}{PCM_{12}}$		

The PCM_5 family is based on the standard deviation, or dispersion, of the fitness values. We consider the unbiased standard deviation in this study. $PCM_{5.1}$ was proposed by Tirronen and Neri (2009) to adapt differential evolution (DE) parameters following heuristic rules. $PCM_{5.2}$

is known as the degree of dispersion, and was proposed by Miao et al. (2009) to adapt particle swarm optimization (PSO) parameter. $PCM_{5.3}$ was promoted by Tirronen *et al.* (2007) to activate local searchers in MA following heuristic rules. In this PCM formulation, as for $PCM_{4.4}$, the minimum operator suggests normalization issues, since the unitary range is not guaranteed. $PCM_{5.4}$ is the VMD normalized version proposed in this study.

PCM_6 to PCM_8 are reformulations of multivariate diversity measurements (Ursem, 2002; Abbass and Deb, 2003; Morrison and De Jong, 2002; Barker and Martin, 2000). In the phenotypic context, PCM_6 describes the mean location of the fitness values with respect to the average fitness of the distribution. PCM_7 is based on the underlying idea of allocating more importance to fitness values away from the mean of the distribution. Finally, PCM_8 corresponds to the mean pairwise distance from all fitness values. The idea behind this PCM was used by Hutter and Legg (2006) to motivate the development of the fitness uniform selection scheme (FUSS).

PCM_9 to PCM_{12} belong to the fitness frequency category. This category involves the entropy concept which, at first sight, could be well suited to being a phenotypic descriptor, since it describes the level of disorder of a distribution. PCM_9 represents the Shannon entropy (Shannon, 1948). Rosca (1995) uses this formulation to correlate GP statistical measures to the phenotypic state with the aim of controlling the EEB, whereas Darwen (2000) uses it to compare problem-specific learning strategies involved in a GA optimizer. PCM_{10} and PCM_{11} are two other entropy families ($\alpha > 0$ and $\alpha \neq 1$) (Havrda and Charvát, 1967; Rényi, 1961). By letting $\alpha \rightarrow 1$, PCM_{10} and PCM_{11} tend toward PCM_9. In contrast, PCM_{12} is an approximation of PCM_9 (Wineberg and Oppacher, 2003). The variable u, shared by PCM_9 to PCM_{11}, stands for the normalization part, as $u = \min\{M, N\}$ (section 3.3.2). For PCM_{12}, there is a similar normalization. Nevertheless, in the original formulation, a correction term ($r = N \bmod M$) was considered for cases where M is not a common divisor of N.

As we have shown, most of the phenotypic indicators formulated in the literature have been used alone to describe the population's EEB state, and no performance analysis was conducted to assess the suitability of these various formulations.

3.4 Validation of phenotypic formulations

Validation of the phenotypic formulation selected is mandatory, in order to ensure the reliability of the EEB diagnostic tool. Since no framework is available in the literature, three diversity requirements are proposed to determine the relevance of phenotypic formulations. These requirements are validated by means of a deterministic frozen diversity case framework, which is a simple framework that can represent them efficiently. In order to avoid potential issues arising from normalization approaches, phenotypic formulations are considered here solely by studying their characteristics at the family level, which reflect their computed diversity.

3.4.1 Requirements for a suitable diversity measure

In pioneering research, Weitzman (1992) listed 14 principal characteristics of reliable diversity measures. Weitzman acknowledged that these properties are not equally important. Later, Solow and Polasky (1994) identified three of them as natural requirements:

1. *Monotonicity in species*: adding a species (or individuals, in the current context) should not decrease diversity or $D(P') \leq D(P)$, if *P'* is a subset of population *P*.
2. *Twinning*: the addition of an individual or a species already in the population should not increase the diversity or $D(P \cup i) = D(P)$, if $d(i, j) = 0$, where $j \in P$ and $i \notin P$.
3. *Monotonicity in distance*: an unambiguous increase in distance between individuals should be reflected in the measurement or $D(P') \leq D(P)$, if $d(i', j') \leq d(i, j)$.

The ideas governing these requirements apply to phenotypic measurement. In reality, the diversity measurement should be understood as a description of the coverage of the search

space. This concept is completely and rigorously expressed by those diversity requirements. Nevertheless, the three requirements must first be reformulated in terms of fitness distribution.

Species monotonicity will be referred to here as *monotonicity in fitness varieties*. This first quality specifies that diversity will increase with the addition of new fitness values. This implies that the maximum phenotypic diversity is produced by a uniform distribution (*U(F)*) over the fitness range. Therefore, the mathematical formulation is: $D(F') \leq D(F) \leq D\left(U\left(F\right)\right)$, where *F'* is a subset of the fitness distribution *F*.

The initial definition of the *twinning* requirement is directly transferable to the present context. However, for fixed population sizes, the presence of duplicate individuals inevitably reduces the diversity of a population. The mathematical form then becomes $D((F \setminus f_k) \cup f_i) < D(F)$, if $d(f_i, f_j) = 0$, where $f_j \in F, f_i \notin F$. Here, f_k is a non duplicated individual removed from the population *F*.

The *monotonicity in distance* requirement also corresponds to the *shuffling dependence property* (Cha and Srihari, 2002). This requirement states that permutation of fitness values impacts the phenotypic measurement directly. In this context, the mathematical formulation becomes: $D(F') \leq D(F)$, if $d(f_i^{'}, f_j^{'}) \leq d(f_i, f_j)$.

Table 3.2 lists and describes the final phenotypic formulation requirements, which will be shown to be sufficient for evaluating their relevance.

Table 3.2 Requirements of the phenotypic formulation

#	REQUIREMENT	BRIEF DESCRIPTION
1	Monotonicity in fitness variation	- Adding a fitness value should not decrease diversity - Uniformly distributed fitness provides upper bound diversity
2	Twinning	Implicitly fitness values should reduce diversity as the distribution moves away from the uniform distribution case (diversity upper bound)
3	Monotonicity in distance or shuffling dependence	Diversity should decrease as fitness values move closer together

3.4.2 Validation framework for the requirements analysis

Six deterministic cases of frozen fitness diversity are proposed to evaluate and illustrate the phenotypic formulation responses, as follows:

Case 1: All the individuals are located at f_{best}.

Case 2: 50% of the population is located at the mid-point between f_{worst} and f_{best}, while the remaining portion is at f_{best}.

Case 3: N-1 of the population is located at the mid-point between f_{worst} and f_{best}, while the remaining individual is at f_{best}.

Case 4: 50% of the population is located at f_{worst}, while the remaining portion is at f_{best}.

Case 5: N-1 of the population is located at f_{worst}, while the remaining individual is at f_{best}.

Case 6: The individuals are uniformly distributed over a predefined fitness range (*VMD* case).

The first case corresponds to a converged situation. Cases 2 and 3, and Cases 4 and 5 present equivalent phenotypic diversities. However, Cases 4 and 5 present higher diversities than Cases 2 and 3. Furthermore, Cases 2 to 5 have low phenotypic diversity (two fitness values). In contrast, Case 6 corresponds to the highest phenotypic diversity state.

During the tests, a population size of 100 and a total number of intervals of 100 are used. In addition, the fitness range is defined between 150 for f_{worst} and 50 for f_{best}.

3.4.3 Relevance of the phenotypic formulations

Table 3.3 presents the diversity levels obtained from each phenotypic family in Table 3.1 and applied to the validation framework.

Results indicate that all the phenotypic formulations identify the converged distribution (Case 1 = 0). However, none of them conforms to the diversity requirements.

In fact, families 1 to 8 violate the monotonicity in fitness varieties; Case 6 is not identified as the highest diversity level. All the descriptors found Case 4 or Case 5 to represent the highest diversity condition, even though they each involve only two fitness values.

Table 3.3 Behavior of phenotypic formulations over the six frozen case framework

FAMILY	DIVERSITY FORMULATION	PHENOTYPIC DISTRIBUTION					
		Case 1 N_{best}	Case 2 $50_{middle}/50_{best}$	Case 3 $(N-1)_{middle}/1_{best}$	Case 4 $50_{worst}/50_{best}$	Case 5 $(N-1)_{worst}/1_{best}$	Case 6 VMD
1	$1-f_{best}/f_{avg}$	0.00	0.33	0.50	0.50	0.66	0.50
2	$1-f_{worst}/f_{best}$	0.00	0.25	0.005	0.33	0.007	0.33
3	$1-f_{best}/f_{worst}$	0.00	0.50	0.50	0.67	0.67	0.67
4	$\lvert f_{avg} - f_{best}\rvert$	0.00	25.00	49.50	50.00	99.00	50.00
5	$\hat{\sigma}_f$	0.00	25.13	5.00	50.25	10.00	29.30
6	$\frac{1}{N}\sum_{i=1}^{N}\lvert f_i - f_{avg}\rvert$	0.00	25.00	0.99	50.00	1.98	25.25
7	$\sum_{i=1}^{N}(f_i - f_{avg})^2$	0.00	6.25E+04	2.48E+03	2.50E+05	9.90E+03	8.50E+04
8	$\frac{2}{N(N-1)}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\lvert f_i - f_j\rvert$	0.00	25.25	1.00	50.51	2.00	34.01
9	$-\sum_{m=1}^{N} p_m \log(p_m)$	0.00	0.69	0.06	0.69	0.06	4.59
10	$\frac{1}{\alpha-1}\left(1-\sum_{m=1}^{N} p_m^{\alpha}\right)$	0.00	0.96	0.70	0.96	0.70	68.34
11	$\frac{1}{1-\alpha}\log\left(\sum_{m=1}^{N} p_m^{\alpha}\right)$	0.00	0.69	0.54	0.69	0.54	4.59
12	$\frac{N^2}{2}\sum_{m=1}^{N} p_m(1-p_m)$	0.00	2500.00	99.00	2500.00	99.00	4949.00

In addition, all but the third phenotypic family violate the twinning requirement; their state evaluations are dependent on the number of individuals located at a given fitness position (Case 2 $\neq$ Case 3, and Case 4 $\neq$ Case 5). In fact, the third family is able to fulfill this requirement only because its extreme ratio takes advantage of the proposed fitness value distribution.

Finally, none of the fitness frequency families (9 to 12) is capable of adequately describing the monotonicity in distance requirement. This is because they all show identical phenotypic measurement, since they do not take into account the location of the intervals over the fitness distribution (section 1.5.1) in cases where the individuals are different distances apart (Cases 2 $\not<$ 4, or Cases 3 $\not<$ 5).

Table 3.4 summarizes the behavior of the phenotypic families over the diversity requirements identified.

Table 3.4 Summary of the diversity requirement fulfillment by the phenotypic formulations

FAMILY	DIVERSITY FORMULATION	REQUIREMENT : CASES :	#1 (6 > all others)	#2 (2>3 and 4>5)	#3 (4>2 and 5>3)
1	$1-f_{bc}/f_{ge}$		NO	NO	YES
2	$1-f_{bc}/f_{cave}$		NO	NO	YES
3	$1-f_{bc}/f_{aave}$		NO	(YES)	YES
4	$\lvert f_{cge}-f_{bac}\rvert$		NO	NO	YES
5	$\hat{\sigma}_f$		NO	NO	YES
6	$\frac{1}{N}\sum_{i=1}^{N}\lvert f_i-f_{avg}\rvert$		NO	NO	YES
7	$\sum_{i=1}^{N}(f_i-f_{avg})^2$		NO	NO	YES
8	$\frac{2}{N(N-1)}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\lvert f_i-f_j\rvert$		NO	NO	YES
9	$-\sum_{n=1}^{M}p_n\log(p_n)$		YES	NO	NO
10	$\frac{1}{\alpha-1}\left(1-\sum_{n=1}^{M}p_n^{\alpha}\right)$		YES	NO	NO
11	$\frac{1}{1-\alpha}\log_2\left(\sum_{n=1}^{M}p_n^{\alpha}\right)$		YES	NO	NO
12	$\frac{N^2}{2}\sum_{n=1}^{M}p_n(1-p_n)$		YES	NO	NO

3.5 New phenotypic formulation proposed

To control the EEB within the orthogonal framework, a reliable PCM is required. The previous section revealed that no phenotypic formulation available in the literature offers a perfect description of the scattering of the fitness distribution. Therefore, our aim in this section is to present a new formulation that meets the requirements listed in Table 3.2.

This new formulation is based on multiplication of the phenotypic value differences established between neighbors. Once the fitness distribution has been sorted, the computation can start from any side of the sorted distribution. The formulation ensures that the state of maximum phenotypic diversity occurs when all the individuals are uniformly spread out within the fitness range, which leads to the VMD case and fulfillment of the monotonicity in fitness varieties requirement.

To demonstrate, Figure 3.2 a) depicts the diversity level of a phenotypic distribution with three individuals located within a 10-unit fitness range. One individual is located at f_{best} (0), another at f_{worst} (10), and the third between the boundaries of this range. This example shows that the behavior of the proposal is generally good. The maximum diversity state appears when the third individual is located at 5, and the performance deteriorates as the third

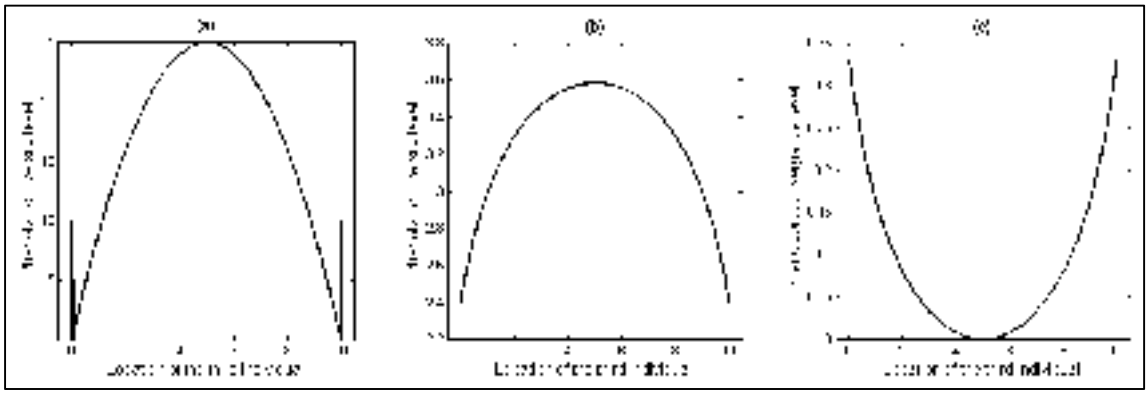


Figure 3.2 Phenotypic diversity level of the new proposal registered over a population of three individuals: a) Multiplicative formulation, b) Logarithmic formulation, c) Phenotypic convergence level of PCM_{13}

individual approaches the fitness range boundaries. This response degradation comes about as a result of the multiplication effect, and so it appears when the third individual is located closer than 1 unit from any other individual. Furthermore, the multiplication of the fitness difference between neighbors could rapidly lead to very high numbers, as the population and fitness range increase. The multiplication is therefore replaced by the addition of the logarithms of the neighbor differences. Moreover, the addition of 1 in the logarithmic operator automatically eliminates duplicate fitness values, which ensures that the twinning requirement is met. Equation 3.1 gives the descriptor formulation. The previous example is repeated in Figure 3.2 b) with this enhanced formulation.

$$\sum_{i=1}^{N-1} \ln\left(1+\left|f_i - f_{i+1}\right|\right) \tag{3.1}$$

Finally, equation 3.1 can be rewritten as a PCM formulation, as follows:

$$\boxed{PCM_{13} = 1 - \frac{\sum_{i=1}^{N-1} \ln\left(1+\left|f_i - f_{i+1}\right|\right)}{VMD}} \tag{3.2}$$

Figure 3.2 c) presents the behavior of PCM_{13} over the previous example. It confirms that the lowest convergence state (or highest diversity state) is achieved when the population is uniformly distributed (with the third individual located at 5). It can also be observed that the maximum achieved convergence level is 0.33, since at least two individuals are always differentiated by the maximum distance allowed from the fitness range.

3.5.1 Analysis of the new phenotypic formulation over the diversity requirements

The formulation can be evaluated by means of the validation framework introduced in the previous section, the results of which are provided in Table 3.5. As with the preceding descriptors, the new proposal in its non normalized version (equation 3.1) detects the

converged fitness distribution (Case 1 = 0). Moreover, Table 3.5 reveals that the new formulation conforms to the three diversity requirements: Monotonicity in fitness varieties is respected, as Case 6 presents the highest diversity level. Twinning is followed, since a different distribution of fitness values has no impact on the diversity level (Case 2 = Case 3, and Case 4 = Case 5). Finally, the monotonicity in distance requirement is met, as the distance between individuals is accurately taken into account (Case 2 < Case 4, and Case 3 < Case 5).

Table 3.5 Diversity level of the new proposal over the validation framework

FAMILY	DIVERSITY FORMULATION	PHENOTYPIC DISTRIBUTION					
		Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
[illegible]	$\sum^{[illegible]} \ln(1+\lvert [illegible] \rvert)$	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

3.6 Analysis of PCMs over specifically designed landscapes

Now that the new phenotypic formulation has been proved to perform in accordance with the diversity requirements, this section examines its behavior over the course of a search process. PCM_1 to PCM_{12} are also included in the investigation. However, this analysis requires that the phenotypic state be known quantitatively, or at least qualitatively, throughout the optimization process. This would become a serious issue if the search were based on an EA, since the sampled fitness distribution depends on the search path followed, which is a stochastic process. The result would be to hide the phenotypic distribution structure of a chosen benchmark. Furthermore, replications of the simulations, which are essential for validating the reliability of a PCM, would be useless.

In order to circumvent this problem, a generic benchmark is proposed to ensure uniform fitness distribution sampling, as well as control of the phenotypic states by the landscape definition and the search dynamic. Furthermore, with this benchmark, no genetic operator is involved in the evolution of the population. Instead, at each iteration, a new fitness

distribution is sampled over the landscape. Phenotypic convergence is simulated by reducing the sampling boundary as the process evolves. Consequently, the process begins in a full phenotypic diversity state (PCM = 0) and proceeds to a convergence state (PCM = 1) following a predefined schedule. As a result, the phenotypic distribution is known throughout the evolution process.

We propose three landscapes here; a linear landscape, a double-slope landscape, and a saw tooth landscape, as depicted in Figure 3.3. The analysis is conducted with a population size of 100, while an interval number of 100 is assigned and applied for PCM_9 to PCM_{12} inclusive. All the results are averaged over 50 repetitions.

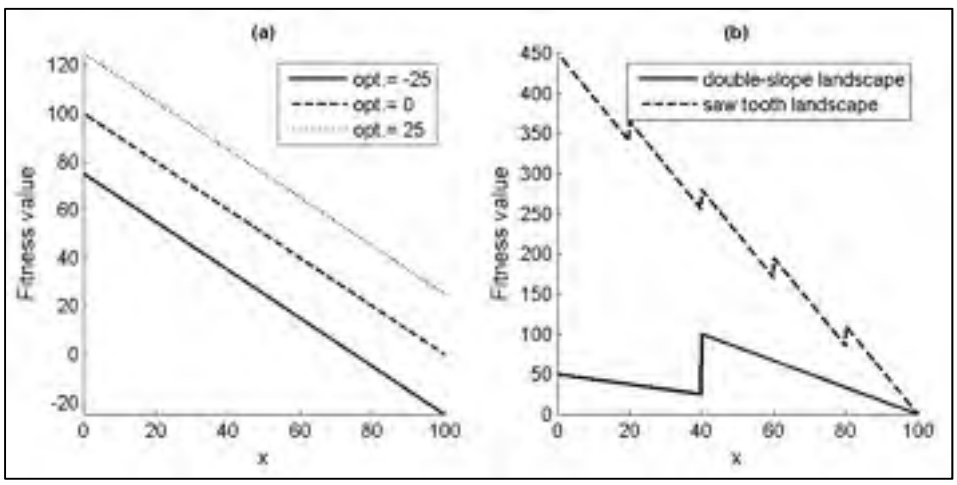


Figure 3.3 Fitness functions of the generic benchmark: a) Three translating cases of the linear landscape, b) Double-slope landscape and the saw tooth landscape

3.6.1 Linear function

3.6.1.1 Landscape definition

The first landscape includes a linear function (Figure 3.3 a)). A good PCM has to reflect the intended linear convergence pattern. The fitness function is given by:

$$f(x) = -x + b \tag{3.3}$$

The variable x denotes the genotypic position of the individuals, and ranges from 0 to 100. Since the study considers only the fitness of the individuals, a univariate genotype is sufficient. Three translations of equation 3.3 are considered: b equal to {75, 100, 125}. These cases allow the assessment of PCM coherence. Indeed, an accurate PCM should provide a constant evaluation, regardless of the fitness value sign. This is important, as the fitness range is often unknown for real-world problems.

The population is uniformly generated within the genotypic range. At each iteration, the convergence of the lower genotypic boundary is increased by 2% toward the optimum, and the process goes from PCM = 0 to PCM = 1 in 51 iterations. In addition, in order to ensure that f_{best} always represents the optimum value $f(x^*=100)$, an elite individual is inserted at the optimum position. This landscape simulates the dynamics of a search process over a unimodal landscape.

3.6.1.2 Behavioral results of the PCMs

As demonstrated by the characteristic response curves in Figure 3.4, PCM_1 to PCM_3, $PCM_{4.4}$,

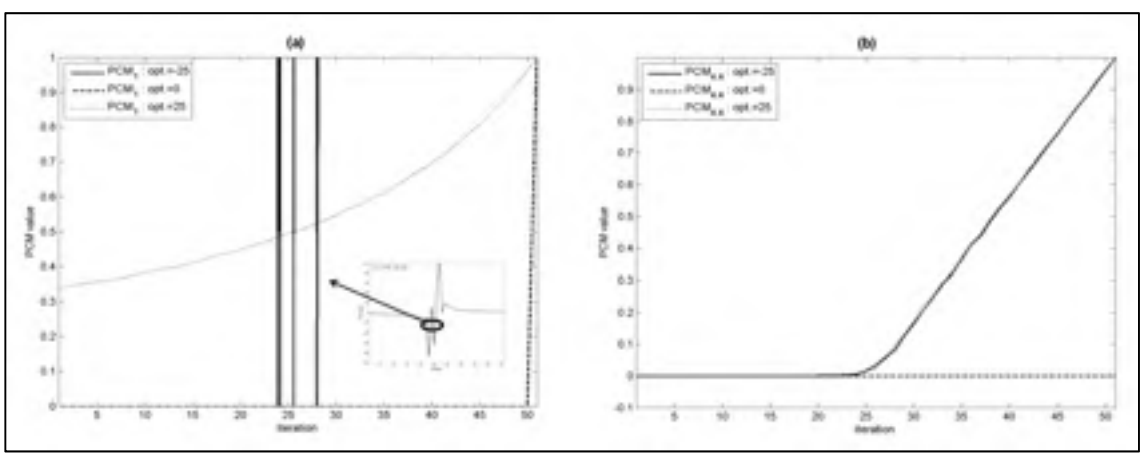


Figure 3.4 PCM behavior observed over the linear landscape: a) PCM_1, b) $PCM_{4.4}$

and $PCM_{5.3}$ appear to be unreliable. The first three PCMs fall outside the unitary range for the negative optimum case. However, $PCM_{4.1}$, $PCM_{4.2}$, $PCM_{5.1}$, and $PCM_{5.2}$ (Figure 3.5 a)) are unable to describe the convergence progression, as their value remains constant throughout the entire process. This behavior is generated by the numerator and the denominator decreasing at the same rate over the process.

The fitness frequency measures PCM_9 to PCM_{12} present a similar trend, and do not monitor phenotypic diversity well, as demonstrated in Figures 3.5 b)-3.6 a). Since the descriptors do not meet the monotonicity in distance requirement, their convergence values remain quite low for a significant part of the process. PCM_9 and PCM_{12} are not explicitly presented, as their behaviors are similar to those of PCM_{10} (α=1.1) and PCM_{10} (α=2.0) respectively.

Figure 3.6 b) presents the evolution of the state of convergence of PCM_{13}. The curves reveal good coherence and show a generally good trend. Nevertheless, the linear pattern is not perfectly represented, as the process does not start in a state of full diversity. This may be a result of the sampling error of the population. We investigate this in a section below, as part of a discussion on a sensitivity analysis procedure. The phenotypic state estimation error at the beginning of the process could also be linked to the random number generator (RNG) imprecision.

Finally, the remaining PCMs ($PCM_{4.3}$, $PCM_{4.5}$, $PCM_{5.4}$, and PCM_6 to PCM_8), although they are not included in a figure, provide excellent descriptions of the linear function convergence pattern.

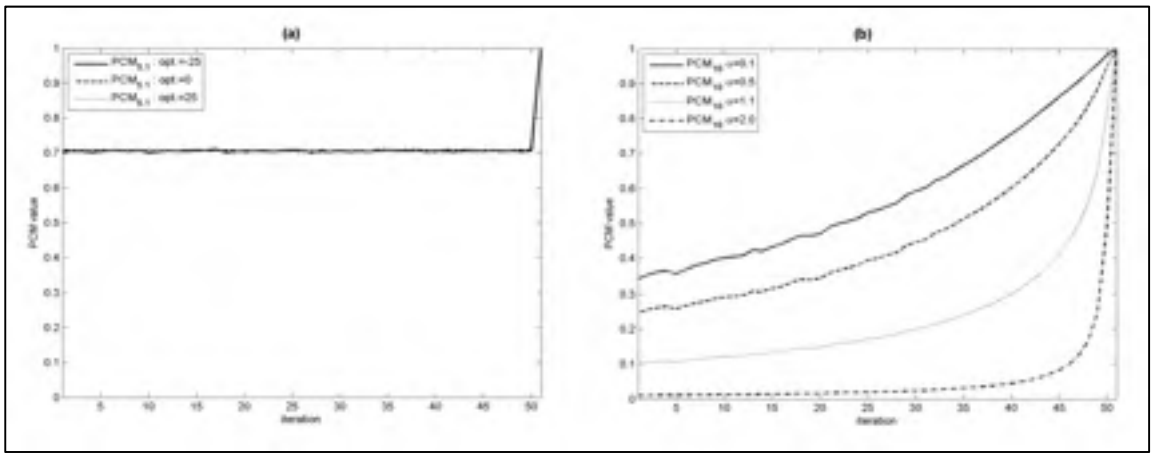


Figure 3.5 PCM behavior observed over the linear landscape: a) $PCM_{5.1}$, b) PCM_{10}

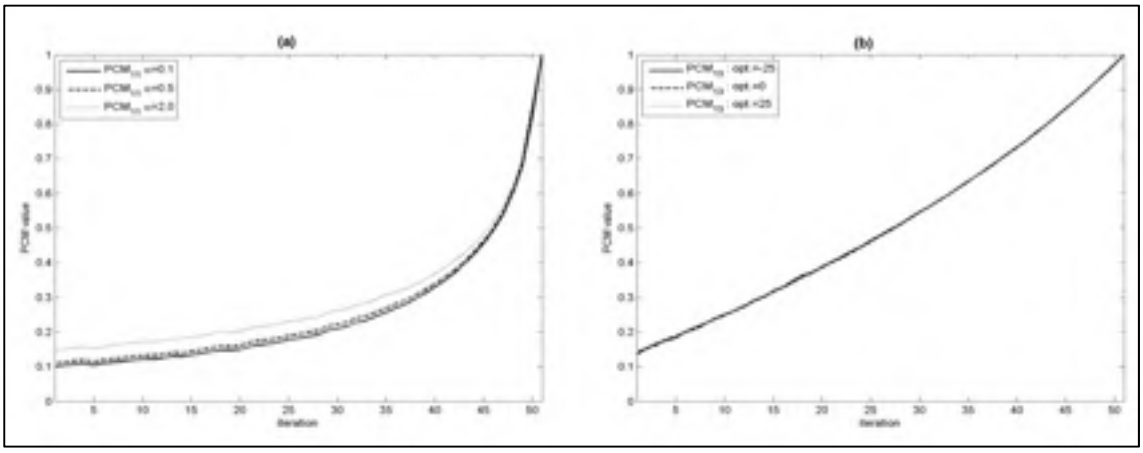


Figure 3.6 PCM behavior observed over the linear landscape: a) PCM_{11}, b) PCM_{13}

3.6.2 Double-slope landscape

3.6.2.1 Landscape definition

The second landscape (Figure 3.3 b)) is proposed in order to study the impact of an increasing fitness range. The fitness function is given by:

$$f(x) = \begin{cases} -25/40\, x + 50 & \text{if } x \leq 40 \\ -5/3\, x + 166.67 & \text{otherwise.} \end{cases} \tag{3.4}$$

Again, the genotypic position (x) ranges from 0 to 100. An elite individual, located at $f(x^*=100) = 0$, is kept in the population, and the remaining individuals are uniformly generated within the genotypic boundaries.

The following convergence schedule is adopted for the analysis: during the first 20 iterations, individuals (except the elite one) are located only on the first slope ($0 \leq x \leq 40$). Thereafter, they are located solely on the second slope ($40 < x \leq 100$) (31 remaining iterations). The lower genotypic bound is brought closer to the global optimum location by 2% of the total range at each iteration. Therefore, the jump to the second slope is implicitly controlled by the lower boundary, while the upper genotypic bound is constrained as previously defined.

Intuitively, reliable PCMs should reflect the two linear patterns. Locating the elite individual at the global optimum ensures that the first convergence pattern does not cover the total fitness range. So, the first pattern should start with a phenotypic convergence level slightly below 0.5, due to the fact that half the fitness range is covered (fitness $\in$ [25, 50]) with the presence of an elite individual at the global optimum. The same pattern should end with a phenotypic convergence at around 1, since there are only two fitness values in the population at iteration 20. The second pattern should start in a full phenotypic diversity state, as the fitness range grows and the population is distributed over the entire area. This pattern should end with a PCM value of 1, due to the full convergence of the population at the global optimum position ($x^*=100$).

The double-slope landscape emulates the dynamics of a search process, which might follow a tuned restart and move the population into a second region.

3.6.2.2 Behavioral results of the PCMs

Figures 3.7-3.8 draw the responses over the double-slope landscape of the PCMs that presented a good description of the linear landscape ($PCM_{4.3}$, $PCM_{4.5}$, $PCM_{5.4}$, PCM_6 to PCM_8, PCM_{13}).

The curves in Figure 3.7 a) reveal incorrect descriptions resulting from the NMDF normalization used by $PCM_{4.3}$, where the first iteration is assumed to offer the highest diversities. Moreover, the first convergence pattern ends far from the convergence state. $PCM_{4.5}$ also appears to be imprecise, as it does not demonstrate the two intended patterns. More fundamentally, the response of $PCM_{4.5}$ does not remain in the unitary range. Since $PCM_{4.5}$ is based on VMD normalization, the negative diversity estimations indicate that $\left|f_{avg} - f_{best}\right|$ (the PCM_4 family) violates the requirement of monotonicity in fitness varieties.

$PCM_{5.4}$ (Figure 3.7 b)) shows a relatively good convergence pattern. Nevertheless, when the fitness range is increased (iteration 21), $PCM_{5.4}$ generates negative values. Even though the error remains small, the underlying problem is similar to the one described for $PCM_{4.5}$.

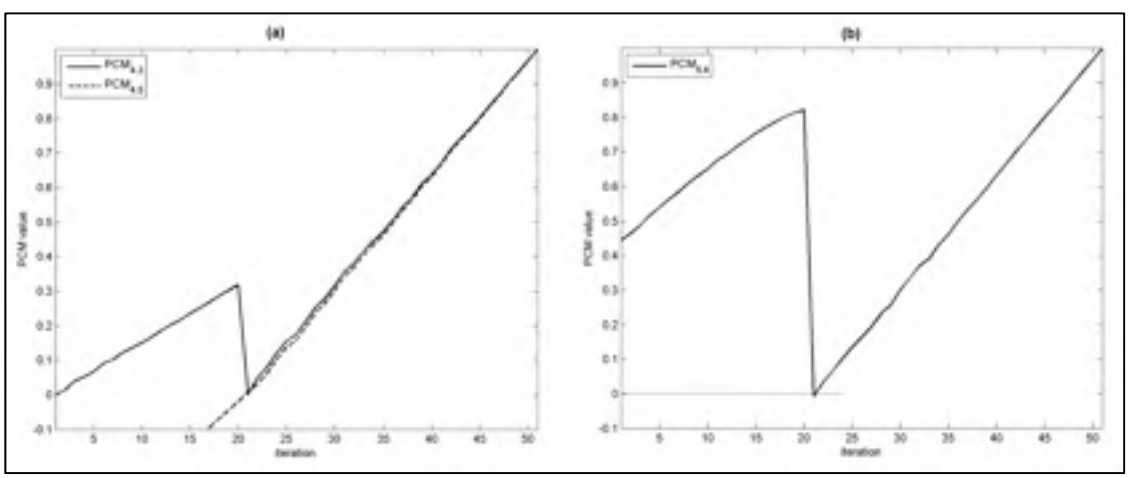


Figure 3.7 PCM behavior observed over the double-slope landscape: a) $PCM_{4.3}$ and $PCM_{4.5}$, b) $PCM_{5.4}$

Therefore, phenotypic formulations based on standard deviation (the PCM_5 family) are not recommended, since they contravene the monotonicity in fitness varieties requirement. The same conclusion applies to PCM_6 to PCM_8 (Figure 3.8 a)).

PCM_{13} shows a good trend over the two patterns (Figure 3.8 b)). Nonetheless, the descriptor cannot perfectly predict a null convergence state when the fitness range is increased (iteration 21). Again, the deviation can be attributed to the population sampling error and RNG inaccuracy.

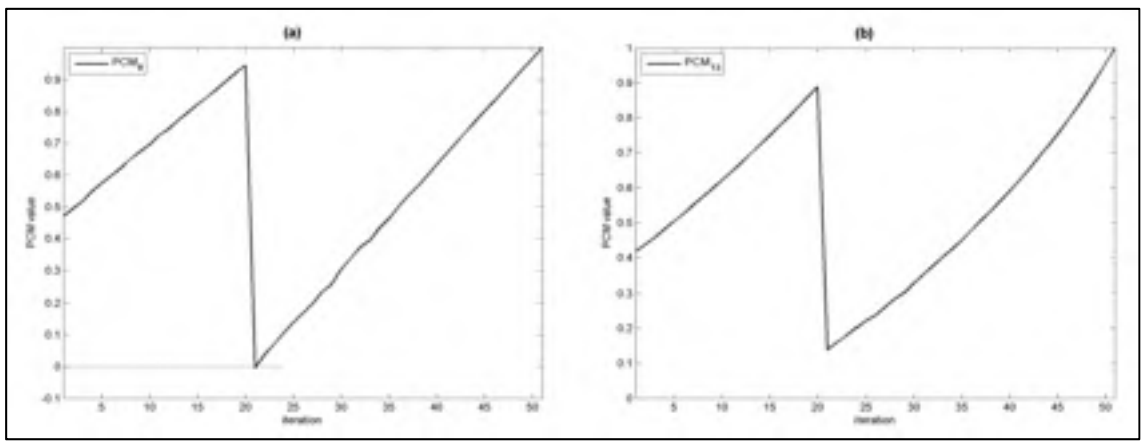


Figure 3.8 PCM behavior observed over the double-slope landscape: a) PCM_8, b) PCM_{13}

3.6.3 Saw tooth landscape

3.6.3.1 Landscape definition

The third landscape (Figure 3.3 b)) reproduces a multimodal fitness distribution. The fitness function is given by:

$$f(x)=\begin{cases} -11/2\,x+450 & \text{if } x\le 20 \\ -11/2\,x+475 & \text{elseif } x\le 40 \\ -11/2\,x+500 & \text{elseif } x\le 60 \\ -11/2\,x+525 & \text{elseif } x\le 80 \\ -11/2\,x+550 & \text{otherwise.} \end{cases} \tag{3.5}$$

Once again, an elite individual is located at the global optimum, and the genotypic position (x) ranges from 0 to 100.

The population is divided into five equivalent groups, and each group is located over a different tooth. This means that the genotypic boundaries are relative to a group. Moreover, the process is divided into five phases, each with 11 iterations. The dynamics of the landscape demands that every group converge toward its local tooth optimum. After this first convergence, the phase is considered completed, and the groups jump to the next tooth, where the process is repeated. After completion of the last tooth, the groups remain frozen at the global optimum position ($f_{best} = f(x^*=100) = 0$) until the end of the process. This structure is presented in Table 3.6.

To summarize, at the end of phase 1 (iteration 11), the fitness distribution is evaluated over five modes. At the end of phase 2 (iteration 22), the fitness distribution is evaluated over four

Table 3.6 Convergence schedules for Group 1 (G1) to Group 5 (G5) over the saw tooth landscape

		SAW TOOTH NUMBER					
PHASE	ITERATIONS	1	2	3	4	5	Frozen
		$x \in [0\ 20]$	$x \in [20\ 40]$	$x \in [40\ 60]$	$x \in [60\ 80]$	$x \in [80\ 100]$	$x = 100$
Phase 1	1 to 11	G1	G2	G3	G4	G5	
Phase 2	12 to 22		G1	G2	G3	G4	G5
Phase 3	23 to 33			G1	G2	G3	G4 to G5
Phase 4	34 to 44				G1	G2	G3 to G5
Phase 5	45 to 55					G1	G2 to G5

modes, and so on, up to the end of phase 5 (iteration 55), where all the individuals are located at the global optimum. The evolution of the process proceeds from a 10% increase in the lower genotypic bound at each iteration, while the upper bound corresponds to the tooth's local optimum. The groups are generated following a uniform distribution between their boundaries.

Evaluation of the PCMs should present the following: the first convergence phase should start at around 0 (the complete fitness range is covered); the second phase should start at around 0.2 (the fitness distribution covers 4/5 of the fitness range), and so on. All phases should end near convergence, with a value increasing toward 1 as the process evolves.

The search process simulates an algorithm that clusters its resources or individuals over different regions of the landscape.

3.6.3.2 Behavioral results of the PCMs

None of the PCMs introduced in Table 3.1 was able to adequately describe the intended phenotypic convergence pattern of the saw tooth landscape. However, some characteristic behaviors of these PCMs are depicted in Figure 3.9. PCM_1, PCM_3, and $PCM_{4.4}$ show a constant state of full diversity throughout the process, due to the influence of f_{best}, which is fixed here at 0. PCM_2 presents a completely misleading trend.

Regarding the PCM_4 family, $PCM_{4.1}$ and $PCM_{4.2}$ show the same behavior, and their formulations (similar to those of PCM_2) are not able to describe the convergence progression within each phase. In contrast, $PCM_{4.3}$ and $PCM_{4.5}$ show a convergence progression, but are unable to describe the intended convergence peak at the end of each phase. $PCM_{4.5}$ starts below 0 (-0.004), as it does for the double-slope landscape.

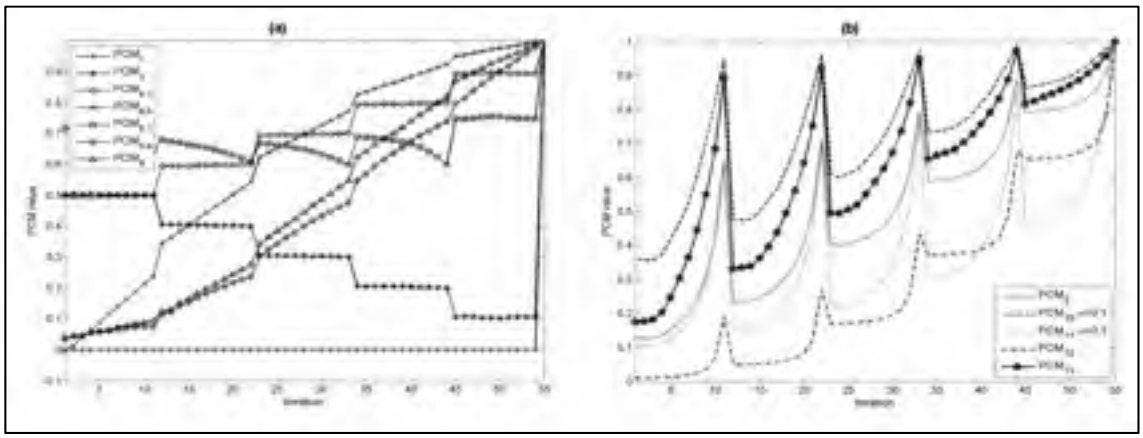


Figure 3.9 PCM behavior observed over the saw tooth landscape: a) PCM_1 to PCM_8, b) PCM_9 to PCM_{13}

The PCM_5 family provides very diversified patterns. This family appears to be incapable of characterizing the convergence of each phase, since no converged peak is observable over the five phases. A similar conclusion can be drawn from the results of PCM_6 to PCM_8.

The PCM_9 to PCM_{12} estimations show clear evidence of the convergence peaks. As demonstrated in Figure 3.9 b), PCM_{10} with $\alpha = 0.1$ is the best option for estimating diversity at the end of each phase, whereas PCM_9 is best for estimating the diversity at the beginning of each phase.

Finally, PCM_{13} accurately represents the convergence pattern over the five phases (Figure 3.9 b)). The converged peaks of each phase are well established. However, as for the two preceding landscapes, the convergence state at the beginning of the first phases is slightly overestimated.

In summary, the proposed landscapes ensure a detailed description of the PCM_1 to PCM_{12} response. PCM13 showed the best overall description of the phenotypic distributions.

3.7 Assessment of desirable PCM qualities

Following the very good performance demonstrated by PCM_{13} over the validation framework and the proposed landscapes, it is now relevant to establish the quality criteria for a reliable PCM. These qualities are applicable to any PCM formulation.

3.7.1 Definition of desirable PCM qualities

The three following characteristics are proposed as desirable qualities:

1. *Reliability*: A PCM should be reliable over similarly scattered phenotypic distributions;
2. *Sensitivity*: A PCM should be as insensitive as possible to the simulation parameters;
3. *Outlier influence*: A PCM should adequately consider the impact of phenotypic outliers.

The performance of PCM_{13} is evaluated in the following section with respect to the desirable qualities. The assessment makes use of the landscapes introduced in the previous section.

3.7.2 PCM_{13} reliability analysis

The reliability of PCM_{13} is studied through a dispersion analysis, based on 50 repetitions per iteration. In a preliminary step, normality tests were carried out in this study using the Kolmogorov-Smirnov test (0.05 significance level), which indicated that, for each iteration, the 50-repetition sample does not follow a normal distribution. Therefore, to picture PCM reliability, it is preferable to compute the dispersion for 96% of the repetition data. For the sake of clarity, the dispersion values are averaged over the whole process. Table 3.7 illustrates the stability analysis for four common EA populations size ($N \in \{50, 100, 300, 500\}$).

The results clearly indicate that PCM_{13} gives a stable phenotypic state description. On average, for a relatively small population, 96% of all repetition data are stacked with a range

smaller than 0.06. Moreover, the analysis reveals that the stability rapidly increases as the population increases, which validates the influence of the sampling error associated with the population size mentioned in section 3.6.

Table 3.7 Average range among 96% of the repetition data for PCM_{13} over the three landscape designs

LANDSCAPE	POPULATION SIZE (N)			
	50	100	300	500
Linear	0.062	0.035	0.024	0.008
Double-slope	0.050	0.028	0.0[illegible]	0.006
Saw tooth	0.038	0.018	0.012	0.019

3.7.3 PCM$_{13}$ sensitivity analysis

The second experiment constitutes an analysis of the sensitivity of a PCM to the simulation parameters. Considering the PCM_{13} formulation, and since the fitness distribution is univariate, the population size is the only parameter involved. The high reliability of PCM_{13} shown during the stability analysis suggests that the mean curves of the 50 repetitions are representative of the convergence process. Figure 3.10 presents the mean curve results obtained for the previous population size samples ($N \in \{50, 100, 300, 500\}$) over the three landscapes described in section 3.6.

The trends observed over the three different landscapes are very similar. The population size parameter shows only a slight influence, no matter what the landscape. In all cases, the maximum difference appears between population sizes 50 and 500. The maximum discrepancy values are 0.089, 0.095, and 0.019 for the linear, double-slope, and saw tooth landscapes respectively. More importantly, this analysis verifies that the linearity and coverage of the phenotypic responses increase as the population increases, which validates the explanation given in section 3.6 of the behavior of PCM_{13}.

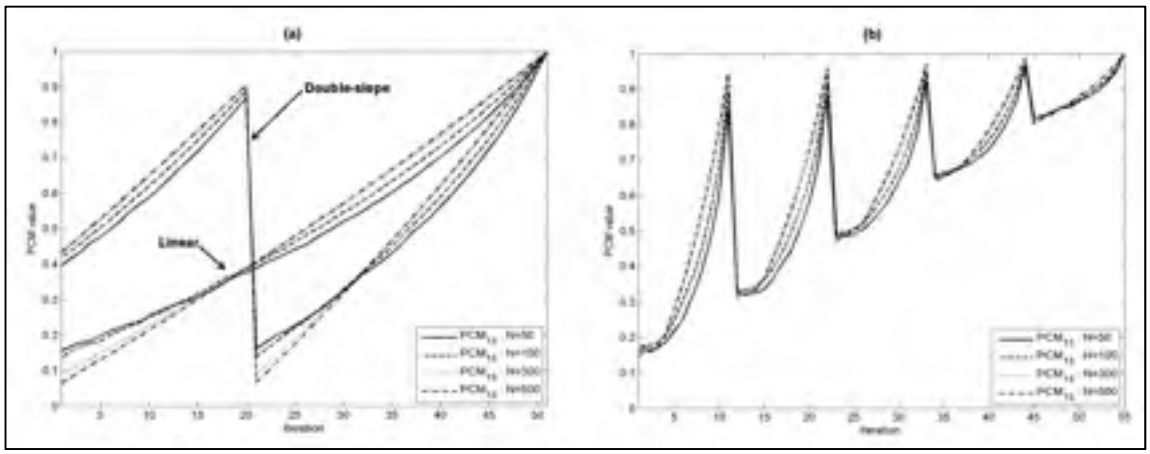


Figure 3.10 Sensitivity of PCM_{13} with respect to the population size (N) observed over: a) Linear and double-slope landscapes, b) Saw tooth landscape

3.7.4 PCM_{13} analysis with outliers

The final experiment assesses the performance of PCM_{13} in the presence of outliers within the fitness distribution. In reality, the impact of outliers on phenotypic distribution remains unclear. While their presence should normally increase the diversity, they could also increase the fitness range, leading to an over-converged state of the remaining population, as compared to a population without outliers. This analysis will therefore help shed light on the effect of outliers.

To conduct the analysis, the landscapes introduced in section 3.6 are adjusted as follows: from the 10th iteration, a given percentage of the population is randomly generated between the lower genotypic bounds of the 1st and 10th iterations. For the saw tooth landscape, this group is restricted to the first tooth. So, as the process evolves, these individuals act as outliers. Herein, the percentage of outliers is set to 1%, 2%, 5%, and 10%.

Figure 3.11 presents the impact of outliers over the linear, double-slope, and saw tooth landscapes. The simulations are repeated 50 times, with the population size kept at 100. PCM_{13} adequately considers the presence of outliers, since the diversity level increases (or

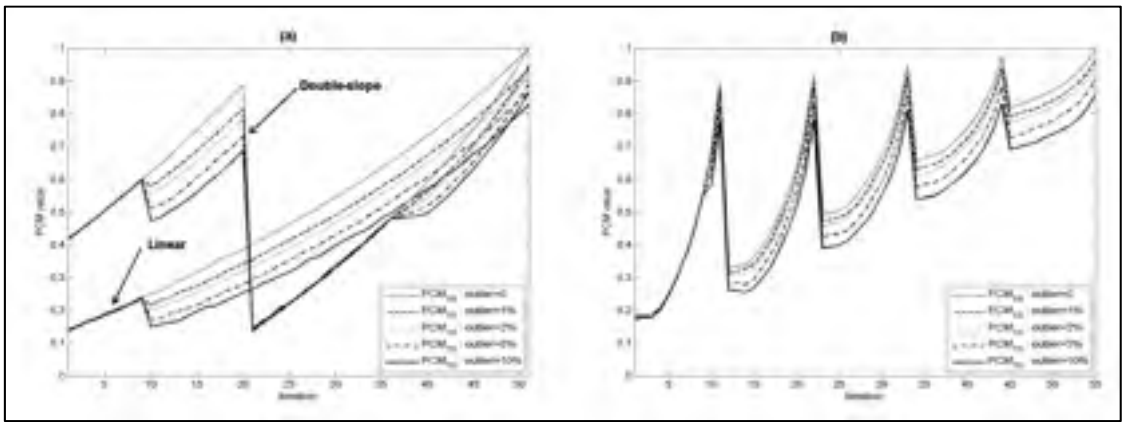


Figure 3.11 Impact of outliers on PCM_{13} observed over the: a) Linear and double-slope landscape, b) Saw tooth landscape

the convergence level decreases) as the number of outliers increases. With the fitness scaling factor set at 1, there is no sign of convergence value overestimation.

In real-world problems, however, outliers may modify the fitness range. This condition is simulated through a scaling factor applied to the fitness of the outliers. Figures 3.12 to 3.14 present the results with 1% and 10% of outliers. Scaling factors of 10 and 100 are compared to both the unit scaling factor and the cases without outliers. Even though the results are obtained for a population of 100, the trend is similar with other population sizes.

The simulations explicitly demonstrate that the enlargement of the fitness range by outliers leads to an overestimation of the convergence levels. More interestingly, increasing the number of outliers that disrupt the fitness range decreases the convergence level overestimation as they increasingly become part of the population. Finally, for all the landscapes studied, even with a scaling factor of 100, the overestimation of the phenotypic convergence level does not lead to an early full convergence state. In other words, even if outliers disturb the convergence pattern, the remaining fitness values continue to play an active role in the computed phenotypic state.

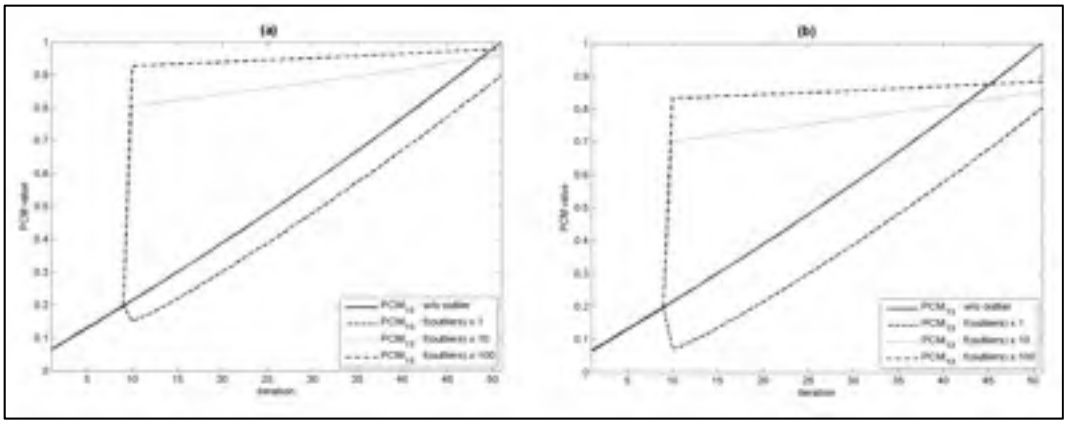


Figure 3.12 Impact on PCM_{13} of outliers that are far away, observed over the linear landscape: a) 1% outliers, b) 10% outliers

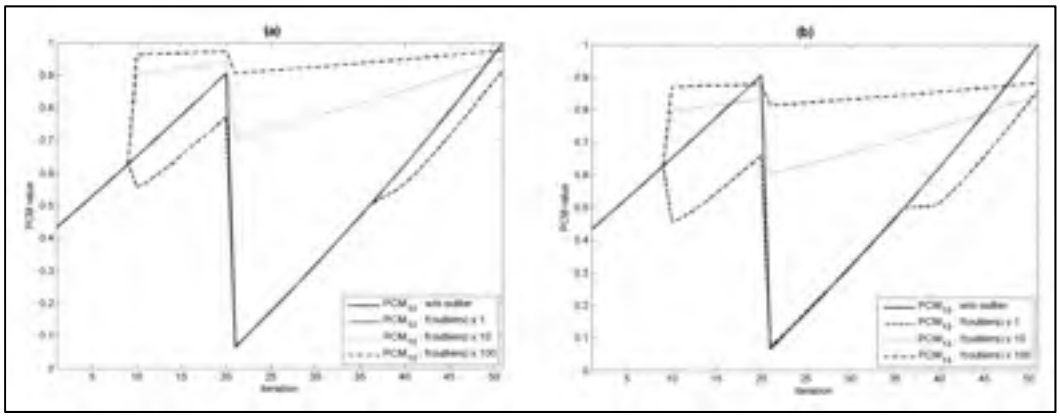


Figure 3.13 Impact on PCM_{13} of outliers that are far away, observed over the double-slope landscape: a) 1% outliers, b) 10% outliers

These observations suggest thus that PCM_{13} is capable of properly managing the presence of outliers within the fitness distribution.

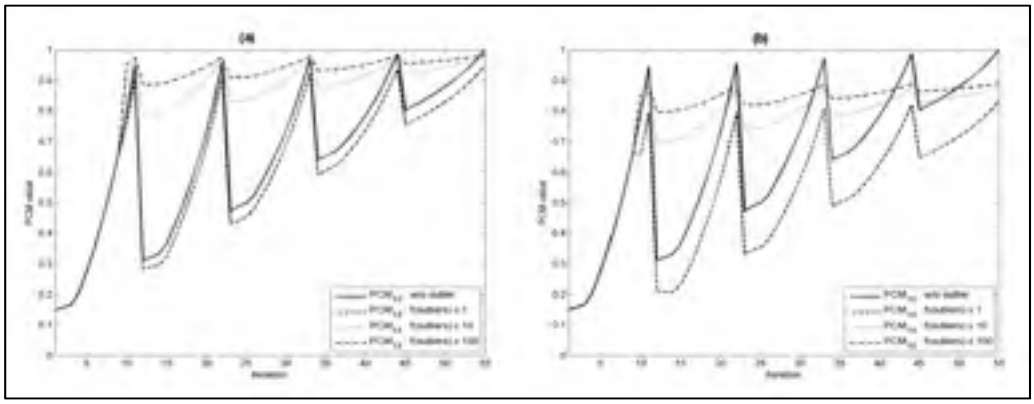


Figure 3.14 Impact on PCM_{13} of outliers that are far away, observed over the saw tooth landscape: a) 1% outliers, b) 10% outliers

3.8 Application of the EEB diagnostic tool

A good way to increase the efficiency of an optimizer over a given problem is to adjust its internal parameters (Eiben, Hinterding, Michalewicz, 1999). Parameter setting can be considered in the broadest sense of the term, so that the number of populations involved, the type of evolution model, the diversity promoting features, and the restart strategies are all viewed as adjustable parameters. Since PCM_{13} reliability for phenotypic convergence description has been demonstrated, this section proposes an efficient diagnostic tool developed based on EEB information to help evaluate the impact of any particular parameter setting procedure. The information acquired from PCM_{13} is completed along the exploration axis of the EEB with D^N_{LN} described by equation 3.6. D^N_{LN} represents a generalized multivariate genotypic descriptor based on PCM_{13}. This measure acts on individual genotypic materials. The difference between individuals is defined by the minimum distance with respect to their neighbors. In this formulation, $x_{i,k}$ and $x_{j,k}$ stand for the value of gene k ($k \in \{1,\ldots,n\}$) of the individuals i and j respectively.

$$D^N_{LN} = \frac{\sum_{i=1}^{N-1} \ln\left(1 + \min_{j\in[i+1,N]} \frac{1}{n}\sqrt{\sum_{k=1}^{n}\left(x_{i,k} - x_{j,k}\right)^2}\right)}{NMDF} \tag{3.6}$$

To illustrate and demonstrate the efficacy of the EEB diagnostic tool, the following experiments integrate and examine various selection plans, genetic operators, replacement plans, and population sizes. The impact of the evaluation of these parameters is measured by means of a real-coded steady-state genetic algorithm (SSGA). SSGA, which allows smooth transitions between generations, can be summarized as follows: two offspring are created at each generation; two individuals are removed from the population, following a selected replacement plan, to make room for the new individuals; thereafter, the best individuals from this temporary pool are inserted back into the population.

Five selection plans are considered here: 1) random selection of the parents; 2-4) a tournament scheme with 2, 5, and 10 competitors; and 5) the recently proposed FUSS approach (Hutter and Legg, 2006), while 4 genetic operators are integrated: 1) parent-centric crossover (PCX) (Deb, Joshi, and Anand, 2002; Deb, Anand, and Joshi, 2002); 2) unimodal normal distribution crossover (UNDX) (Ono and Kobayashi, 1997; Kita, Ono, and Kobayashi, 1998), which requires three parents; 3) uniform crossover (UX) (Sywerda, 1989); and 4) the blended crossover (BLX-α) (Eshelman and Schaffer, 1992), where only two parents are involved. Five values of α (0.1, 0.3, .0.5, 0.7, and 0.9) are considered within the BLX operator, leading to a comparison of eight genetic operators. Note that no mutation is considered in the search process, as all these crossover operators (except UX) have the ability to incorporate new genetic material into the population. Two methods are compared for the replacement plan: 1) randomly removing individuals, and 2) removing the worst individuals from the population. Finally, four common population sizes are analyzed: 50, 100, 300, and 500. The comparisons involve a default setting with random selection and replacement, a PCX crossover, and a population size of 300. In other words, the impact of each choice is evaluated, one at a time.

The following figures (Figures 3.15 to 3.18) present the results of our comparative studies. Continuous curves correspond to PCM information, and dashed curves identify GDM information. The curves present the median run of 25 repetitions. Since each run has a unique convergence history, averaging is difficult. Nonetheless, the curves are completed by a

shaded area indicating the range of values obtained throughout the repetitions. Therefore, narrow shaded areas indicate that the median run is representative. For the sake of brevity, only the CEC'05 benchmark 10-D F2 and 10-D F10 functions from (Suganthan *et al.*, 2005) are illustrated in Figures 3.15 to 3.17, while the genetic operator study (Figure 3.18) utilizes 10-D F2 and 10-D F21 (see appendix I for an overview of these functions). This latter choice provides a better demonstration of the relevance of the orthogonal EEB framework (section 3.2). In fact, Figure 3.18 b) indicates that exploitation (PCM) and exploration (GDM) are not complementary measures, but are complementary concepts portraying the EEB history. This observation is corroborated by the difference between the shaded area patterns of the PCM and the GDM.

In order to control the simulation duration, and since the purpose of the experiments is to monitor the impact of the EA parameters over the EEB (but not performance, in terms of end of solution quality), termination of the process was based on the CEC'05 criterion (100,000 evaluation cutoff) and a threshold applied over the PCM value (>0.99). Consequently, none of the simulated configurations found the global optimum. For instance, Figure 3.18 a) clearly demonstrates that, in some cases, median runs ended due to phenotypic convergence, even though genotypic convergence had not been reached. This condition undoubtedly confirms that the PCM and the GDM have their own role to play. More importantly, it underscores the fact that premature convergence must be evaluated through the GDM instead of the PCM, as identical fitness values came from different locations.

Our experimental results support the conviction of many in the community that the impact of parameter choice is critically important. Regarding selection plans (Figure 3.15), FUSS is shown to be the best option for delaying convergence, even better than the random search. In fact, FUSS was designed to avoid convergence (Hutter and Legg, 2006). However, in the presence of other evolutionary mechanisms, specifically the update plan, which promotes the best individuals, convergence is the inevitable result for any selection scheme. In contrast, compared to the previous schemes, the rate of convergence is higher for tournament

selection. Obviously, this is accentuated as the tournament size increases (Goldberg and Deb 1990).

Figure 3.16 shows the impact of the replacement plan when tested with FUSS selection. As expected, replacing the worst individuals increases the convergence speed. Considering other simulations not included in this chapter, this conclusion may be extended to all other selection methods. As a matter of fact, FUSS appears to be the most reactive selection method, while random selection appears to be the least responsive.

Figure 3.17 describes the impact of population size on the EEB. It turns out, as expected, that increasing the population helps maintain both genotypic and phenotypic diversity.

Crossover types are examined in Figure 3.18. It is observed that over F2, PCX and UNDX converge the fastest, followed by BLX-0.1, UX, and BLX-0.3 to BLX-0.9. The trend is similar for F21. However, no marked difference appears among PCX, UNDX, UX, and BLX-0.1 to BLX-0.3.

Finally, the search paths look very similar, regardless of the problem considered or the parameters selected. This may be surprising at first glance. However, since no mechanism that explicitly promotes diversity is incorporated, all the processes lead to similar search paths directed toward the best individual.

Even though they are constrained to a limited number of problems, the experiments presented here provide valuable insights into the true behavior of particular parameter choices. Globally, the results suggest that monitoring the EEB could serve as a powerful tool for characterizing EA differences and parameter influence, and may ultimately help in the design of better hybrid or improvement strategies.

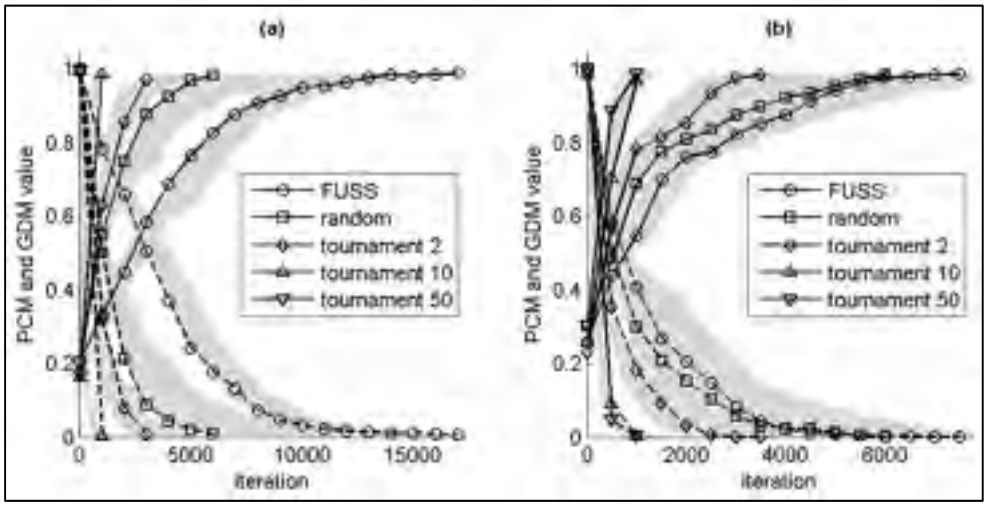


Figure 3.15 Impact of various selection plans over the EEB: a) 10D-F2, b) 10D-F10

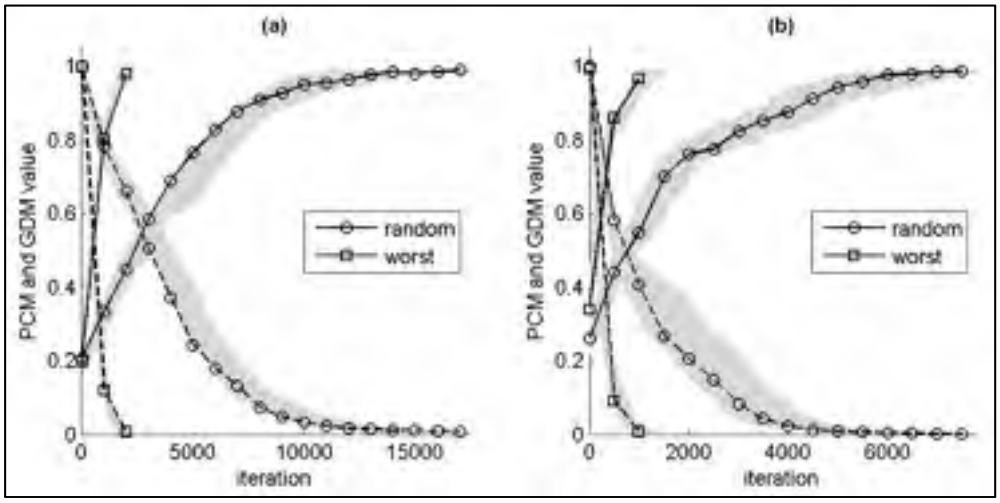


Figure 3.16 Impact of various replacement plans combined with FUSS selection over the EEB: a) 10D-F2, b) 10D-F10

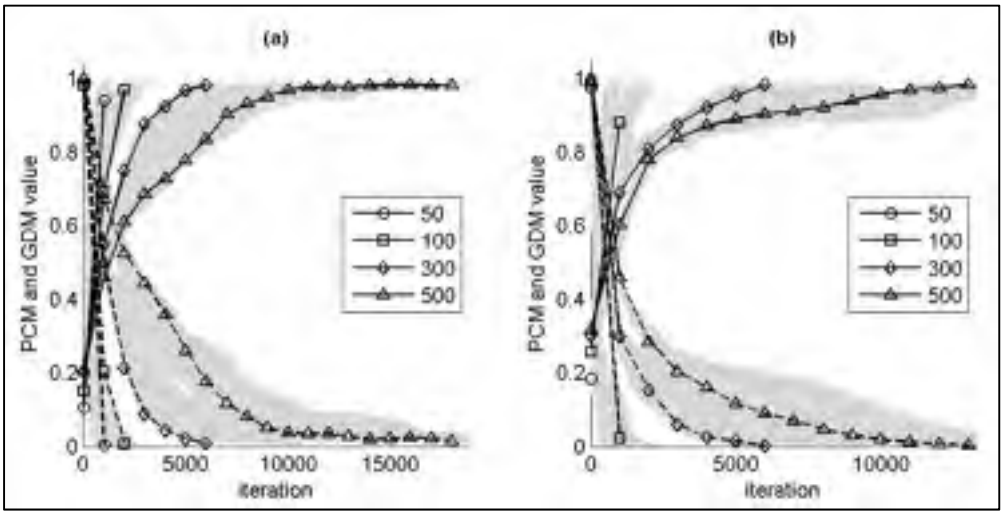


Figure 3.17 Impact of various population sizes over the EEB: a) 10D-F2, b) 10D-F10

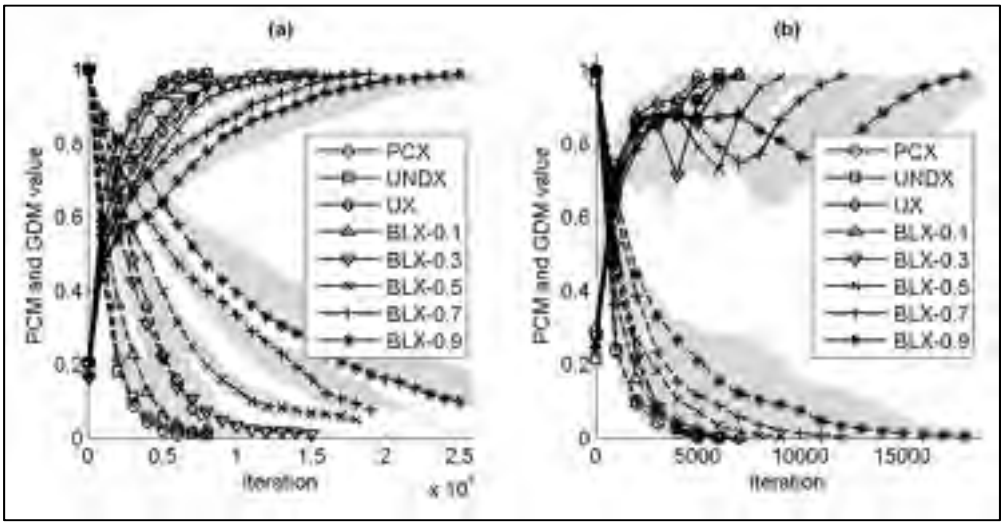


Figure 3.18 Impact of various crossover types over the EEB: a) 10D-F2, b) 10D-F21

3.9 Conclusion

Designing an efficient optimizer for a given problem is an issue that practitioners commonly encounter. It is quite a difficult task, as no single parameter setting procedure can drive the

search path toward its optimal course over any landscape. This issue has led us to develop a diagnostic tool designed to help identify the best optimizer options for the problem at hand. This tool records the EEB of the optimizer throughout the search process. As the EEB is responsible for the way resources are committed, monitoring it is a basic requirement for achieving efficiency for any population-based search method.

This chapter pursued two main objectives. The first objective was to investigate the conceptualization of the EEB framework. Our results show that considering exploration and exploitation as two orthogonal axes offers an effective description of EEB. Subsequently, we identified the GDM as the best description of exploration, and the PCM as an accurate portrayal of exploitation.

The second objective was to assess the performance of the diversity measures. Numerous formulations have been proposed in the literature over the years for doing so. However, to the best of the authors' knowledge, performance assessments of the various diversity formulations have only been conducted for GDM, although this evaluation is an important one. Consequently, we conducted a complete review of the PCM formulations, and compared nearly all the published PCMs over a validation framework involving six test cases with a controlled fitness distribution. With this simple framework, the underlying behavior of phenotypic formulations can be represented based on three requirements that we propose: monotonicity in fitness varieties, twinning, and monotonicity in distance. We proved that these requirements are sufficient for identifying phenotypic formulation weaknesses.

In summary, all the distance-based formulations fail to meet the monotonicity in fitness varieties and twinning requirements, and the fitness frequency formulations fail to meet to the twinning and monotonicity in distance requirements.

To improve the existing descriptive capacities of the formulations, we developed a new formulation (PCM_{13}), based on the neighbor fitness difference. Validation of PCM_{13} proves that it complies with all three requirements. In addition, we compared the phenotypic

formulations over three specifically designed fitness landscapes. The same landscapes also served as a platform for assessing the desirable qualities of PCMs. In fact, PCM_{13} proved to be reliable over similarly scattered fitness distributions, and showed slight sensitivity to population size. However, the observed sensitivity level remained irrelevant for proper convergence evaluation. Furthermore, the influence of outliers was investigated, the results suggesting that PCM_{13} reliably takes into account the influence of outliers, even when they greatly disturb the fitness range.

The new phenotypic formulation, combined with a genotypic formulation based on a generalized version of PCM_{13}, is therefore proposed as the foundation for an EEB diagnostic tool. Its usefulness has been shown by comparing behavior of various genetic operators and parameters over a real-coded SSGA. With this EEB diagnostic tool, it is now possible to compare the underlying mechanisms of various maintaining/promoting diversity approaches, and to better understand them (Das *et al.*, 2011). Finally, the next step would be to leverage the EEB diagnostic tool to develop an EEB management tool, to enable the search process to adapt its own evolutionary path as required, based on the PCM and GDM knowledge gathered.

Before embarking in that direction, we conduct an in-depth study of the new GDM proposal in the next chapter.

CHAPTER 4

EVALUATION OF THE GENERALIZED PHENOTYPIC FORMULATION AS A GENOTYPIC DIVERSITY MEASURE

4.1 Introduction

Chapters 1 and 2 illustrate the need for an improved GDM formulation in real-coded representations, since none of the current formulations acts as a coverage space indicator. In chapter 3, the focus turns to the definition of a phenotypic convergence measure. The outcome of this investigation is the development of a new formulation, represented by equation 3.2 (PCM_{13}), which is shown to be the only one capable of respecting all the defined diversity requirements (Table 3.2). Encouraged by the behavior of PCM_{13} and the fact that no satisfactory genotypic formulation has yet been identified, a multivariate proposal (D_{LN}), based on the underlying idea of PCM_{13}, is introduced in section 3.8. This was required in order to take advantage of the EEB as a search process diagnostic tool. For this application, D_{LN} was assumed to act as a true population diversity descriptor.

The objective of this chapter is therefore to validate D_{LN} by evaluating it on the genotypic framework defined in chapters 1 and 2.

The chapter is organized as follows: the next section provides the details about the multivariate generalization proposal of PCM_{13}; section 4.3 validates the behavior of this new GDM over the validation framework designed in chapter 2 and the benchmark problems defined in chapter 1; section 4.4 demonstrates its performance over a recognized benchmark; and section 4.5 presents our concluding discussion.

4.2 Generalization of PCM_{13} as multivariate diversity measure

The proposed diversity measure development is based on the following two step operation: 1) sort the individuals, and 2) multiply the distance difference between neighbors. This

approach leads to an evaluation of maximum diversity when the individuals are equally distributed. However, in a multivariate framework, the sorting step resulting from the neighborhood concept is a problem. Sorting individuals in a univariate framework directly leads to the notion of neighborhood. However, as indicated in chapter 2 for the Shannon entropy (GF_S), handling the genes independently and averaging their contributions to assess genotypic diversity is not appropriate. A generalization of the neighbor treatment for the multivariate context is therefore needed. We proposed considering the radius of a hypersphere around each individual, the hyperspace being bounded by its nearest neighbor. This will be referred to here as the characteristic distance, or the contribution radius, of each individual.

Two other aspects require further thought. First, the multiplication operator is not appropriate for neighbors closer than one unit, since this would parasite the diversity evaluation (see Figure 3.2 a)). As a consequence, duplicate individuals would overwrite the true diversity estimation with a complete convergence status. Furthermore, multiplication of the contribution radius could rapidly lead to very high numbers. The final value would thus depend greatly on the population size. To control this problem, we replaced the multiplication of the differences with the addition of the logarithm of the neighbor distance differences. Therefore, adding 1 to the distance difference inside the logarithm solves the consideration of duplicate individuals automatically.

The proposed multivariate diversity formulation (D_{LN}) is described in equation 4.1, while the normalized version (D^N_{LN}) is given in equation 4.2. The landscape dimensionality (n) is inserted into equation 4.1 to enhance measurement insensitivity to this parameter. The resulting improvement is illustrated in the next section. Since the population of the first generation of an evolutionary algorithm (EA) results from a random generation process (approximating a uniform distribution over the landscape), the normalization is based on NMDF (Normalization with Maximum Diversity so Far). Indeed, for a population size much larger than the dimensionality of the landscape, the first generation is assumed to provide a good estimation of the maximum diversity state. Nevertheless, the NMDF value is

continuously updated during the evolution process. Thus, the normalization value remains representative for all the EA run.

$$D_{LN} = \sum_{i=1}^{N-1} \ln\left(1 + \min_{j\in[i+1,N]} \frac{1}{n}\sqrt{\sum_{k=1}^{n}\left(x_{i,k} - x_{j,k}\right)^2}\right) \tag{4.1}$$

$$\boxed{D_{LN}^{N} = {D_{LN}}/{NMDF}} \tag{4.2}$$

As indicated by the minimum operator in equation 4.1, only the lower triangle of the pairwise matrix needs to be computed. Moreover, the consideration of only the lower triangle provides a tremendous advantage in the treatment of duplicate individuals. For example, the pairwise matrix (j, i) below represents a four individual population, where three are duplicated and the remaining individual is located at a radius of 10 units from the duplicate individuals.

$$\begin{bmatrix} null & null & null & null \\ 0 & null & null & null \\ 0 & 0 & null & null \\ 10 & 10 & \mathbf{10} & null \end{bmatrix}$$

Considering the minimum radius of each individual, only the third individual (i = 3) presents a diversity value higher than zero (between the duplicate individuals and the fourth individual (j = 4)). That is expected to not overestimate the contribution of duplicate individuals (2nd requirement of Table 2.3), since D_{LN} = ln(1+10). In contrast, if only the three duplicate individuals were considered (dashed box in the matrix), equation 4.1 would indicate a full convergence status (D_{LN} = 0), which would rigorously describe the evolution state.

Finally, the time complexity of D_{LN} is $O(N^2 \cdot n)$, which is equivalent to D_{PW}. However, the computation efficiency of equation 4.1 could be improved with the addition of a condition on the minimum operator. This condition stops the scanning of j when a zero value is found. Therefore, for a converged unimodal population, the time complexity of D_{LN} is reduced to $O(N \cdot n)$, which is the optimal time for a GDM (Wineberg and Oppacher, 2003).

4.3 Performance evaluation of the proposed GDM

4.3.1 Validation framework

First, D_{LN} is analyzed on the reduced population arrangement framework proposed in section 2.4.1. We recall that, in order to respect the diversity requirements established in Table 2.3, the following conditions must be satisfied:

1. Monotonicity in individual varieties: $D\left(P_5 \middle| \hat{x}_5 = 0.5\right) \geq D\left(P_4\right)$;
2. Twinning: $\min D\left(P_5\right) = D\left(P_5 \middle| \hat{x}_5 = 0 \vee \hat{x}_5 = 1\right)$;
3. Monotonicity in distance: $\max D(P_5) = D\left(P_5 \middle| \hat{x}_5 = 0.5\right)$.

The performance results of D_{LN} over this framework are presented in Figure 4.1. The chart shows that it reacts in a similar way to D_{MST} (Figure 2.3 d)) over the reduced population arrangement framework. This suggests that D_{LN} is unable to properly describe the three diversity requirements. The violation of these requirements is related to the fact that the minimum diversity state is achieved with a uniformly distributed population $\left(\hat{x}_5 = 0.5\right)$, while the maximum diversity state is obtained in the presence of duplicate individuals $\left(\hat{x}_5 = 0 \vee \hat{x}_5 = 1\right)$. Furthermore, the diversity function is not monotonic with respect to the location of $\hat{x}_5$. As a matter of fact, the local peaks at $\hat{x}_5 = 0.3$ and $\hat{x}_5 = 0.7$ represent switch points of the characteristic distance of the individuals $\boldsymbol{x}_3$ (1,1) and $\boldsymbol{x}_1$ (0,0) respectively. However, it is worth mentioning that, unlike D_{MST}, D_{LN} penalizes P_5 when it shares the same configuration as P_4 $\left(\hat{x}_5 = 0 \vee \hat{x}_5 = 1\right)$.

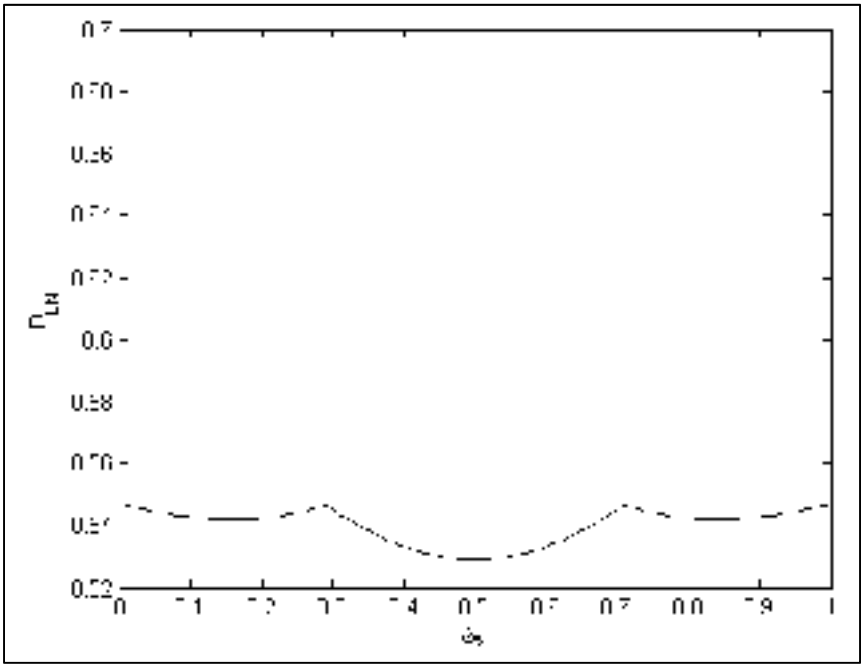


Figure 4.1 Diversity on P_5 (solid curves) and P_4 (dashed curves) with respect to the normalized location of $\boldsymbol{x}_5$ evaluated from D_{LN}

Table 4.1 presents the results of D_{LN} diversity estimation over the seven frozen cases described in section 2.4.2. This framework suggests that D_{LN} conforms to the monotonicity in individual varieties, since Case 7 presents the highest diversity. Moreover, D_{LN} also seems to meet to the twinning requirement, as the diversity levels are identical, regardless of the distribution ratio (Case 2 = Case 3, and Case 4 = Case 5). Finally, the results show that the diversity levels decrease when the optima move closer together (Case 2 < Case 4, and Case 3 < Case 5). Furthermore, the diversity levels of Cases 2 to 5 are close to those of a complete convergence state. Also, these values increase with the number of converging sites.

Table 4.1 Behavior of D_{LN} over the seven frozen cases

GDM	LANDSCAPE	GENOTYPIC DISTRIBUTION CASES						
		1	2	3	4	5	6	7
D_{LN}	2 optima	[illegible]	0.53	0.53	0.88	0.88	1.40	10.13
	4 optima		1.32	1.32	2.08	2.08		

Overall, D_{LN} is no better than the other GDMs tested (section 2.4.3), as demonstrated by the reduced population arrangement experiment. In addition, it is important to note at this point that D_{LN} has an inherent limitation, which may impact the diversity analysis. In fact, application of the minimum operator in equation 4.1 to the lower triangle of the pairwise matrix can be influenced by the order of the individuals in the population. Nevertheless, since no GDM respecting all the diversity requirements has yet been defined, we are forced to select the "least worst" formulation to establish the EEB knowledge and capitalize on it. So, we continue to investigate D_{LN} based on the fact that it demonstrated superior performance on the seven frozen case framework to that of the other GDMs tested.

It is therefore appropriate to evaluate the reaction of D_{LN} with respect to different search processes, and to characterize its response to the desirable quality criteria summarized in Table 2.1. The evaluation is conducted by means of the generic benchmark defined in section 1.4. The first experiment compares D^N_{LN} and D^N_{PW} over a unimodal landscape. Thereafter, the same comparison is made for multimodal landscapes with 2 to 25 converging sites. Finally, the desirable quality criteria (repeatability, robustness, and outlier handling) are analyzed for D^N_{LN}.

4.3.2 Unimodal landscape experiment

For the unimodal landscape experiment, a population size (N) of 100 is used on a two-dimensional landscape ($n = 2$). The results are averaged over 50 repetitions.

Figure 4.2 compares the response of D^N_{LN} with the D^N_{PW} estimations. D^N_{LN} has some difficulty describing the linear pattern associated with this landscape. However, the complete diversity range is adequately covered and the discrimination of the diversity values is correct, even though a small overestimation persists. As demonstrated in a subsequent section (section 4.3.4.2), the observed nonlinear behavior of D^N_{LN} originates from the sampling error related to population size.

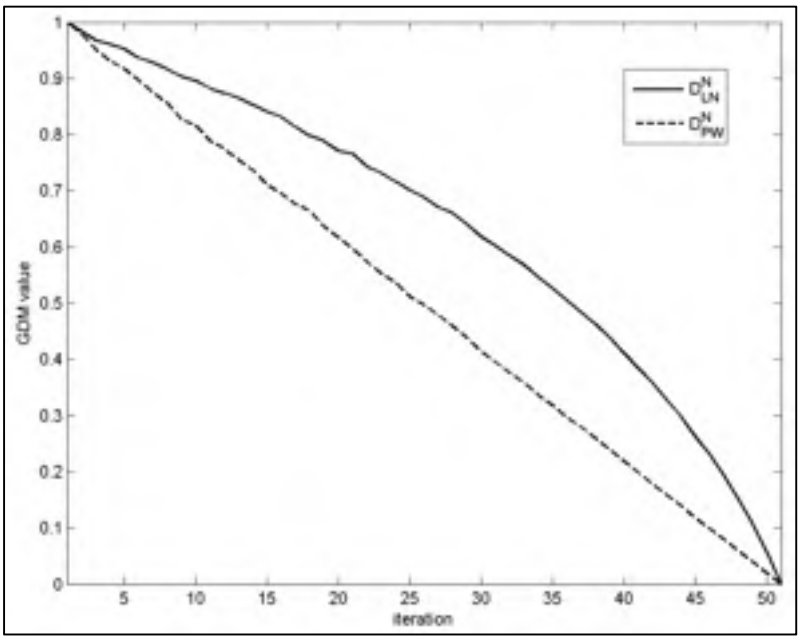


Figure 4.2 Genotypic diversity levels of D^N_{LN} and D^N_{PW} for the unimodal landscape

4.3.3 Multimodal landscape experiments

For the multimodal landscape experiments, the structure (population size, landscape dimensionality, and results averaging) remains identical to that of the previous experiment.

Figure 4.3 compares the results of D^N_{LN} and D^N_{PW} for the various modalities (2 to 25 optima). These experiments involve uniform ratios of individuals attracted by each optimum. As can be seen in Figure 4.3 b), the population end diversities increase with an increase in the optimum. This was expected, since the final diversity corresponds to the distances between the converging sites. However, the population end diversity trend is very dissimilar between D^N_{LN} and D^N_{PW}. D^N_{LN} shows a roughly linear trend up to a diversity level of 32.5%, whereas D^N_{PW} presents a nonlinear trend over the end diversity going up to 94.9%. In reality, the 25 optima landscape converges to a condition where four duplicate individuals are associated with each optimum (N = 100). Consequently, the ultimate diversity is 25% (if the locations of

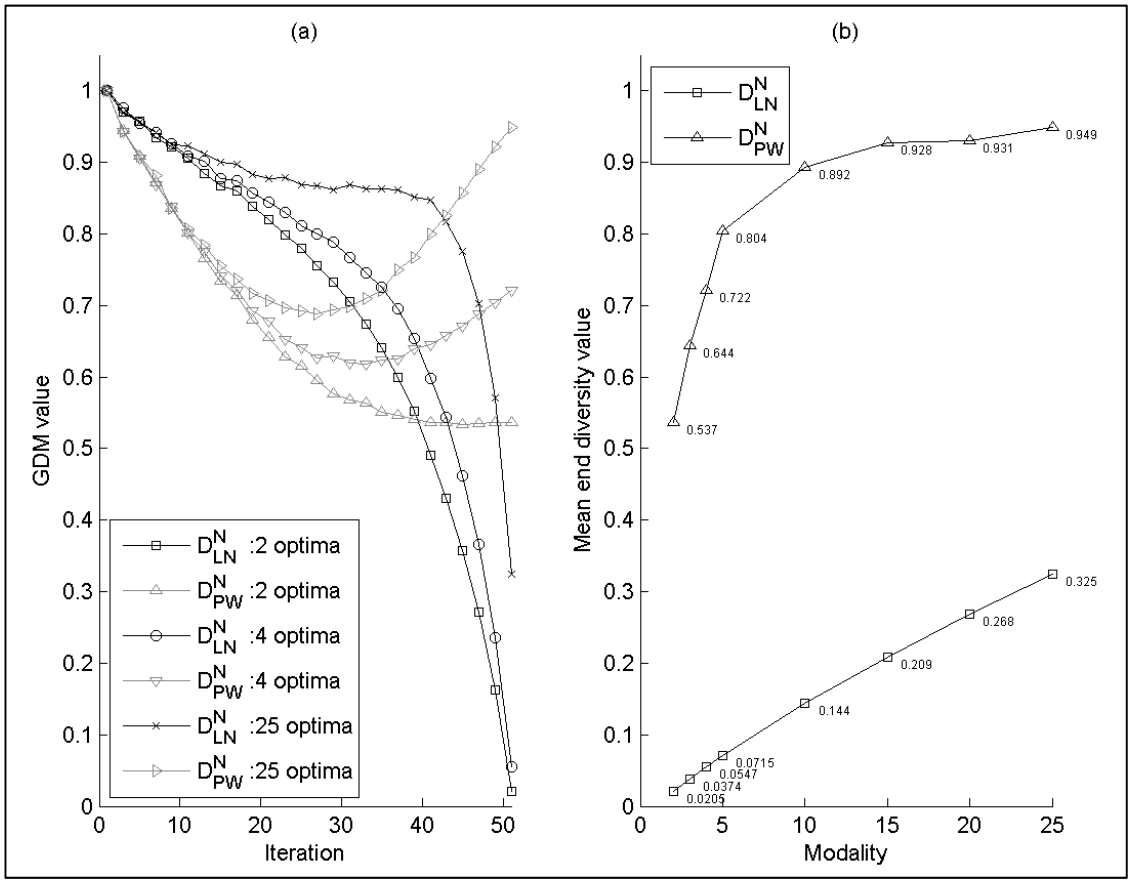


Figure 4.3 Genotypic diversity level of D^N_{LN} (black curves) and D^N_{PW} (gray curves) for a uniformly distributed population: a) Average diversity values over the complete process for three different multimodal structures, b) Average end diversity values for multimodal structures ranging from 2 to 25 optima

the optima are uniformly distributed). The final estimation is 32.5%, which is very close to the true value.

The slope modification observed for D^N_{LN} and D^N_{PW} in the diversity evolutions presented in Figure 4.3 a) results from the hyperspace overlap/separation concept (section 1.5.2) inherent in the generic benchmark. At the beginning of the process, all the hyperspaces around the optima overlap, leading to a diversity level controlled by all individuals. However, when the hyperspaces separate from one another (convergence toward their respective optima), the diversity level of D^N_{LN} becomes controlled by the location of the individuals inside each

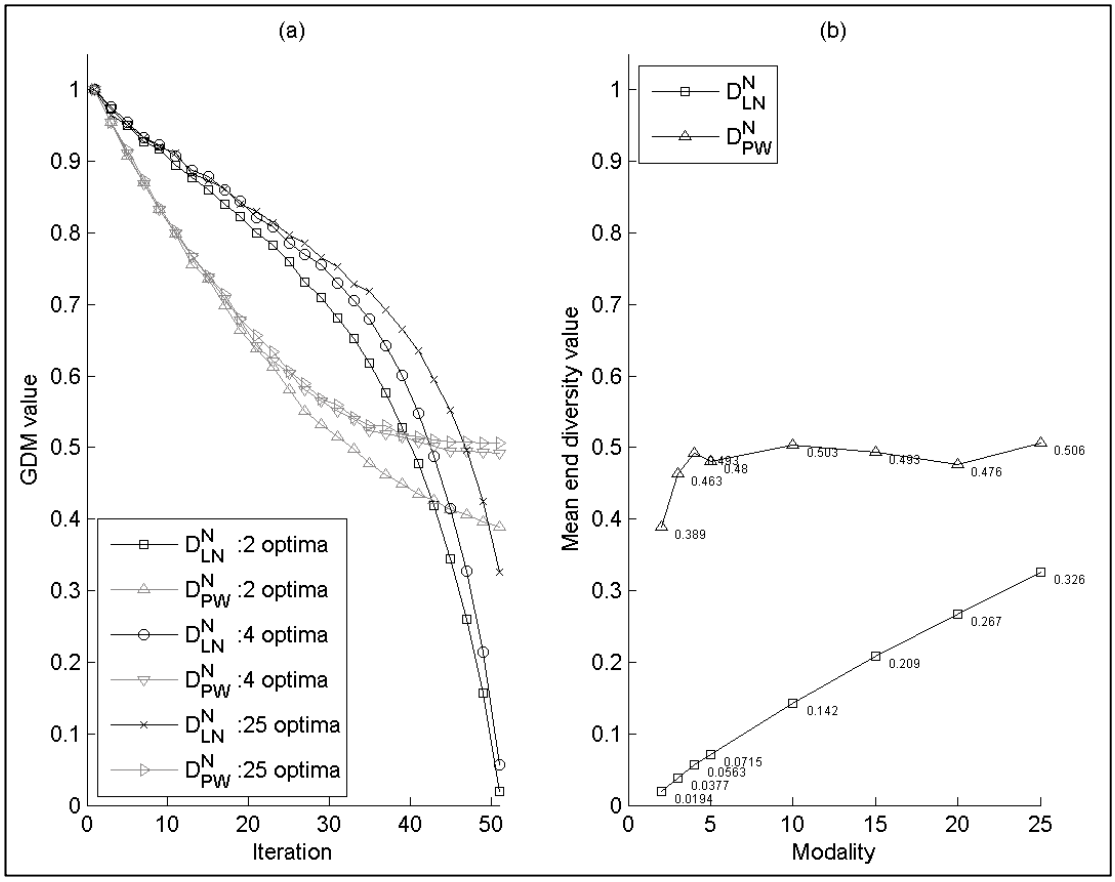


Figure 4.4 Genotypic diversity level of D^N_{LN} (black curves) and D^N_{PW} (gray curves) for a population distributed according to a monopolizing scheme: a) Average diversity values over the complete process for three different multimodal structures, b) Average end diversity values for multimodal structures ranging from 2 to 25 optima

hyperspace (see Figure 1.1). For D^N_{PW}, this latter stage is characterized by an amplification of the differences in optima distances through pairwise measurement.

Figure 4.4 displays the results obtained for the same benchmarks, but with monopolizing ratios. As for frozen Cases 3 & 5 (section 2.4.2), the monopolizing site attracts 70% of the individuals. Figure 4.4 b) indicates that D^N_{LN} is not affected by the ratio of individuals at the convergence state (iteration 51): compared to the uniform case (Figure 4.3 b)), the diversity levels remain unchanged. This is clearly not the case for D^N_{PW}. The ratio of individuals allocated to each hyperspace also has an impact on the diversity evaluation of D^N_{LN} (Figure

4.4 a)). This is expected, however, since the contribution radius of the individuals within the monopolizing hyperspace is reduced, and cannot be entirely compensated for by the other hyperspaces. For instance, compared to the uniform ratio case, the monopolizing hyperspace of the four optimum experiment contains 2.8 times more individuals. In contrast, the three remaining hyperspaces hold 2.5 times fewer individuals.

4.3.4 Desirable quality criteria

4.3.4.1 Repeatability

A reliable GDM should indicate similar diversity levels for a similarly scattered population. Although the results are not included in this chapter, the initial normality assumption of the samples has been tested and invalidated using the Kolmogorov-Smirnov test (0.05 significance level). Stability is therefore evaluated by considering the dispersion range among 96% of the repetition data as defined in section 1.5.3. The dispersion range analysis presents the averaged values for the 51 iteration process. The multimodal landscape stability analysis is not included, since the random positioning among optima would invalidate any interpretation.

Table 4.2 presents the results for D^N_{PW} and D^N_{LN} with four common EA population sizes ($N \in \{50, 100, 300, 500\}$). The analysis is conducted on the two-dimensional unimodal landscape. Since D^N_{LN} shows lower dispersion than D^N_{PW}, which was considered stable in section 1.5.3, D^N_{LN} is also considered stable. Of more interest is the fact that the analysis

Table 4.2 Stability analysis – unimodal landscape, with $n = 2$

GDM	POPULATION SIZE (N)			
	50	100	300	500
D^N_{PW}	[illegible]	[illegible]	[illegible]	[illegible]
D^N_{LN}	[illegible]	[illegible]	[illegible]	[illegible]

clearly reveals the sampling error phenomenon, since the stability of D^N_{LN} (as for D^N_{PW}) increases with a growth in population size.

4.3.4.2 Robustness

This analysis investigates the sensitivity of D^N_{LN} to population size and to the dimensionality of the landscape. The robustness of D^N_{LN} is based on the Friedman statistical test, where each of the samples is characterized by a different population size or a different dimensionality value.

The analysis procedure includes the following considerations (see section 1.5.4):

- At each iteration (uni- and multimodal experiments), a p-value is calculated to validate (or not) the null hypothesis (H_0) stipulating that the sampling distributions are identical.
- A p-value below a predefined level of significance ($\alpha = 0.05$) leads to the rejection of H_0.
- Since the convergence process is repeated 50 times over 51 iterations, the sampling distributions contain the 50 repetition data of a given iteration, resulting in the calculation of 51 p-values. Therefore, the results correspond to the percentage of p-values below α.
- A high percentage indicates that the GDM is sensitive to the sampling parameter or that H_0 was rejected more often than accepted during the convergence process.
- No percentage threshold is defined, since such a figure would be difficult to justify. In addition, as demonstrated below, a postprocessing of this statistical test may be required to accurately determine the sensitivity of a GDM in the case of a high rejection rate.

The results are presented in Table 4.3 for the uni- and multimodal experiments. Population size $N \in \{50, 100, 300, 500\}$ represent the samplings. A two-dimensional landscape is used for this analysis, and each sampling is repeated 50 times over the convergence process. D^N_{LN} appears to be sensitive to population size, whereas D^N_{PW} is quite robust to this parameter. The

results show that the robustness of D^N_{LN} seems to increase with the modality of the landscape structure. Indeed, close to 49% of the iterations with the four optimum landscape accepted H_0, while that percentage was less than 20% for the unimodal experiment. To better illustrate the sensitivity of D^N_{LN} to population size, Figure 4.5 a) presents the mean curves of the four samplings on the unimodal landscape. The curves reveal that this sensitivity is, in reality, not significant. In fact, as shown in Figure 4.5 b), a maximum difference of 0.086 appears during iteration 43, between the two extreme population sizes (50 and 500). A similar trend is observable for the multimodal experiments. Moreover, Figure 4.5 a) clearly indicates that, as the population size increases, the diversity pattern tends to become more linear. In fact, this can be viewed as a direct indication of the influence of the population sampling error on D^N_{LN}. Nevertheless, as demonstrated in Figure 4.5 b) this condition does not jeopardize the diversity analysis. For comparative purposes, the sensitivity behavior of D^N_{PW} is presented in Figure 4.6.

Table 4.3 Sensitivity analysis – population size {50, 100, 300, 500}, with $n = 2$

[illegible]	[illegible]		
	1 optimum	2 optima	4 optima
[illegible]	[illegible]	1.96	[illegible]
[illegible]	[illegible]	68.63	50.98

Table 4.4 presents the sensitivity to dimensionality for D^N_{LN} and D^N_{PW} over the uni- and multimodal experiments. For comparative purposes, the response of D^N_{LN} without the $1/n$ division in its formulation is added to the table. The dimensionalities n are {2, 10, 30}, the population size is fixed at 100, while 50 repetitions are generated during the convergence process for each sampling.

This analysis indicates that D^N_{LN} is sensitive to the dimensionality of the landscape, whereas D^N_{PW} is robust.

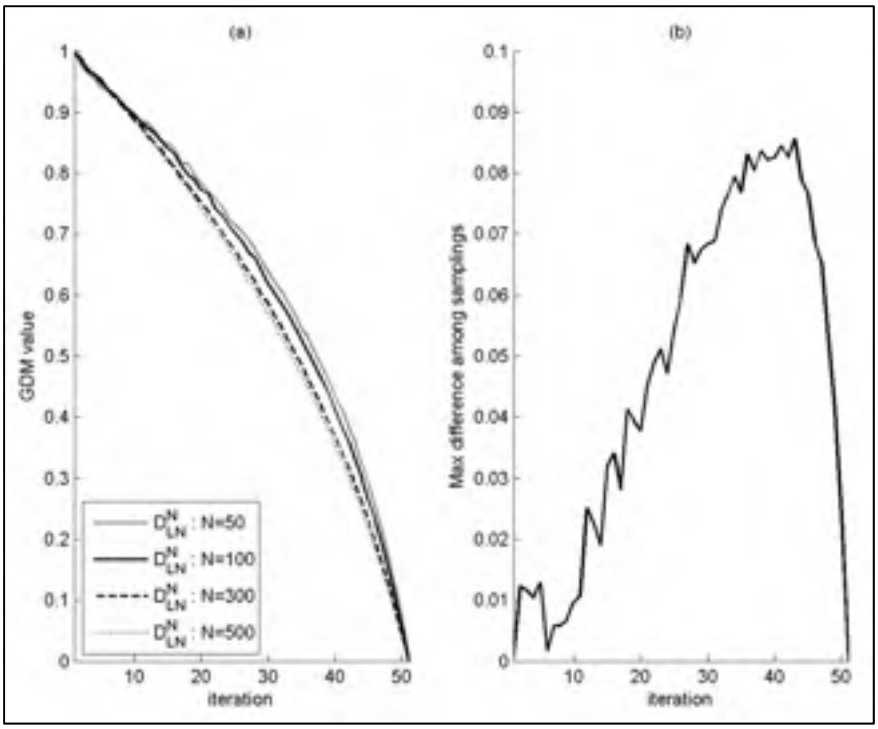


Figure 4.5 a) Effect of population size (N) on D^N_{LN} for the unimodal landscape, b) Maximum difference among the four samplings over the unimodal landscape

Table 4.4 Sensitivity analysis – landscape dimensionality {2, 10, 30}, with $N = 100$

GDM	% p values [illegible]		
	1 optimum	3 optima	4 optima
$D^F_{[illegible]}$	9.61	1.76	3.92
$D^F_{[illegible]}$	92.16	90.20	86.27
$D^F_{[illegible]}$	96.08	96.08	96.08

To better illustrate this sensitivity, Figure 4.7 presents the mean diversity curves for the unimodal landscape, and Figure 4.8 presents the same curves for the multimodal landscape (two optima). In Figures 4.7 a) and 4.8 a), the D^N_{LN} formulation includes the $1/n$ factor, whereas in Figures 4.7 b) and 4.8 b), the formulation does not include this factor.

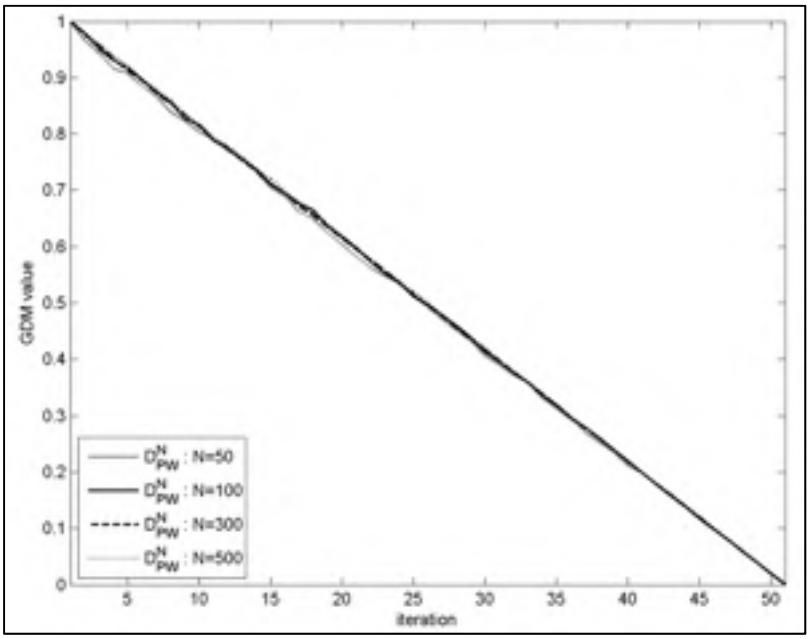


Figure 4.6 Effect of population size (N) on D^N_{PW} for the unimodal landscape

Figure 4.7 a) shows that D^N_{LN} sensitivity remains unimportant. The observation is also valid for the multimodal landscapes (Figure 4.8 a)). Indeed, Figure 4.9 a) reveals that the maximum difference among the three samplings is at most 0.05 and 0.08 for the unimodal benchmark and the multimodal benchmark (2 optima) respectively. Therefore, the impact on the diversity analysis remains very low.

In addition, Figures 4.7 and 4.8 reveal the contribution of the $1/n$ factor. In Figures 4.7 b) and 4.8 b), D^N_{LN} shows a higher sensitivity. Figure 4.9 contributes to the illustration, as follows: a rapid comparison of the curves presented in Figure 4.9 a) and b) indicates that the equation 4.2 is 5.6 times less sensitive than the expression not including $1/n$ for the unimodal landscape, while the sensitivity reduction approaches 2.7 times for the multimodal landscape.

Figure 4.7 a) shows that the n=30 curve is located between the other two dimensionality curves. Additional tests with a dimensionality between 2 and 50 suggest that the n = 2 and n = 10 curves are the lower and upper bound curves respectively. This observation better

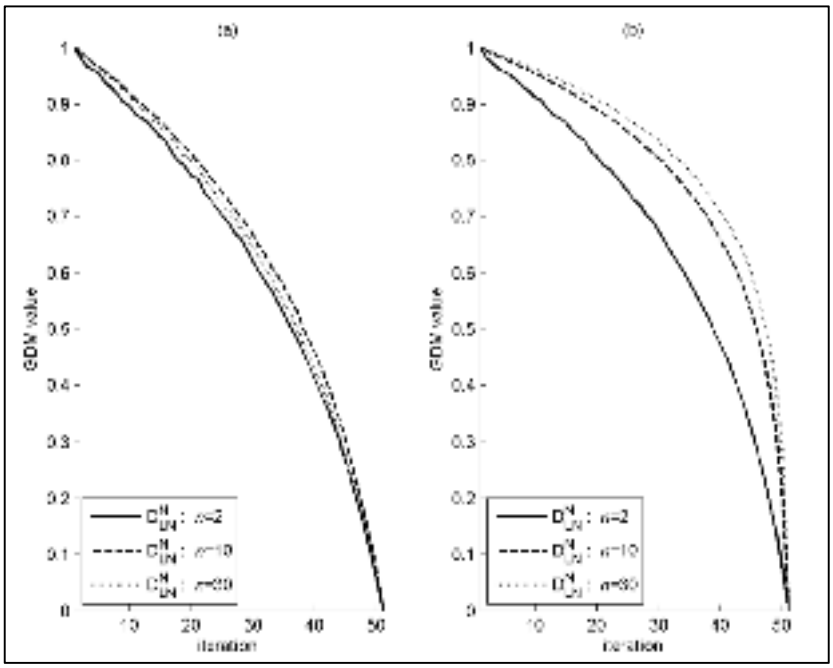


Figure 4.7 Effect of landscape dimensionality (n) on D^N_{LN} for the unimodal landscape: a) with $1/n$ in the D^N_{LN} formulation, b) without $1/n$ in the D^N_{LN} formulation

depicts the real sensitivity of D^N_{LN}, which is actually very low. However, as revealed in Figure 4.8 a), this phenomenon is limited to the single-site convergence process.

For comparative purposes, the sensitivity of D^N_{PW} is presented in Figure 4.10. As shown by the statistical test (Table 4.4), the mean D^N_{PW} curves reveal insensitive behavior.

In reality, the analysis of the mean curves reveals the limitation of the statistical test, which is that it does not account for the sensitivity amplitude. A 1st degree analysis would have portrayed D^N_{LN} as a sensitive diversity descriptor. However, in light of Figures 4.7 to 4.9, the sensitivity of D^N_{LN} appears not to be a dominant characteristic.

Consequently, we have to conclude that, even though D^N_{LN} is more sensitive than D^N_{PW} to landscape dimensionality and population size, it offers the robustness required for excellent diversity estimation, regardless of the values of those factors.

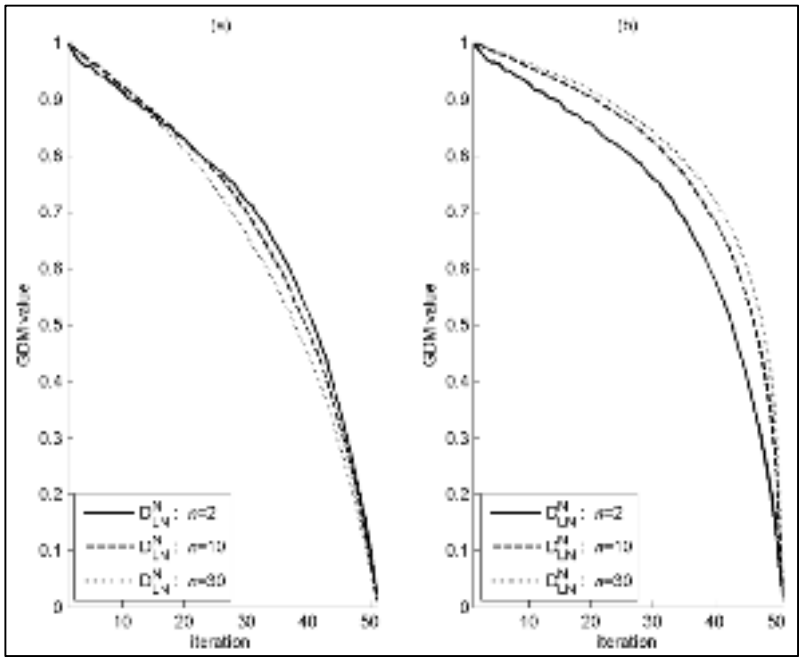


Figure 4.8 Effect of landscape dimensionality (n) on D^N_{LN} for the multimodal landscape (two optima): a) with $1/n$ in the D^N_{LN} formulation, b) without $1/n$ in the D^N_{LN} formulation

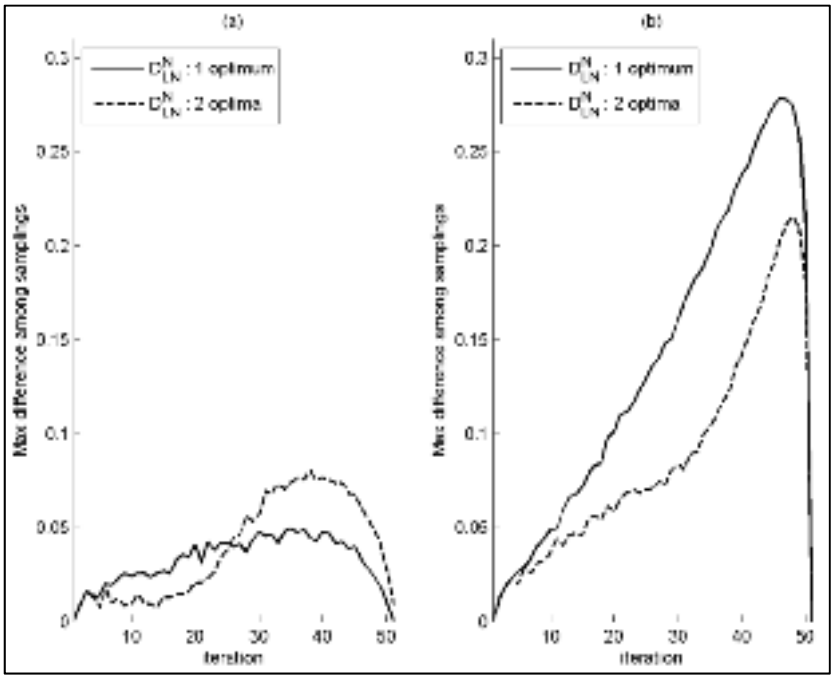


Figure 4.9 Maximum GDM value difference among the three samplings over the unimodal and multimodal (two optimum structure) landscapes: a) with $1/n$ in the D^N_{LN} formulation, b) without $1/n$ in the D^N_{LN} formulation

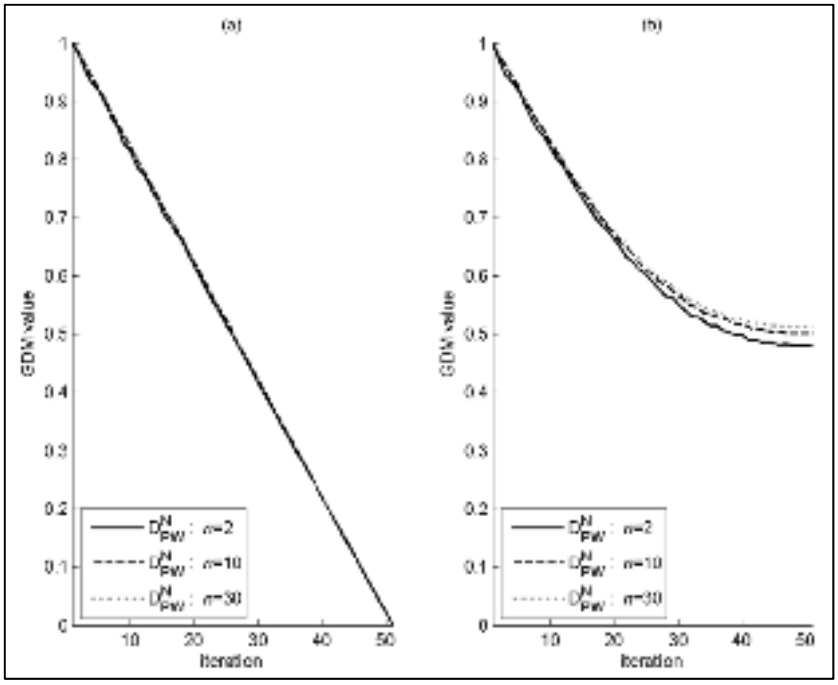


Figure 4.10 Effect of landscape dimensionality (n) on D^N_{PW} for: a) the unimodal landscape, b) the multimodal landscape (two optima)

4.3.4.3 Influence of outliers

Outliers increase the diversity level of a population. However, their final influence is determined by their number. Suitable diversity estimations should reflect this condition. Moreover, since, by definition, outliers represent a small fraction of the population, they should under no circumstances dominate the diversity measure. In order to portray the D^N_{LN} response to the presence of outliers, the previous benchmark is employed with no modification up to the 10th iteration. Then, a fraction of the population (1%, 2%, 5%, and 10%) is randomly repositioned between the 1st and 10th iteration boundaries. The outlier positions are rearranged inside these limits at each iteration.

Figure 4.11 gives the mean D^N_{LN} response for the four test cases applied to the two-dimensional unimodal landscape with a population size of 100. The results are similar for

different population sizes and multimodal landscapes. The curves clearly describe an efficient representation of the outlier influence: the diversity pattern is translated from the 0-outlier position by a value related to the percentage of outliers. Consequently, the diversity level ended at 0.02, 0.04, 0.07, and 0.12 for the 1%, 2%, 5%, and 10% outlier cases respectively.

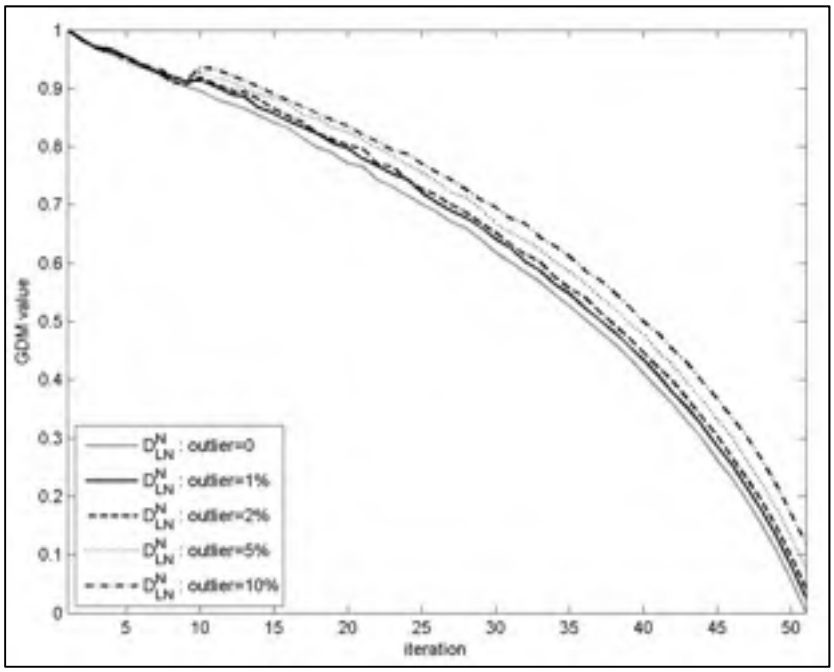


Figure 4.11 Effect of outliers on D^N_{LN} for the unimodal landscape

For comparative purposes, Figure 4.12 presents the same experiment for D^N_{PW}. It appears that D^N_{PW} overestimates diversity in the presence of outliers compared to D^N_{LN}. In fact, the end diversity increases by a factor of up to 2.6 as the outlier percentage grows by 10%. This condition emanates from the formulation of D^N_{PW}, where the outlier distance is considered in combination with the complete population. In other words, the imperfect incorporation by D^N_{PW} of the outlier contribution and the resulting overestimation of the diversity is related to the duplicate individual treatment discussed in section 2.4.

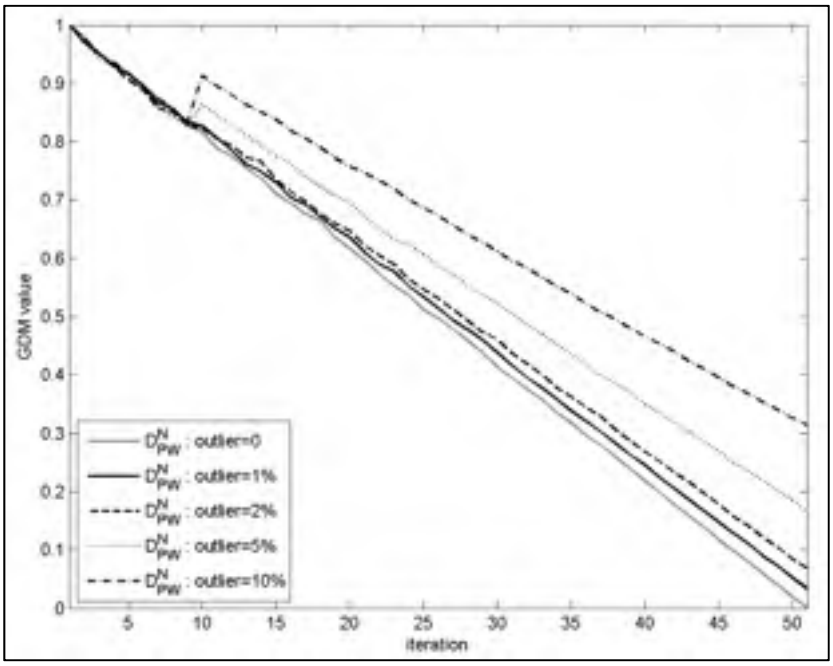


Figure 4.12 Effect of outliers on D^N_{PW} for the unimodal landscape

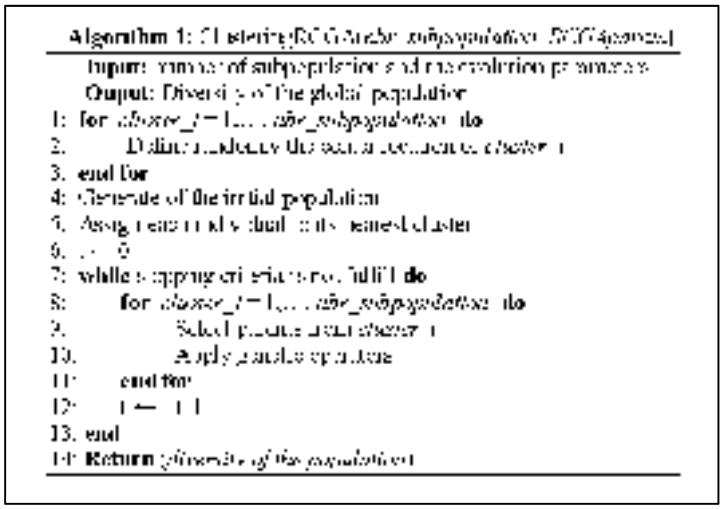
4.4 CEC'05 benchmark GDM comparison

In order to complete the study, D^N_{LN} and D^N_{PW} were tested on the CEC'05 benchmark (Suganthan *et al.*, 2005) (see appendix I for an overview) and a real-coded genetic algorithm (RCGA) with common parameters (Herrera, Lozano, and Verdegay, 1998). Like any other search method, RCGA is affected by a search bias. However, it is possible to implement strategies promoting diversity. The concept of subpopulations without immigration appears to be well adapted to the test conducted here. The strategy is implemented through clustering, which leads to better tracking of the treatment of duplicate individuals by GDMs. In this way, the overestimation of the population diversity by D^N_{PW} should be accurately depicted. At the same time, as previously illustrated with the generic unimodal landscape, if only one population is used instead of multiple subpopulations, D^N_{PW} and D^N_{LN} should result in similar diversity estimations.

Since performance achievement is not the aim of this comparison, no specific parameter setting approach is introduced for RCGA response improvement (Lobo, Lima, and Michalewicz, 2007). The population size is fixed at 100, and a binary tournament is chosen as the selection process. BLX-0.5 serves as the crossover operator, and non uniform mutation with b = 5 is considered. The crossover probability (p_c) is 0.6, and the probability for mutation (p_m) is 0.005. The purpose of this study is to compare and illustrate the response of D^N_{PW} and D^N_{LN} over a recognized benchmark.

Algorithm 4.1 shows the clustering RCGA implementation. At the beginning, the clusters are chosen by a random positioning of their centers over the landscape. The initial population is generated, and the individuals are allocated to the nearest cluster. Thereafter, the evolution follows the normal EA process. However, the subpopulations evolve in parallel, and no immigration is allowed. Consequently, a subpopulation owning a lone individual would show practically no sign of evolution. The process ends when the function evaluation exceeds $10\ 000 \cdot n$. This limit is prescribed in CEC'05. A condition where each subpopulation would present only duplicate individuals within a 10^{-6} tolerance also stop the process. Finally, an elite is preserved within each subpopulation.

Algorithm 4.1 Clustering RCGA search procedure



In order to control the length of this section, only three CEC'05 benchmark problems are covered: 10-dimensional F2 (shifted Schwefel problem 1.2), 10-dimensional F8 (shifted-rotated Ackley's function with global optimum on bounds), and 10-dimensional F11 (shifted rotated Wieierstrass function). Function F2 is a unimodal problem, and the others are multimodal. The simulations are conducted for different numbers of subpopulations, and each is repeated 25 times. The results are presented for the median run of the simulations with 1 population and 10 subpopulations, while the mean of the final diversity is monitored for 9 different subpopulations (1, 2, 3, 4, 5, 10, 15, 20, and 25). Note that none of the simulations reached the global optimum within the allowed CEC'05 tolerance. Table 4.5 summarizes simulation performance, and Figures 4.13 to 4.15 present the results for F2, F8, and F11, respectively. Table 4.5 includes, in parenthesis, the CEC'05 error tolerance associated with each function to consider having reach the global optimum and the number of runs terminated by the duplicate individuals criterion.

Table 4.5 Mean error of the best solutions with respect to the global optimum over the 25 repetitions

Problem	Numbers of subpopulation								
(10-D)	1	2	3	4	5	10	15	20	25
F2	1.69	24.59	95.23	160.53	212.01	316.52	1261.72	1960.93	2619.99
(6.1e-6)	(0)	(0)	(0)	(0)	(1)	(2.5)	(17)	(10)	[illegible]
F8	20.18	20.18	20.[illegible]	20.[illegible]	20.17	20.35	20.32	20.35	20.27
(7.1e-2)	(0)	(0)	(0)	(1)	(9)	(22)	(16)	(11)	(4)
F11	[illegible]	[illegible]	4.45	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
(7.1e-2)	(25)	(24)	(25)	(25)	(25)	(25)	(15)	(12)	(4)

Table 4.5 reveals that, even if the optimum is not properly located, runs often terminate in a convergence condition (number in parenthesis close to 25). By relaxing the duplicate individuals condition threshold to 10^{-2} instead of 10^{-6}, most of the run terminate in converged subpopulations state. Figures 4.13 and 4.15, representing F2 and F11, indicate, for the simulations with 1 cluster (without subpopulations), that the two GDMs present similar diversity trends and end in a full convergence state. Only F8 (Figure 4.14) maintained non null diversity levels at the end of the process for the single cluster simulations. At the same time, for cluster numbers greater than one, the results clearly show final diversity levels

higher than zero in all cases. The ten cluster curves in Figures 4.13 to 4.15 a) illustrate this condition, and, after a given number of generations, the diversity stabilizes at a constant level. Moreover, Figures 4.13 to 4.15 b) show the influence of the cluster number on the diversity estimations: increasing the number of subpopulations causes a related increase in final diversity. Comparable trends are shown by both D^N_{PW} and D^N_{LN}. As mentioned earlier, this relationship originates from the distance between the cluster locations.

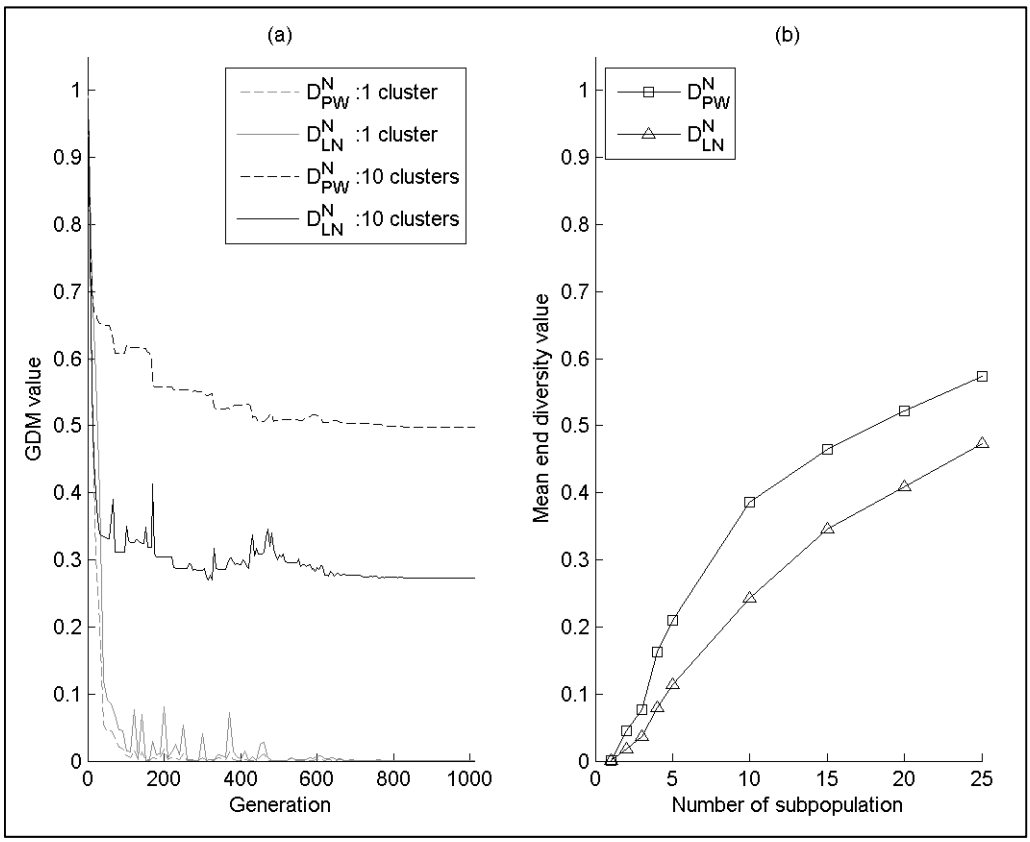


Figure 4.13 Genotypic diversity level of D^N_{LN} and D^N_{PW} for the 10-D F2 function: a) Median run with 1 and 10 subpopulations, b) Mean of the end diversity for different numbers of subpopulations

The curves in Figures 4.13 to 4.15 b) also allow a comparative analysis of the descriptors. For example, when the population converges on multiple sites, the final diversity estimated by D^N_{PW} is 1.6 and 1.9 times higher than the D^N_{LN} evaluation for the uni- and multimodal problems respectively. This important difference is attributable to the consideration of duplicate individuals within D^N_{LN}. Finally, the experiments illustrate the strong influence of the underlying benchmark and search strategy on the resulting diversities, and therefore justify the exploitation of generic benchmarks isolating the descriptor influence, such as the one used in this study for GDM comparison.

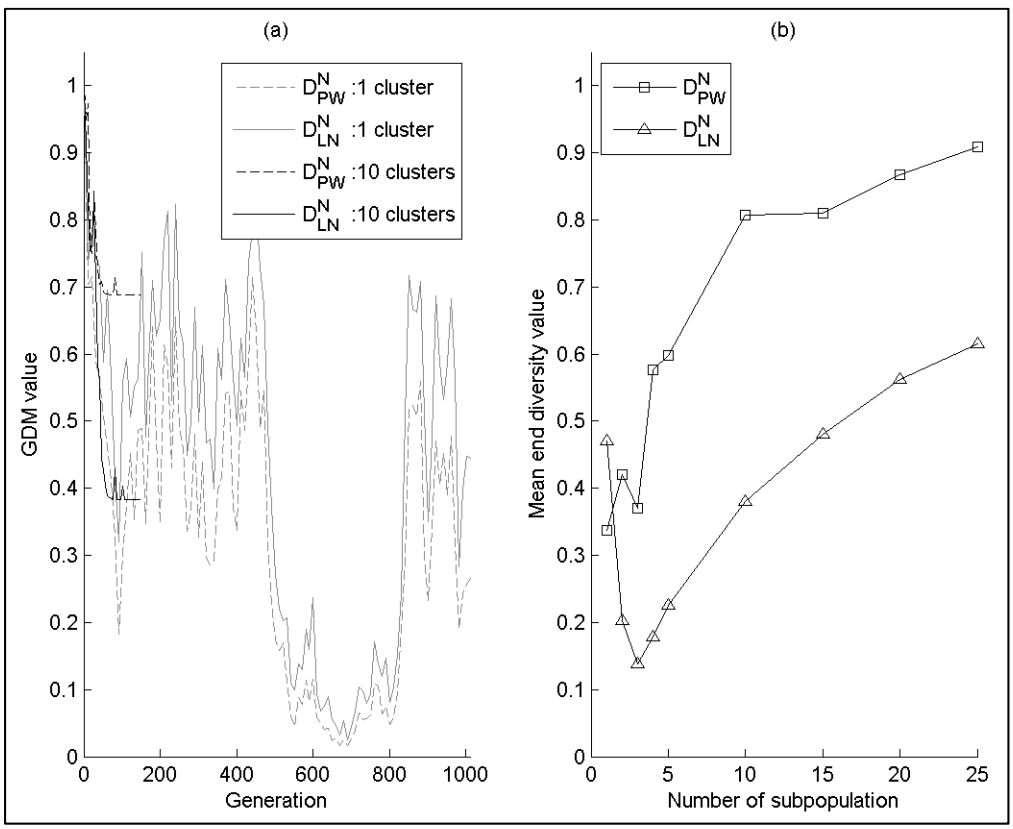


Figure 4.14 Genotypic diversity level of D^N_{LN} and D^N_{PW} for the 10-D F8 function: a) Median run with 1 and 10 subpopulations, b) Mean of the end diversity for different numbers of subpopulations

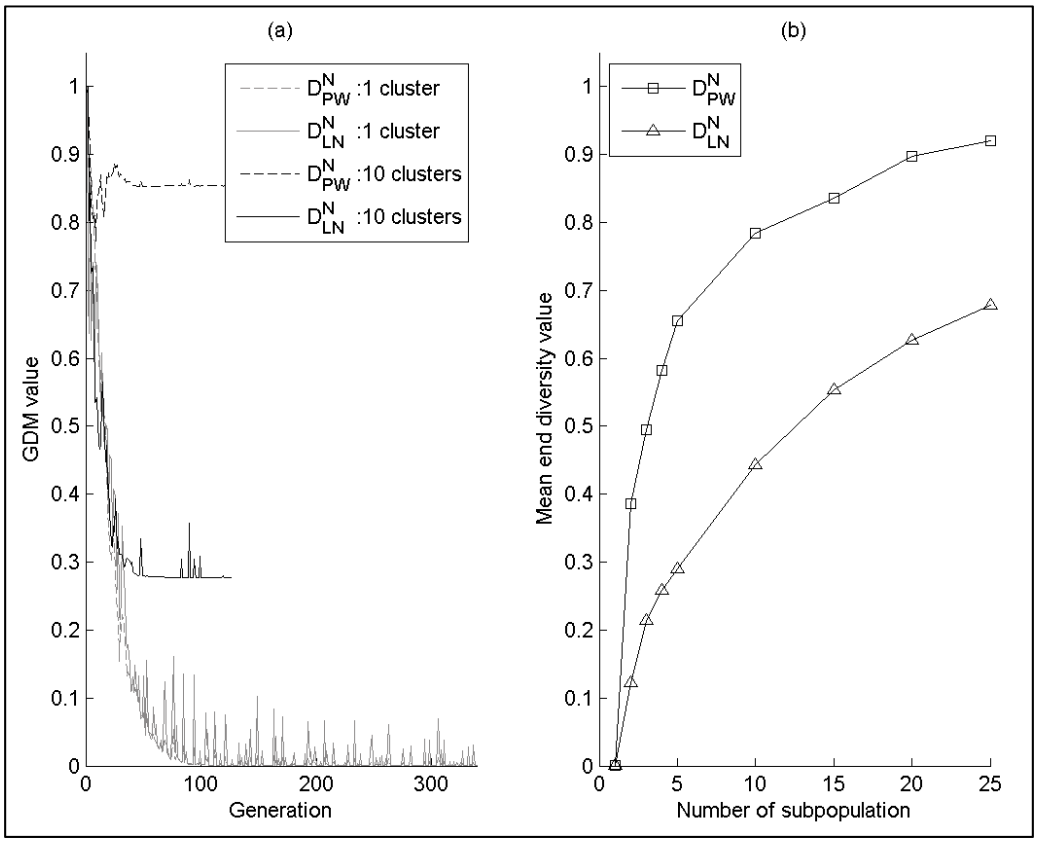


Figure 4.15 Genotypic diversity level of D^N_{LN} and D^N_{PW} for the 10-D F10 function: a) Median run with 1 and 10 subpopulations, b) Mean of the end diversity for different numbers of subpopulations

4.5 Conclusion

Genotypic diversity measurement is a useful concept for monitoring and/or managing the exploration task of an optimization process. Premature convergence towards a suboptimal solution can be minimized based on the information gathered by a GDM. Numerous GDMs have been proposed in the literature over the years. However, a detailed investigation (chapter 2) reveals that none of the potential formulations adequately describes population diversity.

We propose a new distance-based GDM formulation (D_{LN}) to better describe population diversity in chapter 3. This new formulation is based on the central idea of aggregating all the

radii contributed by the individuals (a radius being the distance between an individual and its nearest neighbor). The resulting D_{LN} formulation automatically eliminates the contribution of duplicate individuals. However, this formulation has been introduced based on the premise that it properly reflects population diversity, and so the purpose of this chapter is to test this assumption. To achieve this objective, the performance of D_{LN} is evaluated by means of the validation framework defined in this chapter.

D_{LN} performed poorly on the reduced population arrangement, as it was unable to respect any of the diversity requirements defined in Table 2.3. In fact, it behaves like the other potential GDM formulations. Nevertheless, over the seven frozen cases of controlled population diversity, D_{LN} met all the diversity requirements, which means that its performance is superior to that of the other potential GDMs (Table 2.5). Since these two benchmarks are considered to be complementary, D_{LN} cannot be considered to adequately describe population diversity. In addition, D_{LN} has a design limitation that may impact its population diversity measurement, which is that the radius of each individual to its nearest neighbor is selected only from the lower triangle of the pairwise matrix. Consequently, the ordering sequence of the individuals within the population can have an influence. However, further investigation was conducted with D_{LN} without any apparent sign arising from this limitation.

We then compared the stability of D_{LN} to that of D_{PW} by testing their normalized versions (D^N_{LN} and D^N_{PW}) over the generic benchmark proposed in chapter 1, and found that, over a similarly scattered population, D^N_{LN} offers significantly more stable descriptions than D^N_{PW}. In contrast, the new formulation appears to be slightly more sensitive to the simulation parameters (N and n) than D^N_{PW}. Nevertheless, a detailed analysis of the response of the descriptor has established that this apparent sensitivity has no real influence on the diversity estimation. The new proposal also offers very accurate diversity estimations for populations containing outliers. Finally, experiments conducted with a clustering GA over the CEC'05 benchmark corroborate our finding that duplicate individuals are better treated by D^N_{LN}.

So, our investigation shows that, even though it is difficult for D^N_{LN} to meet all the diversity requirements, it behaves better overall than the other potential GDM formulations. Consequently, until a formulation is developed that respects all three diversity requirements, we recommend the use of D^N_{LN} to represent the exploration axis of a search process EEB.

Having found a way to portray the EEB concept, we can now leverage this knowledge to achieve the main objective of this thesis, which is to develop an adaptive strategy for controlling all the EA parameters at once. We do this in the next chapter.

CHAPTER 5

BAYESIAN NETWORK AS AN ADAPTIVE PARAMETER SETTING APPROACH FOR GENETIC ALGORITHMS

Parameter setting is one of the most active research topics in the evolutionary algorithm (EA) community. This can be explained by the major impact that EA parameters have on search performance. However, parameter setting has been shown to be both problem-dependent and evolution-dependent. Moreover, parameters interact in complex ways. Consequently, the development of an efficient and beneficial parameter setting approach is a rather difficult task, and no widely recognized solution has emerged to date.

In this chapter, we borrow the notion of parameter adaptation with the objective of addressing the parameter setting dependencies mentioned above, using a strategy based on Bayesian network. The adaptive framework is elaborated for a steady-state genetic algorithm (SSGA) to control 9 parameters. To judge parameter state productivities, we consider fitness improvement of the population and management of the exploration/exploitation balance (EEB). The performance of this proposal, Bayesian Network for Genetic Algorithm parameters adaptation (BNGA), is assessed based on the CEC'05 benchmark. BNGA is compared to static parameter setting, a naïve approach, three common adaptive systems (PM, AP, and FAUC-RMAB), and two state-of-the-art EAs (CMA-ES and G-CMA-ES). Our results demonstrate statistically that the performance of BNGA is equivalent to that of FAUC-RMAB, CMA-ES, and G-CMA-ES, and is superior overall to all the other SSGA parameter setting approaches. However, these results also reveal that all the approaches considered have great difficulty finding global optima in a multimodal problem set, which suggests a lack of complementarity and/or synergy among parameter states.

This chapter was submitted to *Applied Soft Computing* in June 2012 (Corriveau *et al.*, 2012c).

5.1 Introduction

Metaheuristics inspired by nature, such as evolutionary algorithm (EA), form a major class of the optimization methods used today (Whitacre, 2011a). One of the reasons for their success is that they provide the flexibility needed to solve diverse engineering problems (Whitacre, 2011b). However, this flexibility comes at the cost of defining and setting multiple internal parameters, which is a difficult task. This is because:

1. The parameters are problem-dependent (Wolpert and Macready, 1997);
2. They can change as the process evolves (Bäck, 1992);
3. The interactions among them can be complex (Rojas *et al*., 2002).

The issue of problem dependency is confirmed by the No Free Lunch (NFL) theorem, as individual parameter configurations represent different search path optimizers. The issues of evolution dependency and interaction dependency are mostly acknowledged through empirical studies.

Parameter setting approaches began to be developed with the emergence of the EA field, with the aim of providing guidelines to practitioners (De Jong, 1975; Grefenstette, 1986; Schaffer *et al*., 1989). In the early 1990s, the emphasis shifted toward control systems and tuning methodologies instead of guidelines, when it became clear to researchers that general recommendations were of little use (Eiben, Hinterding, and Michalewicz, 1999). After more than 35 years, parameter setting is still an important active field of research (Lobo, Lima, and Michalewicz, 2007), since no widely recognized and satisfactory solution has yet been found. Aside from the fact that parameter setting approaches may help alleviate practitioners' efforts to some extent, they play a major role in the achievement of enhanced EA performance. In fact, EA parameters are responsible for providing a specific exploration/exploitation balance (EEB) over a given problem (Eiben and Schippers, 1998), and that balance dictates the search path to be followed. As such, parameter setting is a stepping stone to providing an

optimization framework to practitioners that is both robust (high quality solutions for a wide range of problems) and efficient (high quality solutions delivered as quickly as possible).

Eiben, Hinterding, and Michalewicz (1999) have provided the most comprehensive taxonomy of parameter setting approaches to date. They define four categories of approach: parameter tuning, deterministic control, adaptive control, and self-adaptive control. Parameter tuning methods are designed to fix parameters prior to the optimization process. Parameter control methods change parameters as the process evolves. Deterministic control methods are iteration-based formulations that do not take into account feedback from the process. Adaptive control methods use feedback gathered from the search and apply it in some way to select parameter states. Finally, self-adaptive control refers to techniques in which parameters are directly encoded in individuals with the aim of allowing evolution to identify the most promising parameter states. Of course, each of these categories has its own advantages and drawbacks (Eiben, Hinterding, and Michalewicz, 1999). What we can say, based on the above discussion, is that parameter tuning and deterministic control are not capable of characterizing all three parameter setting dependency issues identified. In contrast, the parameter setting capability of self-adaptive control is based entirely on the fitness of the individuals. However, these raw values are often not sufficient to characterize the search behavior. Population diversity, for instance, is important in guiding the search process (see Chapter 1). Therefore, we conjecture that the remaining option, adaptive control, is the most attractive class for steering the search toward an optimal path.

Over the years, numerous adaptive parameter control approaches have been proposed, but all of them applied to a restricted number of EA parameters. In fact, a review of more than 100 papers on parameter adaptation (Figure 0.2) reveals that more than 88% of the approaches concern the adaptation of at most two parameters, and only 3% involve four parameters. We found no study in which the adaptation of more than four parameters was proposed. Furthermore, only a few of those approaches have the potential to handle parameter interactions, and most approaches in this subset are based on fixed interactions. This limits their adaptation effectiveness in terms of achieving an optimal search path, as parameter

interactions may change as the process evolves or for different problems. This suggests that EA parameter adaptation is a rather complex task, and one that remains an open question. Our goal in this chapter is to propose a new adaptive control framework that enables the management of all three parameter setting dependencies.

This chapter is organized as follows: section 5.2 reviews the adaptive frameworks proposed in the literature; section 5.3 presents our new adaptive control system; section 5.4 describes the methodology we use to compare selected parameter setting approaches; section 5.5 presents the results of this comparative study; and, finally, section 5.6 provides a concluding discussion.

5.2 Review of adaptive parameter control strategies

There are many aspects of an adaptive control strategy that need to be carefully defined. These aspects are described below, and the adaptive control process is illustrated in Figure 5.1. We review the components of the process individually, as they each have a different function.

- The type and states of the parameters involved;
- The feedback indicator used to evaluate the impact of the current state (j) of parameter i;
- The window interval (W) on which the adaptation is conducted;
- The credit assignment scheme required to convert feedback information into a suitable reward;
- The parameter selection rule used to update parameter states.

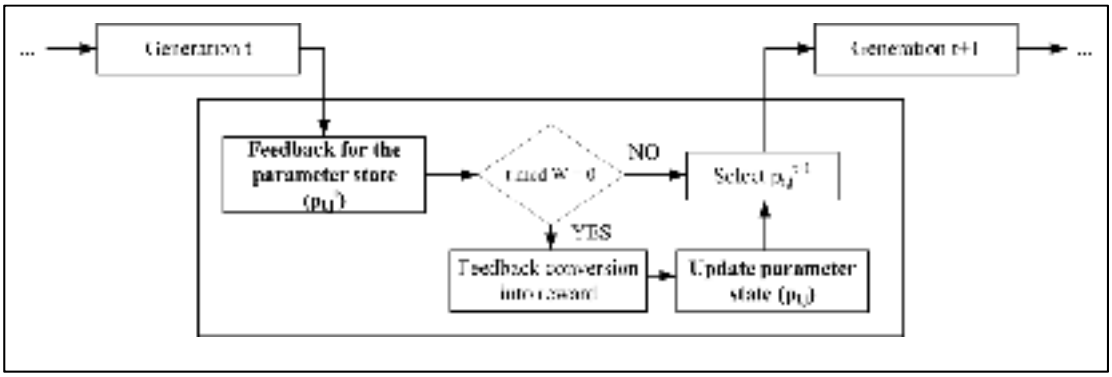


Figure 5.1 General framework describing the adaptive process (steps in bold type refer to fundamental components)

5.2.1 Parameters involved

Here, we focus on genetic algorithm (GA), as this search method is widely used and requires the setting of many parameters. In general, six parameters must be set to conduct a GA search, including population size (*N*), selection mode and genetic operators through crossover and mutation types, as well as their respective activation probabilities, p_c and p_m. This ensemble can become larger if we consider multi-parent selection (μ) and/or multiple offspring creation per couple (λ). In addition, supporting the steady-state evolution model (SSGA) adds replacement and update plans to parameter selection. Throughout this chapter, parameter setting is looked at in the broadest sense of the term, including both parameters and operators.

One of the strongest motivations for conducting parameter adaptation is to control the EEB throughout the search. Each parameter has its own impact on this fundamental search characteristic. For instance, increasing *N* favors exploration of the landscape, whereas increasing selection pressure directs the search toward exploitation of the most promising solutions. Notwithstanding these tendencies, any of the genetic operators can be seen as having an exploration or an exploitation function, depending of its formulation (Eiben and Schippers, 1998). That is why it is important to adapt all the parameters.

5.2.2 Feedback indicators

Feedback indicators can be defined as the evidence on which adaptation takes place (Smith and Fogarty, 1997), since they act as the driving force for adaptation by assessing the productivity of the parameter states. Its choice is fundamental for any adaptive framework, as wrong advice sent to the update mechanism may adversely affect the adaptation process (Tuson and Ross, 1998).

Since the majority of adaptive studies cover genetic operators and their activation probabilities, we must point out that many feedback formulations are based on the improvement achieved by the offspring ($f_{offspring}$) relative to a reference fitness value. This reference could be the best individual in the population (f_{best}) (Davis, 1989), the parents (f_{parent}) (Tuson and Ross, 1998), or any population-based statistics (Julstrom, 1995; Barbosa and Sá, 2000). In almost all applications, a null impact is allocated when no improvement is registered from $f_{offspring}$, instead of penalizing the corresponding parameter state.

Another widely applied approach is to consider diversity as a feedback indicator, since it is a measure commonly used to express the EEB of a search process. Diversity can be formulated from the location of the individuals in the search space (i.e. genotypic formulation) or from their fitness distribution (i.e. phenotypic formulation). Various proposals have been developed regarding genotypic diversity measures (GDMs) (see chapter 1) and phenotypic diversity measures (PDMs) (Herrera and Lozano, 1996) (see chapter 3). It is important to note, however, that PDMs are inherently limited when they are used alone (Motoki, 2002). Consequently, feedback indicators using both GDM and PDM have also been proposed, with the aim of describing the quantity and the quality of the population diversity respectively (Herrera and Lozano, 1996). Finally, it is interesting that in (Maturana and Saubion, 2008) three aspects were considered: a GDM, a quality measure defined as the average fitness (f_{avg}) variation, and the mean execution time of the selected parameter state.

5.2.3 Credit assignment scheme

Credit assignment is used to convert a feedback indicator into a suitable form supported by the adaptation mechanism (i.e. selection rule) and/or for aggregating multiple feedback indicators. Normalization of a feedback indicator can be regarded as a simple credit assignment scheme. In fact, normalization is helpful in reducing the dependency issue related to feedback indicators built from raw values (Fialho, 2011). In contrast, ranking concept has also been proposed as a means to alleviate the concern about raw values (Fialho, 2011). Finally, the proposal in (Whitacre, Pham, and Sarker, 2006) focuses on rare but substantial improvement based on the probability of producing exceptionally good solutions.

Whatever credit assignment scheme is promoted, the frequency rate of the adaptation schedule must be defined. This is reflected by window size (W), a hyperparameter whose impact was investigated in (Tuson and Ross, 1998). Results suggest that W may be sensitive and problem-dependent, although it was found that W is relatively robust overall. Three streams are employed for conferring a reward to a selection rule: instantaneously conferring the reward ($W = 1$), averaging the rewards over W, or conferring the extreme reward encountered along W.

5.2.4 Parameter selection rule

The parameter selection rule can be considered as the core of any adaptive system. It is responsible for incorporating past performance into guidance for current parameter selection, and doing so automatically. Numerous proposals have been put forward over the years. For the sake of clarity, they are grouped into six families.

5.2.4.1 Heuristic rule

This family encompasses various kinds of selection rules. Their common denominator is that they are specifically designed for a given parameter. Because of this limitation, only the most

famous of them is presented here, which is Rechenberg's "1/5 success rule", proposed within the evolution strategy (ES) paradigm. He stipulates the following (Rechenberg, 1973, p.123): "*The ratio of successful mutations to all mutations should be 1/5. If it is greater than 1/5, increase the standard deviation, if it is smaller, decrease the standard deviation.*"

5.2.4.2 Fuzzy Logic Controller (FLC)

The fuzzy logic controller (FLC) allows the definition of fuzzy IF-THEN rules. This approach is useful when knowledge is vague, incomplete, or ill-structured (Herrera and Lozano, 1996). Well-defined fuzzy sets are, however, quite difficult to form, and membership functions can be challenging to value. So, in general, FLCs rely on the knowledge and experience of experts (Herrera and Lozano, 1996).

5.2.4.3 Probability Matching (PM)

The probability matching (PM) method was proposed by Goldberg (1990) as a reasonable adaptive strategy for an environment characterized by uncertainty. The underlying idea of PM is to translate reward values into a corresponding selection probability. PM is formalized in the following mechanism:

$$\hat{q}_j^{t+1} = (1-\alpha)\hat{q}_j^t + \alpha \cdot r_j^t \tag{5.1}$$

$$p_{i,j}^{t+1} = p_{i,\min} + \left(1 - J \cdot p_{i,\min}\right)\frac{\hat{q}_j^{t+1}}{\sum_{j=1}^{J}\hat{q}_j^{t+1}} \tag{5.2}$$

At generation t+1, state j of parameter i is selected following the probability stored in $p_{i,j}^{t+1}$. This selection probability is built from the past empirical quality estimate of state j ($\hat{q}_j^t$) and the current reward (r_j^t) provided by the credit assignment scheme. These two components are connected through an additive relaxation mechanism described by equation 5.1 and

controlled by an adaptation rate parameter ($0 < \alpha \leq 1$). This parameter allows the system memory to be reduced as past knowledge is increasingly forgotten with the rise of α. To ensure that no parameter state is lost in the process, a minimum probability ($p_{i,min}$) is granted to all states (J) of parameter i. This last feature is essential, in order to cope with the evolution dependencies. The PM selection rule has been used extensively (Davis, 1989; Julstrom, 1995; Tuson and Ross, 1998; Barbosa and Sá, 2000; Whitacre, Pham, and Sarker, 2006).

5.2.4.4 Adaptive Pursuit (AP)

Adaptive pursuit (AP) was introduced as an adaptive selection rule by Thierens (2005). It is based on a decision theory approach in which a winner-takes-all strategy is adopted, as the emphasis is placed on the parameter state with the highest reward. AP uses the same additive relaxation mechanism as PM (equation 5.1). However, the updated probability mechanism is given by:

$$p_{i,j}^{t+1} = \begin{cases} p_{i,j}^{t} + \beta\left(p_{i,\max} - p_{i,j}^{t}\right) & if\ j = j^{*} \\ p_{i,j}^{t} + \beta\left(p_{i,\min} - p_{i,j}^{t}\right) & \forall\, j \neq j^{*} \end{cases} \tag{5.3}$$

where $j^{*} = \arg\max_{j=1,\ldots,J}\left\{\hat{q}_{j}^{t+1}\right\}$ and $p_{i,\max} = 1 - (J-1)\, p_{i,\min}$. The learning rate (β ∈ [0, 1]) is used to control the dominance of the best parameter state with respect to the concurrent states. It has been shown that AP outperforms PM (Thierens, 2005; Fialho, 2011), at least over artificial scenarios.

5.2.4.5 Multi-Armed Bandit (MAB)

The multi-armed bandit (MAB) approach was introduced as an EA adaptive selection rule by DaCosta *et al.* (2008). It allows management of the exploration/exploitation balance among the parameter states. According to this approach, no probability is allocated to any individual

parameter state. In fact, the decision is made by selecting the state that provides the highest expectation. The general form is given by:

$$p_i^{t+1} = \arg\max_{j=1,\ldots,J} \left(\hat{q}_j^{t+1} + C \cdot \sqrt{\frac{2 \cdot \log\left(\sum_{j=1}^{J} n_j\right)}{n_j}} \right) \tag{5.4}$$

where n_j represents the number of times state j was selected from the latest W period. Within this formulation, the first term describes the exploitation aspect, while the second term is intended to promote exploration. The balance between these two aspects is controlled through the scaling factor C.

The first MAB implementation was dynamic MAB (DMAB). However, issues related to DMAB hyperparameter setting motivated the development of MAB variants, like the sliding MAB (SLMAB) and the rank-based MAB (RMAB) (Fialho, 2011). The main difference between RMAB and SLMAB is the way $\hat{q}_j^{t+1}$ is defined. Instead of using raw values from the credit assignment scheme, RMAB ranks them through the area under the curves (AUC) concept or the sum of the ranks (SR). Variants of these two ranking processes have also been introduced: FAUC and FSR. These latter approaches are shown to be invariant to monotonous fitness transformation, compared to AUC and SR. Furthermore, a decay factor (D) is integrated into RMAB to give more influence to top-ranked parameter states.

In all the MAB forms, equation 5.4 is applied only when at least one trial is given to each parameter state. Otherwise, the unselected states are randomly chosen. FAUC-RMAB turns out to be the best MAB variants for adapting genetic operator (Fialho, 2011). FAUC-RMAB has also shown better performance than PM or AP. Finally, it is worth noting that none of the MAB variants succeeded in solving most of the multimodal functions considered. To explain this issue, Fialho (2011, p. 170) concluded: "[…] *the maintenance of some level of diversity in the population should also be accounted somehow for the rewarding of operator applications* […]."

5.2.4.6 Covariance Matrix Adaptation (CMA)

The covariance matrix adaptation approach was developed within the ES framework (CMA-ES) by Hansen and Ostermeier (1996). The purpose of this strategy is to maximize the rate of progress of the search by reproducing the successful evolution path through the adaptation of mutation (i.e. step size and direction). The underlying assumption of CMA-ES is that the mutation step size (σ) must be adapted as fast as possible, whereas the covariance matrix (**C**) of the mutation distribution should be modified more slowly. This is reflected in the use of an additive relaxation mechanism, as defined in equation 5.1, for **C**. That said, σ is defined globally within the population, but applied locally, as it is weighted with respect to the principal axis identified by **C**. This characteristic gives CMA-ES an invariant property with respect to rotational and linear transformation of the search space. CMA-ES was introduced to improve the local search performance of ES, but it was shown in (Hansen and Kern, 2004) that increasing N improves CMA-ES performance in multimodal problems. Based on this observation, G-CMA-ES was proposed in (Auger and Hansen, 2005) with a restart feature that doubles the population size every time it is triggered. This could be seen as a second adaptive mechanism, since the restart decision is made based on five criteria characterizing search performance. Up to now, G-CMA-ES has been considered as the state-of-the-art EA (Garcia *et al*., 2009).

5.2.5 Discussion

Regarding the three parameter setting dependencies, it is clear that, by default, all the adaptive control proposals take into account the problem and evolution dependencies. However, not every method can manage the parameter interaction dependency. In fact, among the previously defined families, only the heuristic rule and FLC are able to consider this aspect, as the other families are parameter-independent approaches. Nevertheless, methods relying on user-defined adaptive formulations are prone to encountering generalization issues about parameter interaction owing to the problem dependency issue.

Furthermore, almost all the adaptive methods rely on some hyperparameters. This would be justifiable if it could be demonstrated that these hyperparameters are less sensitive than the controlled parameters of EAs. However, if the ratio of hyperparameters to controlled parameters (which we refer to here as the H/C ratio) is greater than 1, it is easy to lose track of the intended adaptive goal. In fact, the H/C ratio equals 3 for SLMAB, 4 for PM and RMAB, and 5 for AP and DMAB when hyperparameters related to the adaptation schedule (window size and type of reward conferred) are included.

5.3 Parameter adaptation through Bayesian network

We propose the use of a Bayesian network (BN) as a GA parameter adaptation scheme (which we refer to as BNGA) to more effectively tackle the parameter setting dependencies. The BN is a coherent probabilistic framework taken from the machine learning field, and is a causal network built on Bayes' rule.

With a BN, it is possible to update our belief about a parameter state based on new evidence from the search process. However, as the number of parameters increases, it becomes difficult to perform inference directly from Bayes' rule. To address this problem, BN use a graphical model, which can efficiently represent relationships among parameters, and inference from it is conducted using algorithms developed to take advantage of the graphical representation.

BN has been applied in various domains, particularly in the EA field, it is used in the estimation of distribution algorithms (EDA) search paradigm to carry out probability distribution updates of landscape variables (Larranaga and Lozano, 2002). BN has also been used for tuning GA parameters based on the number of evaluations required to achieve suitable solutions (Pavón, Diaz, and Luzón, 2008), and was later extended, with case base reasoning (CBR), to cope with the inherent limitations of tuning (Pavón *et al*., 2009). To the best of the authors' knowledge, BN has never been used as an adaptive parameter control system.

Construction of a BN for a particular application usually follows two steps. First, learning is conducted to define the best structure for describing the relationship among variables or parameters in the explored context. Then, network data or conditional probability tables (CPT) are learned from that BN structure, with the aim of quantifying the strength of these relationships. In this way, the BN framework tackles parameter interaction directly through the definition of the graphical model, and the problem and evolution dependencies are settled with the help of the CPTs.

5.3.1 Graphical model of BNGA

Learning the BN structure for a specific application can be performed following scoring metrics or search algorithms. However, to avoid incurring too high a computational cost, in this study a graphical model is defined a priori. Consequently, BNGA is based on the simplest BN category, where structure is known and the data are all observable.

The graphical model of BNGA is illustrated in Figure 5.2. The structure is developed for a real-coded SSGA. The decoupling of the structure into two-parameter sets is intended to allow the parameters to act on two different canvases. In fact, as observed in section 5.2.2, two kind of evidence are used in the literature for assessing parameter state productivities: solution improvement and population diversity. Both these indicators should be used to assist in the adaptation, as they have a direct impact on search performance. In BNGA, therefore, parameters related to the creation of new solutions (p_i, $i \in$ 1 to 4) are judged on their ability to improve solutions. In contrast, parameters involved in population governance (p_i, $i \in$ 5 to 9) are judged on their ability to manage the EEB through population diversity. Obviously, parameters from the former set have an impact on the EEB. However, the benefit of the steady-state model is that replacement and update plans constitute a gateway for deciding who will take part of the population.

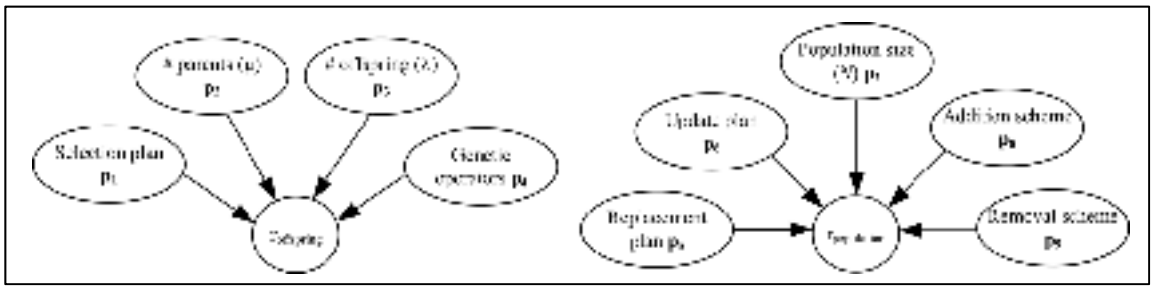


Figure 5.2 Graphical model of BNGA

Regarding the parameters involved, no activation probabilities (p_c or p_m) are considered here as offspring are automatically created through a genetic operator. Furthermore, p_8 and p_9 are added to control the dynamics of population size.

Besides the fact that BNGA is able to encompass all the SSGA parameters, the simplicity of its structures has three advantages (Pearl, 1988). First, the V converging structures, corresponding to singly connected polytrees, allows closed form inference to be performed through a local updating scheme (the message passing algorithm). This is an essential characteristic of BNGA, as it allows the run time impact of the adaptive system on the GA process to be minimized. Second, it is possible to incrementally gather data from this two-level structure, and refine our knowledge about parameter states. As will be shown in the next section, this characteristic constitutes the core of our proposal for building the CPTs. Finally, the decoupling of the parameters into two sets enables the use of belief propagation without any approximation technique, like noisy OR and AND gates. This is because of the small number of parent nodes in each structure, which is important, as no assumption is made about parameter interaction other than the structure by itself. In this way, BNGA is aligned with the statement made by Druzdzel and van der Gaag (2000, p. 483): "*Building a probabilistic network requires a careful trade-off between the desire for a large and rich model on the one hand and the costs of construction, maintenance, and inference on the other hand.*"

5.3.2 Credit assignment schemes

Rewards for p_1 to p_4 ($r_{offspring}$) are based on the fitness improvement of the offspring, while rewards for p_5 to p_9 ($r_{population}$) characterize the search EEB by means of population diversity measures.

Regarding $r_{offspring}$, preliminary experiments have suggested that comparing average fitness of the offspring ($\bar{f}_{offspring}$) to the average fitness of the parents ($\bar{f}_{parent}$) provides the most suitable results (appendix II). Therefore, for minimization problems the reward over the former parameter set is given by:

$$r_{offspring} = 0.5 + 0.5 * \frac{\left(\bar{f}_{parent} - \bar{f}_{offspring}\right)}{\left|\bar{f}_{parent}\right|} \tag{5.5}$$

$$P(r_{offspring} \,|\, p_1, p_2, p_3, p_4) = \begin{cases} r_{offspring} & if \;\; p_{\min} \le r_{offspring} \le 1 - p_{\min}; \\ p_{\min} & elseif \;\; r_{offspring} < p_{\min}; \\ 1 - p_{\min} & otherwise. \end{cases} \tag{5.6}$$

The multiplication by one half in equation 5.5 is accounted for better representing the reward response between -0.5 and 0.5. As a result, we penalize parameter states that are not able to improve parents' solution. This contradicts the mainstream view (see section 5.2.2), but is justified by the fact that applying a bad parameter choice provides relevant knowledge that must not be ranked at the same level as an unused parameter state. Finally, conditions are proposed to ensure that $r_{offspring}$ remains in the [p_{min}, 1-p_{min}] range. We discuss this further at the end of the section.

For $r_{population}$, we rely on both a genotypic and a phenotypic measure to describe the quantity and quality of the population diversity respectively. It has been shown that the EEB orthogonal framework is a generalization of the exploration/exploitation opposing force

concept (see chapter 1). As such, the GDM is responsible for characterizing the exploration axis, and the exploitation axis is represented by a phenotypic convergence measure (PCM). Note that the PCM is the inverse of the normalized PDM. A PCM is used instead of a PDM to comply with the *useful diversity* concept proposed by Goldberg and Richardson (1987), which favors the maintenance of appropriate diversity which potentially brings in good individuals. From this standpoint, both the GDM and the PCM must be maximized. Equations 5.7 and 5.8 formulate the measures employed here for GDM and PCM evaluation, respectively. Even not guaranteeing perfect evaluation in all cases, the descriptors were compared to other available formulations (see chapters 3 and 4), and globally shown significant performance increases.

$$GDM = \frac{\sum_{i=1}^{N-1} \ln\left(1 + \min_{j\in[i+1,N]} \frac{1}{n}\sqrt{\sum_{k=1}^{n}\left(x_{i,k} - x_{j,k}\right)^2}\right)}{NMDF} \tag{5.7}$$

$$PCM = 1 - \frac{\sum_{i=1}^{N-1} \ln\left(1 + \left|f_i - f_{i+1}\right|\right)}{VMD} \tag{5.8}$$

The subscripts used in these formulations are related to the individuals in the population. Therefore, $x_{i,k}$ and $x_{j,k}$ refer to gene k of individuals i and j. The GDM formulation is developed for a real-coded representation, where n corresponds to the dimensionality of the landscape. In the PCM formulation, f_i and f_{i+1} correspond to neighbor fitness taken from the sorted fitness distribution. Finally, *NMDF* and *VMD* are normalization techniques. The former performs normalization with maximum diversity achieved so far, and simply yields the maximum value attained by the numerator since the beginning of the search process. In contrast, the latter performs normalization with virtual maximum diversity. *VMD* is computed through the diversity part, or numerator, of the PCM formula using a virtual distribution of N samples uniformly distributed between f_{worst} and f_{best}. Therefore, *VMD* needs to be updated when population size and/or absolute fitness range change.

From there, we need to determine how these two measures can be combined to produce $r_{population}$. A similar framework was applied in (Maturana and Saubion, 2008), where it was proposed that genetic operators be rewarded based on their ability to bring the EEB closer to 45° (same amount of exploration and exploitation). A more commonly adopted EEB management scheme involves promoting full exploration at the beginning of the evolution process (EEB→90°), and moving this balance toward full exploitation as the process evolves (EEB→0°) (Eiben, Hinterding, and Michalewicz, 1999). However, these EEB management schemes are biased, and they do not convincingly provide an optimal search path over any problem. This issue was summarized by Burke, Gustafson, and Kendall (2004, p. 48-49), as follows: "*The type and amount of diversity required at different evolutionary times remains rather unclear.*" Consequently, we decided to give as much flexibility as possible to EEB management by encouraging parameter states that maximize both exploration and exploitation (i.e. GDM→1 and PCM→1). This is formalized in the following equation:

$$r_{population} = \frac{\sqrt{GDM^2 + PCM^2}}{\sqrt{2}} \tag{5.9}$$

$$P(r_{population} \mid p_5, p_6, p_7, p_8, p_9) = \begin{cases} r_{population} & if \quad \left(p_{\min} \le r_{population} \le 1 - p_{\min}\right) \\ & \quad \wedge \left(GDM > p_{\min} \wedge PCM < 1 - p_{\min}\right); \\ p_{\min} & elseif \quad \left(r_{population} < p_{\min}\right) \\ & \quad \vee \left(GDM < p_{\min} \wedge PCM > 1 - p_{\min}\right); \\ 1 - p_{\min} & otherwise. \end{cases} \tag{5.10}$$

A condition is added from equation 5.10 which penalizes parameter states that produce convergence (i.e. GDM < p_{min} and PCM > 1-p_{min}). This condition is incorporated to move away from the premature convergence state without using restart feature.

Finally, as observed in equations 5.6 and 5.10, a minimum probability (p_{min}) is used to constrain the rewards. The meaning of p_{min} in this case is similar to that involved in the PM

and AP adaptive systems, as it ensures that parameter states ($p_{i,j}$) will not be lost during the search process. However, the role of p_{min} is slightly different in BNGA, in that certainty states (i.e. $p_{i,j}$ = 0.0 or $p_{i,j}$ = 1.0) are prohibited by avoiding the certainty of conditional probabilities. However, $p_{i,j}$ can approach these values without restriction. In addition, this hyperparameter has a miscellaneous function for each parameter set. For $r_{offspring}$, p_{min} allows the reward value to be bounded in a probabilistic range. In contrast, for $r_{population}$, p_{min} is involved in the definition of a hazardous area and is used to penalize parameter states that navigate into it.

5.3.3 Conditional Probability Table (CPT)

To ensure that the same evidence is not counted multiple times, we must exercise caution in the use of the probability update from generation t (BEL(p_i^t)) as a prior probability of generation t+1 (p_i^{t+1}). It is fundamentally important to bring up this point, since this probability updating scheme is the core adaptive mechanism used by BNGA. It is managed by building the conditional probability tables (CPTs) represented by P($r_{offspring}|p_1, p_2, p_3, p_4$) and P($r_{population}|p_5, p_6, p_7, p_8, p_9$).

In BNGA, CPTs are built from a kind of Bayesian estimation. At each generation, a uniform probability is assigned to all the parameter configurations in the CPTs (P($r|p_i$) = 0.5) and the reward ($r_{offspring}$ or $r_{population}$) is subsequently used to update the parameter configuration selected ($j = s$) in the CPT (P($r|p_{i,s}$) = r). This CPT building scheme ensures that knowledge gathered through the reward is considered only once, as its impact is transferred to the parameter state probabilities by setting the next generation, t+1, prior probabilities equal to the posterior probabilities of the previous generation, t. This approach is valid because, as mentioned above, the CPTs are built from scratch at each generation.

Finally, we note that the size of the CPTs increases exponentially with the number of parameters involved. So, a mechanism is added to increase the impact of the single configuration reward (P($r|p_{i,s}$)) in the CPTs. This is achieved by allocating a probability equal

to 1-P($r|p_{i,s}$) to all the configurations that do not involve parameter states included in the selected parameter configuration (i.e. $\forall$ P($r|p_{i,j}$) = 1-P($r|p_{i,s}$), where $j \neq s$). The assumption underlying this reinforcement mechanism is that parameter state performing well over a given generation makes other choices undesirable and vice-versa. In fact, this reinforcement mechanism shows to increase the rate of adaptation.

5.3.4 BNGA process

In summary, the adaptive process of BNGA is as follows: At the beginning of the search over a given problem, all the parameter state probabilities are set uniformly ($p_{i,j} = 1/J$). This is because no knowledge is available to bias, one way or the other, our belief in particular states (problem dependency). At each generation, the reward is computed ($r_{offspring}$ and $r_{population}$) from the feedback indicators chosen, and allocated to the selected parameter configuration. The posterior probabilities BEL(p_i^t) are defined according to the message passing algorithm (Pearl, 1988) given by:

$$\lambda_r(p_i) = \alpha \sum_{p_k : k \neq i} P(r|p_m, ..., p_n) \prod_{k \neq i} \pi_r(p_k) \tag{5.11}$$

$$BEL(p_i) = \alpha \cdot \lambda_r(p_i) \cdot \pi_r(p_i) \tag{5.12}$$

where $\lambda_r(p_i)$ is the message passed by the child node ($r_{offspring}$ or $r_{population}$) to the parent nodes, and $\pi_r(p_i)$ is the message passed by the parent nodes p_i to the child node. Since the nodes p_i do not have a parent, $\pi_r(p_i)$ is exactly equal to their prior probabilities. α is a normalization factor based on the summation of all p_i states. In equation 5.11, m = 1 and n = 4 for the structure converging toward $r_{offspring}$, and m = 5 and n = 9 for the structure involving $r_{population}$. Therefore, parameters p_m to p_n are conditionally dependent on knowing r (interaction dependency). Finally, BEL(p_i^t) are used as prior probabilities for the next generation (evolution dependency). Readers are invited to consult appendix III (Algorithm A III-1) for more details about BNGA and the inference process used.

Consequently, BNGA allows us to control all the SSGA parameters, while at the same time taking into account the three dependencies. Furthermore, no additive relaxation or fading mechanism is needed by BNGA, as the memory of past knowledge is automatically managed through modification of the prior probabilities at each generation. Finally, an adaptation schedule is not needed for BNGA, aside from the time reduction allocated to inference, as it takes advantage of new evidence as soon as it is released.

5.4 Comparative study

We conducted a two steps comparative study to assess the performance of BNGA. First, we evaluated BNGA by means of an SSGA, looking at static parameter setting, a naïve adaptive method, PM, AP, and FAUC-RMAB. For static parameter setting, the parameters were defined from the most promising states identified by BNGA for each problem considered (see Table A IV-4). Survey of the selected parameter states from the other adaptive methods is also provided in appendix IV. For the naïve approach, the parameter states were defined with uniform probabilities throughout the course of the evolution. For the other approaches, a hyperparameter study was performed prior to the comparison to determine the impact of these parameters and select the best configuration for them.

Then, we compared the best parameter setting approach found in the first step to CMA-ES and G-CMA-ES in their default settings[3]. We propose this two-step approach, as these latter methods follow different EA paradigms with different genetic operators, and so they may overshadow the objectives of the analysis over the SSGA framework.

5.4.1 Methodology

The comparison is performed based on the 25 functions (F1 to F25) defined in the CEC'05 benchmark (Suganthan *et al*., 2005) (see appendix I for an overview). We only conduct the

[3] CMA-ES version 3.51.beta was used to conduct this analysis. It can be accessed at http://www.lri.fr/~hansen/cmaes.m

10-dimensional experiments here (n = 10), but rigorously adopt the methodology provided by the CEC'05 benchmark. Twenty-five runs are conducted for each problem and the search is stopped if the 100 000 function evaluation limit is reached, or if the optimizer is able to locate the global optimum within the prescribed tolerance of 10^{-8}.

5.4.2 Parameter states involved

The SSGA process is summarized as follows: λ offspring are created at each generation by using a specific crossover operator over the set of μ selected parents; following a chosen replacement plan, r individuals are removed from the population to make room for the new individuals; subsequently, an update plan is used to reinsert individuals from the temporary pool into the next generation.

Table 5.1 summarizes the parameters considered in this study. In terms of the selection plan, we propose five alternative states: 1) the best individual is selected as the parent, and the remaining individuals are chosen randomly; 2) one individual is selected randomly from among the best individuals in ten groups in the fitness range, and the remaining parents (μ-1) are picked at random from the population (Sinha, Tiwari, and Deb, 2005); 3) parents are randomly selected from the population; 4) parents are selected from a binary tournament; and 5) parents are selected using FUSS (Hutter and Legg, 2006). For the genetic operator, only crossovers are covered here, as they all have the ability to introduce new genetic material. Again, five potential states are taken into account: 1) a parent-centric crossover (PCX) with

Table 5.1 Parameters involved and their potential states

	Parameter i								
state j	1 Selection plan	2 Number of parents (μ)	3 Number of offspring (λ)	4 Genetic operator	5 Replacement plan (r)	6 Update plan	7 Population size (N)	8 Addition scheme	9 Removal scheme
1	Best + random	2*	2	PCX	Fatoro	Best	50	Random	Worst
2	Group-10		5	UNDX	1 random	BT1	100	Child+random	Tournament-2
3	Random		10	SPX	2 random	BT2	300		Tournament-5
4	Tournament-2			BLX-0.5	5 random		500		Tournament-10
5	FUSS			PNX					
P_{ij}	0.2	1.0	0.33	0.2	0.25	0.33	0.25	0.5	0.25

$\sigma_\eta = \sigma_\zeta = 0.1$ (Deb, Anand, and Joshi, 2002); 2) a unimodal normal distribution crossover (UNDX) with $\sigma_\xi = 1/\sqrt{\mu-2}$ and $\sigma_\eta = 0.35/\sqrt{n-\mu-2}$ (Ono and Kobayashi, 1997); 3) a uniform crossover (UX) with the addition of normally distributed noise $N(0,0.01)$; 4) a blended crossover (BLX-0.5); and 5) a parent-centric normal crossover (PNX) with $\eta=1.0$ (Ballester and Carter, 2004). Four choices are provided for the replacement plan: 1) parents are added to the temporary pool, and 2) to 4) $x \in \{1, 2, 5\}$ individuals are randomly selected from the population. For the update plan, three strategies are proposed: 1) the best individuals are reinserted into the population; 2) the best individual is selected and the remaining individuals are picked from binary tournaments (BT1); and 3) the same as BT1, except that a diversity control mechanism is considered, since the selected individuals replace the nearest genotypic individual already in the temporary pool (BT2). Finally, the addition and removal schemes are required to manage the dynamics of population size. Two addition methods are considered when N increases: 1) randomly created individuals are added to the population; and 2) all the non-inserted offspring of the current generation are added to the population, and the remaining required individuals are randomly created. Four removal methods are considered when N decreases: 1) the worst individuals are removed from the population; and 2) to 4) a tournament is made up of $x \in \{2, 5, 10\}$ competitors, and the worst is removed from the population.

All the SSGA adaptive control systems (naïve, PM, AP, FAUC-RMAB, and BNGA) use these parameter states. At the beginning of each run, the probability of application of each parameter state ($p_{i,j}$) is set uniformly, as presented in the last row of Table 5.1. The adaptive control of those systems, except for the naïve approach, is based on the credit assignment schemes described in section 5.3.2. For PM, AP, and FAUC-RMAB, the parameters are controlled independently, since these systems do not manage parameter interaction. In contrast, BNGA integrates parameter interaction, as depicted in Figure 5.2.

5.4.3 Hyperparameter sensitivity analysis

To compare the adaptive systems fairly, a sensitivity analysis is conducted over the hyper-parameter setting as a prerequisite. Table 5.2 summarizes the hyperparameter values considered for each adaptive system. Note that for decoupled parameter control systems (PM, AP, FAUC-RMAB), the same hyperparameters are used for each parameter, even though different optimal settings are likely. This choice is made to maintain a suitable H/C ratio. Regarding p_{min}, the $1/2J$ option comes from Thierens' proposal, which is to apply, at most, the best parameter state half the time (Thierens, 2005).

Table 5.2 Hyperparameter values considered for the sensitivity analysis

Adaptive control system	H/C ratio	Hyper-parameter							Total combination (TC)
		p_{min}	α	β	C	D	Reward type (RWD)	Window size (W)	
PM	0.44	{1/(2J), 0.01}	{0.1, 0.3, 0.5, 0.7, 0.9}				{0.instantaneous, 1.mean, 2.myrmax}	{10, 100, 250}	60
AP	0.55	{1/(2J), 0.01}	{0.1, 0.3, 0.5, 0.7, 0.9}	{0.1, 0.3, 0.5, 0.7, 0.9}			{0.instantaneous, 1.mean, 2.myrmax}	{10, 100, 250}	350
FAUC-RMAB	0.33				{0.01, 0.1, 0.5, 1, 5, 10}	{0.25, 0.4, 0.5, 0.75, 0.9, 1}		{10, 50, 100, 250}	144
BNGA	0.11	{0.001, 0.005, 0.01, 0.03, 0.05, 0.1, 0.15, 0.2}							8

The sensitivity analysis is made up of 5 repetitions from F1 to F10 by considering all the hyperparameter combinations. Since the goal is to find robust hyperparameters over various problems, as well as configurations that provide high quality solutions, we suggest a multi-objective approach. Both objectives are evaluated by the equation:

$$obj_i = \sum_{j=1}^{10} 1 - \frac{\min_{k \in [1,\, TC]} f_{k,j}}{f_{i,j}} \tag{5.13}$$

In the case of the first objective, f represents the mean best fitness error (MBFE), and for the second objective, f corresponds to the best fitness error (BFE). Both are defined from all the repetitions over the evaluated problem j. Subscript i stands for the hyperparameter

configuration studied, which ranges from 1 to the total number of combination (TC). Using this normalized formulation, responses can be aggregated, regardless of the complexity of the test functions. Consequently, each hyperparameter configuration is assigned a result ranging from 0 to 10.

Figures 5.3 and 5.4 present the results for the four adaptive systems. For PM, the configurations with instantaneous reward application (RWD) and p_{min} = 0.01 stand out relative to all the other hyper-parameter choices. Of these, three are non-dominated configurations (α = {0.5, 0.7, 0.9}). To perform the comparative study, PM with: p_{min} = 0.01, α = 0.9, and RWD = instantaneous (MBFE = 4.4, BFE = 4.6), is selected. For AP, no hyper-parameter family emerges. However, seven configurations present non-dominated behavior. Of these, AP with p_{min} = 0.01, α = 0.3, β = 0.7, and RWD = instantaneous (MBFE = 5.8, BFE = 6.7), is favored.

For FAUC-RMAB, favored configurations suggest large window size, but only four configurations are found to be non-dominated. Of these, FAUC-RMAB with C = 0.5, D = 0.5, and W = 250 (MBFE = 5.3, BFE = 5.9), is favored. Finally, the performance of BNGA shows improvement with a decreasing p_{min} value. From that result, we decided to adopt BNGA with p_{min} = 0.01 (MBFE = 1.9, BFE = 3.5), as it is one of the non-dominated configurations and is more apt to reintroduce configurations that had almost been forgotten.

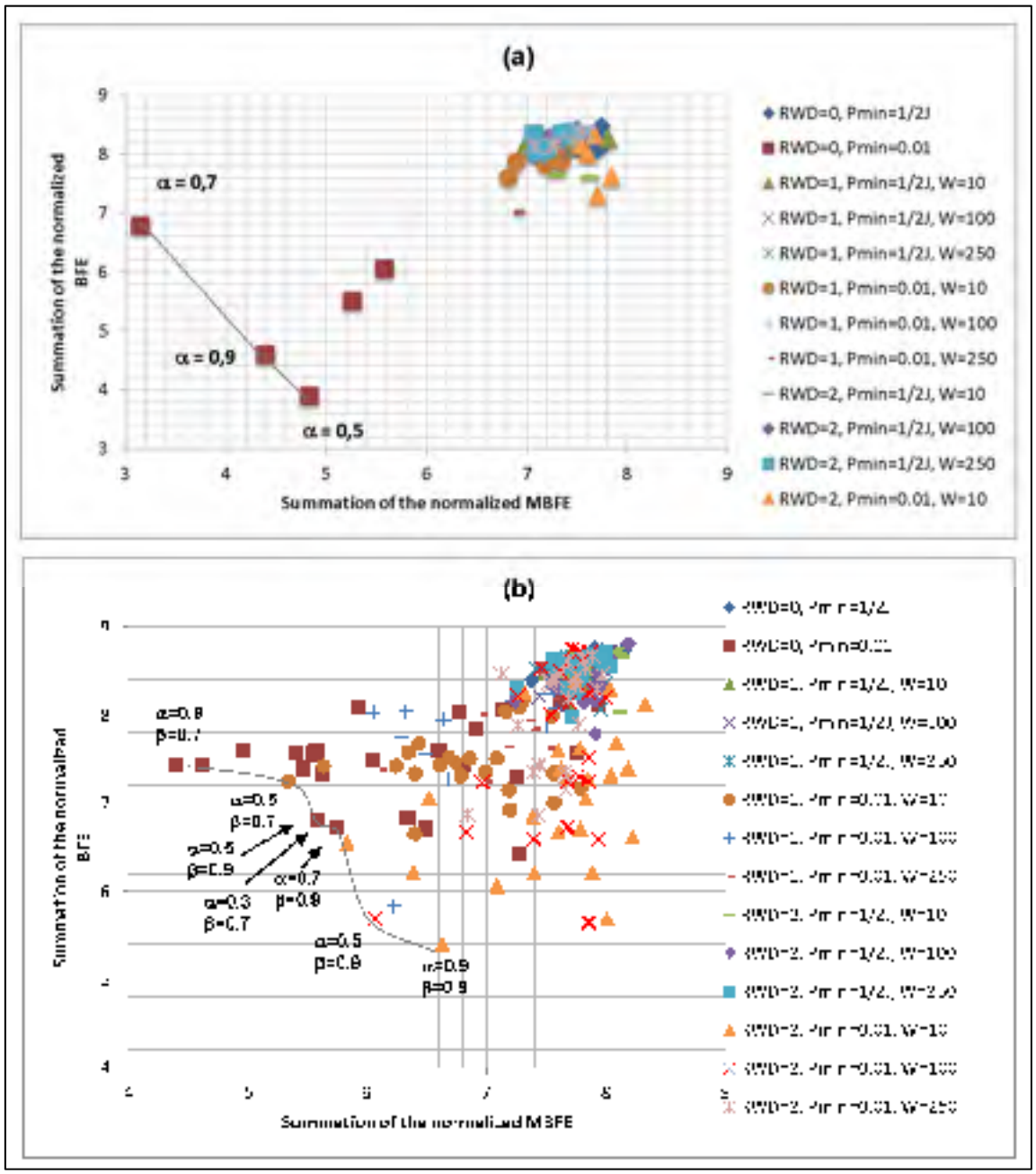


Figure 5.3 Results of the hyperparameter sensitivity study: a) PM, b) AP

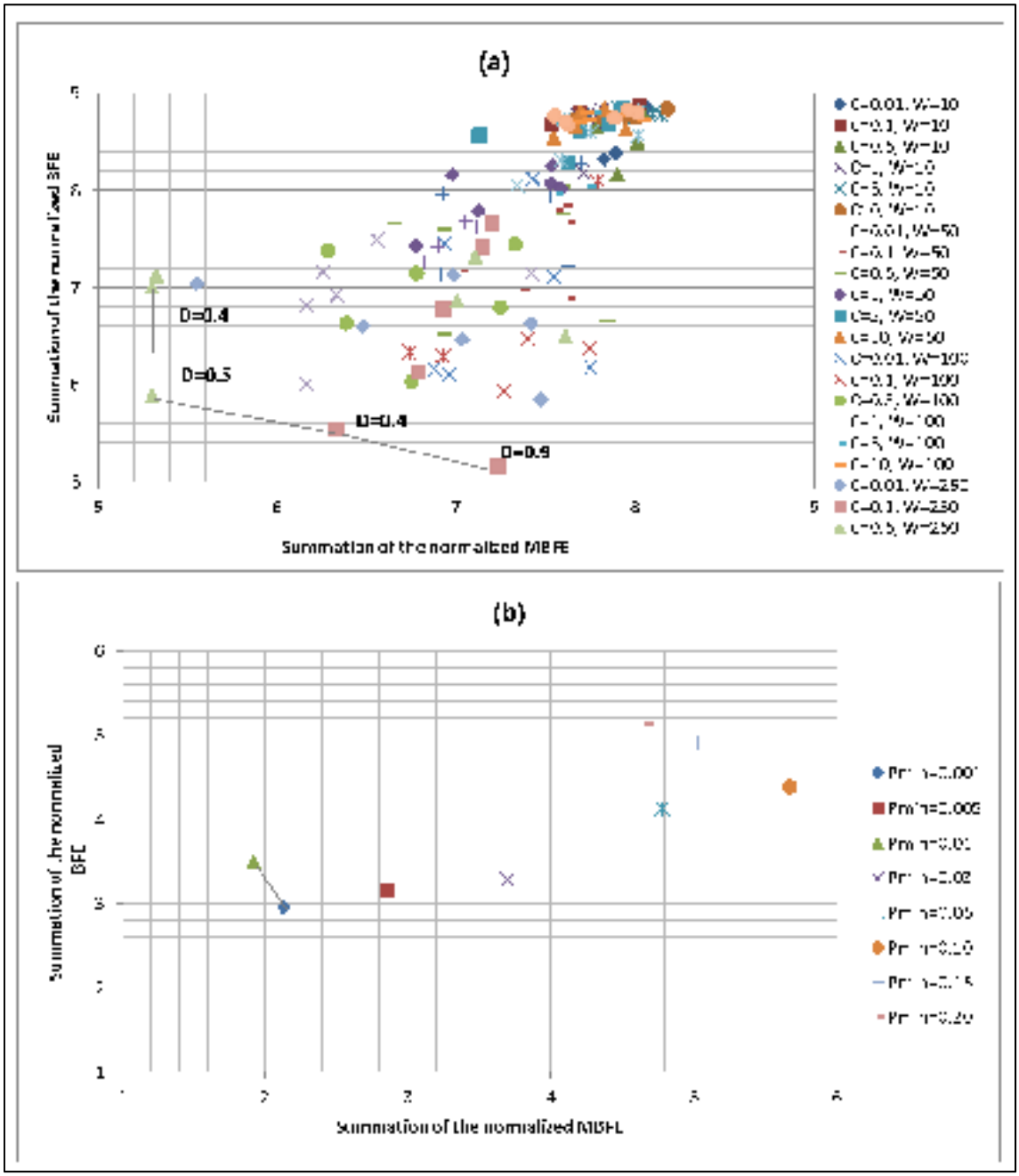


Figure 5.4 Results of the hyperparameter sensitivity study: a) FAUC-RMAB, b) BNGA

5.5 Results

To analyze the results of the parameter setting approaches we compared, the statistical procedure proposed by García *et al.* (2009) is followed, according to which Friedman's non parametric test is used to reveal any significant performance difference from at least one approach. Then, a post hoc test, following Hochberg's procedure, is applied to identify any concrete difference between each parameter setting approach and a control algorithm. This control algorithm is represented by the best ranked approach over the sample considered. For each approach, the sample is built from the MBFE characteristics over a predefined group of functions. As proposed in (Garcia *et al.*, 2009), three groups are examined; F1 to F14, F15 to F25, and F1 to F25. This statistical procedure is also applied for each test function independently, where BFEs from all the repetitions serve as a sample. This addition is included to increase the confidence level on the conclusions arising from the statistical comparison.

Table 5.3 presents the format used to display the results of each parameter setting approach over each individual test function. The MBFE with respect to the global optimum is recorded with its standard deviation (STD) and the best fitness error (BFE) achieved over the 25 repetitions. Furthermore, the success rate (SR), as well as the success performance (SP), are registered. The former is defined as the percentage of runs performing better than the prescribed tolerance, whereas the latter describes the mean required number of function evaluations for the successful runs divided by SR. SP is intended to compare, on the same basis, algorithms with a low probability of success but a high convergence rate, and vice-versa. The last entry corresponds to the statistical test p-values. This information is used for both individual test functions and groups of functions. Shaded p-values indicate Friedman's

Table 5.3 Results format

	[illegible]	
F_j	MBFE STD	SR SP
(tolerance)	[illegible]	p-value

test results, and are located in the control algorithms column. The other p-values represent Hochberg's post hoc test. An underlined p-value indicates that the performance of a given approach is worse than that of the control algorithm, following a 0.05 significance level. Finally, bold entries designate the best values achieved for a given test function.

5.5.1 Results of the SSGA parameter setting approaches

Table 5.4 presents the results of the six parameter setting approaches compared within the SSGA framework. Overall, the statistical results show that the static parameter setting approach is outperformed by other approaches over 44% of the test functions (number of underlined p-value / number of functions). This value increases to 48%, 64%, and 68% for the naïve approach, PM, and AP respectively. Regarding FAUC-RMAB, poorer performances were observed over 40% of the test functions. BNGA is thus the most robust approach in this comparison, as only 16% of the test functions demonstrated weaker performances compared to the other parameter setting approaches. This result confirms the importance of considering parameter interaction dependency.

Regarding the unimodal or low multimodal test functions (F1 to F6), our results indicate that the static parameter setting approach performs the best. In fact, BNGA shows comparable results over those problems. However, it can be observed from the SP measures that BNGA converges more slowly to the optimum. This behavior is expected, as BNGA, like any other adaptive parameter control method, needs to identify the most promising parameter states, whereas the static approach takes advantage of a kind of off-line tuning. It is interesting to note, however, that the high conditioning characteristic of F3 poses great difficulty for all adaptive parameter control systems.

With the multimodal problems (F7 to F25), the general trend of the behavior of the approaches compared is less obvious. In fact, FAUC-RMAB and BNGA turn out to be the best approach, or the control algorithm, over most of these test functions. Nevertheless, except for F12, none of the approaches was able to find the global optimum within the

prescribed tolerance. For the most complex subset of multimodal problems, which are the hybrid composition functions (F15 to F25), it is interesting to note that the naïve approach is ranked among the best. These observations suggest that over highly multimodal problems, SSGA adaptive systems are no better than a pure exploration strategy for selecting parameter states. However, investigation of the population EEB reached throughout the simulations indicates that none of the compared approaches was able to allocate resources following a useful diversity pattern (GDM → 1.0 and PCM → 1.0) (appendix V). This lack of valuable EEB states suggests a missing complementarity and/or synergy among the parameter states involved. This introduces the exploration/exploitation dilemma, but this time, at the parameter level. Too many parameter states could hide their performances, as they are not sufficiently involved in the search. At the same time, without enough states, the probability of finding robust and efficient parameter configurations is reduced. As a result, a kind of balance is advised here too. However, investigation of this aspect is not considered, since it is beyond the scope of this thesis.

Table 5.5 presents the performance of the SSGA parameter setting approaches from the perspective of the groups of functions sampled. The outcomes of these statistical tests are similar to those obtained from the independent functions sampled. This allows us to confirm that the static parameter setting approach is as good as the best adaptive control systems over the most simple test functions (F1-F14), whereas the naïve approach is competitive over complex test functions (F15-F25). This group perspective does not, however, show the difference between FAUC-RMAB and BNGA performance, if any, although it reveals that, over all the problems considered (F1-F25), these two stand out as the best SSGA parameter setting approaches.

Table 5.4 Results of the SSGA parameter setting approaches over the CEC'05 benchmark

Table 5.5 Statistical comparison by function group

	STATIC	NAIVE	PM	AP	FAUC-RMAB	BNGA
F1-F14	0.480	0.002	0.014	0.079	0.259	0.003
F15-F25	<1E-3	0.610	0.001	0.004	0.909	<1E-3
F1-F25	0.001	0.002	<1E-3	0.001	0.226	<1E-3

5.5.2 Results of the EA parameter setting approaches

The FAUC-RMAB and BNGA results are now compared to state-of-the-art EAs (CMA-ES and G-CMA-ES), as they have both demonstrated statistically the highest overall performance over F1 to F25 within the SSGA comparison framework. Furthermore, a variant of BNGA has been added to this comparison ($BNGA_1$), with the aim of leveraging knowledge gained, over a given problem, as the run number increases. In fact, the difference between the BNGA variants is that $BNGA_1$ uses the average probabilities of the parameter states computed from the previous runs at beginning of each run, instead of uniform prior as is the case for $BNGA_0$ (section 5.5.1). Note that other knowledge integration mechanisms were also investigated in this research (appendix VI). However, they are not included here, as no additional performance benefit was observed.

Table 5.6 presents the results of this comparison. Overall, it shows that FAUC-RMAB is statistically outperformed by other approaches over 60% of the test functions, compared to 36% for $BNGA_0$, 72% for $BNGA_1$, 40% for CMA-ES, and 20% for G-CMA-ES. This confirms the robustness of G-CMA-ES and ranks $BNGA_0$ second.

It is worth mentioning that the recorded performance of G-CMA-ES is different from that provided for functions F9 to F11 by Auger and Hansen (2005). This discrepancy can be explained by different initialization seed numbers and the stochastic nature of EAs.

As for the SSGA parameter setting approaches, CMA-ES and G-CMA-ES are not able to allocate resources in accordance with the useful diversity concept (appendix V). This may

explain, at least partially, their difficulty in reaching the global optimum over highly multimodal test functions.

At the same time, compared to $BNGA_0$, $BNGA_1$ is effective in terms of increasing the quality of the solutions or the convergence rate over test functions F1 to F7 only. Over more complex problems, the integrated knowledge mechanism of $BNGA_1$ generally results in performance deterioration. This mechanism is therefore not recommended in its current form.

Table 5.7 presents the statistical comparison from the point of view of groups of functions. FAUC-RMAB and $BNGA_1$ are the only adaptive control systems dominated by other approaches on F1 to F14 and F1 to F25 respectively. It is interesting to note that, in contrast, G-CMA-ES is not statistically better than CMA-ES or $BNGA_0$, whatever group of functions is considered. This is somewhat surprising, as G-CMA-ES was introduced specially to improve CMA-ES performance over multimodal problems. However, the lack of performance enhancement may related to the strong asymmetry of the hybrid composition functions (Hansen and Kern, 2004) and the maximum number of prescribed function evaluations preventing the identification of an effective λ value (Lunacek and Whitley, 2006). Finally, the fact that $BNGA_0$ is statistically equivalent to G-CMA-ES is encouraging, as it suggests that SSGA with an appropriate parameter setting approach may be viewed as a competitive EA strategy.

Table 5.6 Results of the EA parameter setting approaches over the CEC'05 benchmark

	[illegible]		[illegible]		[illegible]		[illegible]		[illegible]	
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Table 5.7 Statistical comparison by function group

	FAUC-RMAB	$BNGA_0$	$BNGA_t$	CMA-ES	G-CMA-ES
F1-F14	0.005	3.094	0.112	0.633	0.008
F15-F25	1.000	2.086	0.061	0.842	1.000
F1-F25	0.119	3.421	0.029	0.474	0.061

5.6 Concluding discussion

Parameter setting is a longstanding issue in the EA field. On the one hand, specific parameters drive the performance of the optimizer, as they are responsible for the search path followed. On the other, they are very difficult to determine, as optimal parameter states are problem-dependent and evolution-dependent, and they take part in complex interactions.

The objective of this chapter was to propose a new adaptive system capable of coping with all the parameter setting dependencies, in order to achieve search performance improvement. The adaptive approach developed is based on Bayesian network and is applied to genetic algorithms (BNGA). This strategy addresses all three dependencies: 1) problem dependency, by ensuring uniform parameter state probabilities at the beginning of the search; 2) evolution dependency, by updating parameter state probabilities as new evidence appears; and 3) parameter interaction, by supporting relationships through the graphical model. Compared with the other available adaptive parameter setting approaches, BNGA is the only one capable of acknowledging parameter interaction, while autonomously managing the strength of the interactions as the search evolves. Furthermore, BNGA requests only one hyper-parameter, compared to as many as five for comparable systems. These characteristics are clearly promising, in terms of enhancing search robustness and efficiency.

A comparative study was then initiated following the CEC'05 benchmark. The first step was to compare BNGA to static parameter setting, a naïve approach, and three recognized adaptive methods: PM, AP, and FAUC-RMAB. This comparison was performed within an SSGA framework involving the control of 9 parameters. All the adaptive control approaches were based on the same credit assignment scheme. This means that four parameters related to

offspring creation were judged on their ability to improve fitness, and five parameters related to population governance were evaluated on their ability to manage the EEB.

In summary, results from the statistical test suggest three interesting conclusions. First, management of parameter interaction is now recognized as valuable over some test functions, which has led to BNGA being awarded the highest rank in terms of robustness. Second, FAUC-RMAB and BNGA stand out as the best SSGA adaptive approaches. Their performances were similar to those of static parameter setting over unimodal test functions, and statistically equivalent to the naïve approach over complex multimodal problems. Considering all the test functions, they were superior to all the parameter setting approaches compared. These observations are very interesting, in that they clearly attest to the importance of the role that an adaptive control system must perform. Indeed, adaptive systems must be proficient in matching exploitative strategy or explorative strategy when they are best suited. Finally, the hyperparameter sensitivity study confirms the implementation advantage of BNGA over other adaptive systems, like FAUC-RMAB. In fact, the minimal number of hyperparameters involved in BNGA gives it the greatest generalization power of all the adaptive systems.

We further compared BNGA and FAUC-RMAB to state-of-the-art EAs: CMA-ES and G-CMA-ES. Summarizing, our results demonstrate that, statistically, BNGA performs as well as CMA-ES and G-CMA-ES. From a practitioner's standpoint, this general conclusion is attractive, as it indicates that SSGA with appropriate parameter setting is highly competitive, while offering a more flexible framework than CMA-ES for customizing search components.

This study also reveals that none of the approaches considered is able to reach the global optimum over complex multimodal problems. We can conjecture that this may be related to a lack of useful diversity in the search process, based on our investigation of the population EEB, and, consequently, a lack of complementarity and/or synergy among parameter states. Determination of the best combination of parameter states involved may result in marked performance gain.

Finally, it is important to mention that the proposed adaptive approach is not limited to GA applications, but is rather a generic control system that can be effective for any population-based search algorithm.

CONCLUSION

One of the greatest challenges to using evolutionary algorithms (EA) is to adequately set their parameters. EA parameters are problem-dependent, they are prone to change during the course of the evolution, and they interact in complex ways. Despite these difficulties, appropriately setting them is critical. As explained at the beginning of this thesis, they directly impact the performance of the process by driving the search path pursued. We believe that an adaptive parameter control strategy is the best choice, since knowledge from the search can be included, and so the potential of parameter setting can be maximized. In order to achieve this, we envisage integrating the exploration/exploitation balance (EEB) concept into the feedback indicator by evaluating population diversity. However, there is no consensus in the evolutionary computation field about the best way to model the EEB concept and which diversity measurements to use: a genotypic measure, a phenotypic measure, or both. We address these issues directly in this thesis, as its objective is to develop an adaptive parameter setting approach to handle all the EA parameters at once, which has not been achieved to date.

Our initial working hypothesis was that the search EEB, which serves as a feedback indicator, should be represented by means of the orthogonal framework, where genotypic measurement is used to express the exploration axis, and phenotypic measurement is adopted to characterize the exploitation axis. Our discussion in chapter 1 helped to support this assumption by illustrating that the orthogonal framework is a generalization of the *opposing force* concept, as it allows us to represent any type of search space structure. That discussion was extended in section 3.2, because the orthogonal framework is the only EEB representation that can justify the use of a phenotypic formulation. In addition, that discussion highlights the fact that the orthogonal EEB framework, unlike the opposing force concept, can depict the *useful diversity* notion introduced by Goldberg and Richardson (1987), as both exploration and exploitation can be maximized. Finally, section 3.2 also provides the motivation for using a genotypic diversity measure (GDM) for the exploration axis and a phenotypic convergence measure (PCM) for the exploitation axis. These choices

were put forward considering that the GDM provides information about the scattering of the population over the search space, while the PCM describes the fitness distribution used to target promising regions. Nonetheless, numerous formulations have been proposed in the literature to describe the GDM and the PCM without comprehensive study of their behavior.

From there, we set about answering the questions posed in this thesis:

1. What is the best genotypic formulation for estimating the exploration provided by the search process?
2. What is the best phenotypic formulation for outlining the exploitation supplied by the search process?
3. How can the exploration and exploitation knowledge for adapting the EA parameters be converted into a reward?
4. Which adaptive system can best handle the parameter setting dependencies?
5. How can the performance of an adaptive parameter control proposal be assessed?

Regarding the first question, a thorough investigation was initiated in chapter 1 aimed at identifying the most suitable genotypic formulation for the real-coded representation context. The comparison involves 15 different formulations representing two broad families; distance-based measures, and gene frequency measures. The behavior of the measures is evaluated over a specifically designed benchmark by means of three quality criteria, which refer to their repeatability, their robustness with respect to the simulation parameters, and their ability to consider outliers within the population. Results show that the mean pairwise measurement (D^N_{PW}) dominates all the genotypic formulations considered, but, like the other descriptors, it is unreliable for reflecting diversity over convergence processes with multiple sites. That raises the question of whether or not the established criteria are sufficient for judging the performance of GDMs. Furthermore, the ability of the available measures to truly reflect population diversity is challenged, and, among other things, the non special treatment they afforded to duplicate individuals is identified as a potential issue.

These results point to the need to develop a GDM validation framework. This step is addressed in chapter 2 using three diversity requirements extracted from the literature, but adapted to the present context. The requirements are the following: monotonicity in individual varieties, twinning, and monotonicity in distance. They enable the propensity of GDMs to be judged in order to monitor population diversity, or, more fundamentally, to determine whether or not the genotypic formulations act as coverage space indicators. The three diversity requirements are evaluated by means of a reduced population arrangement framework and seven cases of controlled population diversity. Four representative GDMs are involved in this experiment, the mean pairwise measurement (D_{PW}) that characterizes most common distance-based GDMs, the Shannon entropy (GF_S) that describes gene frequency GDMs, the hypervolume distance measure (D_L) that is considered from other studies as probably the best way to represent population diversity, in spite of its prohibitive computational cost, and, finally, the minimum spanning tree measure (D_{MST}) that is intended to approximate D_L behavior efficiently. The validation framework has made it possible to conclude that none of the genotypic formulations tested can be considered as a proper way to measure population diversity. Overall, it is difficult for D_{PW}, GF_S, and D_{MST} to respect any of the established diversity requirements. We know this because of their failure to identify a population that is distributed uniformly across the landscape as being in a maximal diversity state. In addition, they all have trouble managing duplicate individuals, and they cannot efficiently account for the relative distance between individuals in a population. At the same time, our gene frequency measurement results provide us with strong evidence that generalization of a univariate measurement into a multivariate framework can generate inaccurate evaluations. In contrast, D_L meets the requirement of monotonicity in individual varieties, as well as the twinning requirement. However, it suffers from the same issue concerning the description of the relative distance between individuals, which leads to violation of the monotonicity in distance requirement. These findings support the commonly held belief that the way to estimate genotypic diversity is not obvious. Although this chapter does not resolve question #1 as expected, the proposed validation framework lays the foundation for the potential development of any new genotypic formulation. Consequently, it illustrates the fact that genotypic diversity evaluation deserves more attention.

In chapter 3, we turn our attention to the second question, which involves identifying the best phenotypic formulation for characterizing the exploitation provided by the search process. This change of direction was motivated by our idea that resolving a relatively simpler problem might provide insight into the way to tackle the previous question, which has not yet been resolved. In fact, the phenotypic formulation, or PCM, should be more straightforward to study, since it corresponds to a univariate diversity measure instead of a multivariate one, like GDM. To perform the analysis, 19 phenotypic formulations retrieved from the literature are reviewed. They are then evaluated over a validation framework involving six controlled fitness distribution cases to assess the diversity requirements reformulated within the phenotypic context. These requirements are the following: monotonicity in fitness varieties, twinning, and monotonicity in distance. The cases proved to be sufficient to highlight the weaknesses of these formulations, as they all failed to meet two of the three requirements. These results led to the development of a new formulation (PCM_{13}), based on the summation of the logarithm of the fitness difference between neighbors. Validation of PCM_{13} shows that this formulation complies with all three requirements. To strengthen these conclusions, all the PCMs are compared over three specifically designed fitness landscapes. The same landscapes also serve as a platform to illustrate that PCM_{13} is reliable over similarly scattered fitness distributions and provides good robustness with respect to population size. Finally, we investigate the influence of outliers, our results suggesting that the PCM_{13} formulation provides an effective way to account for their presence, even when they greatly disturb the fitness range.

Encouraged by this achievement, we generalize a multivariate diversity measure (D_{LN}) from PCM_{13} to serve as a GDM. The fundamental difference between D_{LN} and PCM_{13} is the way the distance between individuals is evaluated. Instead of sorting the population to determine the nearest neighbors, the multivariate generalization takes a diversity contribution radius for each individual by computing its minimum distance from the lower triangle of the pairwise matrix.

Also in chapter 3, we introduce a diagnostic tool for monitoring the search process EEB, which is now possible because we have both an exploration measure and an exploitation measure at our disposal. Our initial purpose in proposing such a tool was to further illustrate that GDM and PCM constitute complementary concepts, not complementary values. Having achieved this purpose, we realized that this diagnostic tool much handier than we first thought, as it provides a means to help practitioners identify the impact that customized strategies may have over search performance. Specifically, it is one of the few tools available to help us better understand the way an optimizer works over real-world problems. For example, it has been shown to be useful for comparing the behavior of various parameters and genetic operators over a real-coded steady-state genetic algorithm (SSGA). Overall, this application illustrates that the choice of specific EA parameters has a great impact on how much search convergence is delayed. However, none of the parameters considered was able to significantly modify the SSGA search path. In fact, a similar EEB trend was observed over both unimodal and multimodal landscapes, underscoring even more the need to develop a parameter setting strategy with the ability to manage the EEB.

In summary, an optimizer diagnostic tool, a genotypic diversity measure, and a rigorous framework developed for evaluating the performance of various phenotypic formulations are proposed in chapter 3. However, the main contribution of this study is undoubtedly our proposed answer to question #2 about the way to characterize the exploitation provided by the search process (PCM_{13}).

Before pursuing our main objective further, we believe that it is essential to present our evaluation of the performance and reliability of the new GDM (D_{LN}), which we do in chapter 4, using the validation framework introduced in chapter 2 and the quality criteria assessment benchmark defined in chapter 1. This investigation allows us to point out the fundamental shortcoming of D_{LN}, which is that the new proposal does not respect any of the three diversity requirements over the reduced population arrangement environment. However, compared to the other representative GDMs, D_{LN} is the only formulation capable of properly representing all seven cases of controlled population diversity. Based on this fact, we decided

to continue the analysis of this measure. As a result, we show that D_{LN} is more stable than D_{PW} over similarly scattered populations, and that it behaves properly over the multi-site convergence process. In addition, D_{LN} shows low sensitivity with respect to the simulation parameters (N and n) and very accurate estimates of the diversity for populations containing outliers. Consequently, we recommend using it until a formulation is developed that can cope with all the diversity requirements. These results constitute our answer to question #1.

All the above investigations lay the foundation for our EEB assessment and representation. From there, we focus on the development of an adaptive parameter control strategy by addressing the last three questions to be answered in this thesis. This part of the research is presented in the final chapter of this thesis.

In terms of the third thesis question, which is related to the conversion of the exploration and the exploitation knowledge into a reward for adapting the EA parameters, we propose a very flexible strategy. We consider the qualities of the parameters based on their ability to maximize the exploration and exploitation. In addition, to prevent the occurrence of the premature convergence phenomenon, we are adding a constraint region where exploration is very moderate and exploitation is very intense. This credit assignment scheme is applied to the EA parameters responsible for population governance, whereas the parameters involved in the creation of individuals are rewarded based on their fitness improvement capability.

Regarding the fourth thesis question, we develop a new adaptive parameter setting approach, since a thorough review of the adaptive systems currently available shows that none of them can support the parameter interaction dependency concurrently to the problem and evolution dependencies. As our proposed approach is based on a Bayesian network and applied to GA, we call it BNGA. All three parameter dependencies in this system are managed automatically: problem dependency is considered through the initial uniform probably allocated to each parameter state; evolution dependency is managed through the inference process; and parameter interaction is accounted for by the graphical model describing the BNGA structure. The decomposition of GA parameters into two subsets, as defined by their

reward mechanisms, leads to a system in a quest for short- and long-term performance. Performance assessment of BNGA is evaluated over a real-coded SSGA, where nine parameters are controlled. To the best of the author's knowledge, this is the first EA study that has managed this number of parameters at one time.

To answer the last thesis question, a comparison study is conducted over the CEC'05 benchmark. Static parameter setting, a naïve adaptive approach, PM, AP, FAUC-RMAB, CMA-ES, and G-CMA-ES are all evaluated in this study. To ensure fair competition, this study was preceded by a hyperparameter analysis over PM, AP, FAUC-RMAB, and BNGA. These experiments confirm the generalization potential of BNGA with respect to the other adaptive approaches, since it requires the smallest number of hyperparameters. In fact, only the minimum probability for each parameter state (p_{min}) is required. The overall comparison is performed by enhancing the statistical procedure proposed by Garcia *et al*. (2009). Very interesting conclusions arise from the results. First, BNGA demonstrates a performance that is statistically equivalent to that of G-CMA-ES, which is a state-of-the-art EA. From a practitioner's point of view, this is appealing, as it shows that competitive results can be achieved with a framework that is susceptible to customization, like SSGA. Note that, overall, FAUC-RMAB achieves similar performance to BNGA, but requires more hyperparameters to be set. Furthermore, the robustness of BNGA is demonstrated, as it provides equivalent or better results, in terms of solution quality, than the most competitive SSGA strategies over different kind of problems. This shows the usefulness of dealing with parameter interaction in an adaptive parameter control system, since it improves the ability of the optimizer to track the optimal search path, whatever the nature of the landscape.

However, BNGA requires significantly more function evaluations than static parameter setting or (G-)CMA-ES, in order to obtain the optimum over unimodal test functions. This is the disadvantage of most adaptive approaches, since, to some extent, robustness and efficiency act in opposition to one another. To remedy this problem, strategies integrating knowledge from previous runs have been introduced into BNGA, with the aim of biasing the initial parameter probabilities towards their most promising state. As expected, efficiency is

significantly improved over the unimodal functions. Nevertheless, this comes with a cost in the form of performance degradation over complex multimodal functions. Consequently, knowledge integration mechanisms are not recommended in their current form, but they do represent a promising approach to improving the overall search performance.

More importantly, our results illustrate that BNGA, like the other approaches we compare, is not capable of finding the global optimum over complex multimodal test functions. In fact, through an EEB footprint analysis provided in appendix V, we observe that none of the approaches compared is capable of reaching or sustaining useful diversity, which characterizes the maximization of both exploration and exploitation. This issue might suggest a missing synergy among the parameters, and demand an in-depth study on the choice of the parameter states involved.

All things considered, this research helps improve our knowledge of the EEB, and the way to measure this key search process characteristic. In addition, our development of the diagnostic tool contributes to the enhancement of EA customization. Finally, the BNGA proposal makes it possible to integrate all the knowledge gained, while at the same time BNGA is the first adaptive system with the capability to control all the EA parameters at once. The performance of this system has been shown to be very competitive. We therefore believe that the objectives of this thesis have been met, even though there is still a long way to go before a new generation of optimizers emerges.

RECOMMENDATIONS

Based on the results of this thesis, numerous recommendations are formulated below to support advanced research on performance enhancement of evolutionary algorithms.

First, the development of a more suitable multivariate measurement for population diversity evaluation should be the first priority in the quest to manage the exploration/exploitation balance (EEB). Even though no dubious observation was noted as a result of D_{LN} measurement in this thesis, it would be prudent to suggest the application of an indicator capable of coping with all the diversity requirements. This will ensure that no bias in the EEB analysis will affect the implementation of the other recommendations. To achieve this, we advise starting the validation of any new GDM proposal with the procedure presented in section 4.3.

We further recommend mapping the characteristics of as many EA parameters as possible over various test functions with the help of the EEB diagnostic tool. This will enable categorization of the kinds of synergies/complementarities that may be expected among the parameters, as well as their potential states. Also, we recommend the use of the EEB diagnostic tool to more accurately measure the performance of diversity promoting strategies, as well as that of the memetic algorithm proposal.

From there, the third step should be to validate these synergies/complementarities using BNGA (or another adaptive approach that supports all the parameter dependency). This step will be fundamental to examining the extent to which the EEB can be influenced. More importantly, it will make it possible to confirm or reject the notion that a relation exists among parameter synergies/complementarities, useful diversity, and search performance over multimodal problems.

In parallel, investigation of the EEB at the parameter level (as opposed to investigation at the search level, which is discussed throughout this thesis) should be initiated. This could

provide some interesting guidelines about the tradeoff between the exploitation of few parameter states and the exploration of numerous parameter states. For this purpose, we encourage readers to take note of the Blacksmith proposal of Maturana, Lardeux, and Saubion (2010).

We also propose that BNGA be extended over other metaheuristic approaches, like PSO and DE, leading to BNPSO and BNDE respectively. The rationale behind this proposal is to exploit the attractive features of BN adaptive systems, like their support of parameter dependencies and their generality, to help other population-based methods to enhance their search performance.

Another promising research direction could be to integrate the adaptive Bayesian network approach at the algorithm selection level. This would lead to a BN tailored for the selection of the best metaheuristic for a particular problem and a particular evolutionary stage, as well as a specific BN for every metaheuristic with the aim of selecting their most promising parameter state. In so doing, the search process itself might become a hyperheuristic (Burke, Kendall, and Newall, 2003; Ozcan, Bilgin, and Korkmaz, 2008) or a portfolio framework (Peng *et al.*, 2010). This recommendation is related to the search for a new generation of optimizers.

Finally, below we list some brief recommendations for further development of BNGA:

- Its performance over landscapes with higher dimensionality ($n > 10$) must be assessed;
- Other credit assignment schemes based on the EEB framework could be investigated;
- Other knowledge integration mechanisms could be developed;
- A surrogate model could be implemented with BNGA to reduce the computational cost of the global search strategy;
- A hyperparameter study could be conducted by weighting objectives based on problem complexity, instead of uniformly aggregating their outcomes.

APPENDIX I

OVERVIEW OF THE CEC'05 TEST FUNCTIONS CHARACTERISTICS

Table A I-1 provides the principal characteristics of the CEC'05 test functions, our aim being to provide a quick overview of the type and nature of the test functions used in this thesis. Further details about these test functions and the experimental methodology used are available in the report by Suganthan *et al*. (2005).

Table A I-1 Overview of the CEC'05 test functions characteristics

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

1 – Separable near the global optimum (Rastrigin)
2 – A local optimum is set on the origin
3 – Solved easily if initialization at the bounds

APPENDIX II

SELECTION OF THE CREDIT ASSIGMENT SCHEMES INVOLVED IN THE ADAPTIVE CONTROL SYSTEMS

In the preliminary design of BNGA, various attempts were made to define the best feedback indicators and credit assignment schemes for $r_{offspring}$ (equation 5.5). Based on these results, the most relevant approaches were reassessed with the final version of BNGA. This appendix presents a comparison of these reassessed approaches, which are described in equations A II-1 to A II-4. The first two equations are based on the fitness improvement of the offspring relative to that of the population. In this set, the first of them is based on the average fitness, while the second is based on the best fitness. In contrast, the last two equations are based on the fitness improvement of the offspring relative to that of the parents. Again, within this set, the first is based on the average fitness, while the second is based on the best fitness. Note that these formulations are defined for minimization problems.

$$r_{offspring} = 0.5 + 0.5 * \frac{\left(\bar{f} - \bar{f}_{offspring}\right)}{\left|\bar{f}\right|} \qquad \text{(A II-1)}$$

$$r_{offspring} = 0.5 + 0.5 * \frac{\left(f_{best} - f_{offspring,best}\right)}{\left|f_{best}\right|} \qquad \text{(A II-2)}$$

$$r_{offspring} = 0.5 + 0.5 * \frac{\left(\bar{f}_{parent} - \bar{f}_{offspring}\right)}{\left|\bar{f}_{parent}\right|} \qquad \text{(A II-3)}$$

$$r_{offspring} = 0.5 + 0.5 * \frac{\left(f_{parent,best} - f_{offspring,best}\right)}{\left|f_{parent,best}\right|} \qquad \text{(A II-4)}$$

Comparison of these credit assignment schemes is performed over the CEC'05 benchmark following the same methodology as described in section 5.4.1. However, to reduce the

computational burden, only 5 runs were conducted over each test function. Table A II-1 presents the results obtained for each scheme, which are described with the format introduced in section 5.5 (see Table 5.3).

To compare the performance of the credit assignment schemes, the statistical analysis procedure described in section 5.5 is applied. However, this statistical analysis is not performed on the independent test function, since not enough runs were conducted over each test function to achieve reasonable power from the statistical test. In other words, the sample size is too small (5 runs). As a matter of fact, Garcia *et al*. (2009) have proposed as a rule of thumb, that the sampling size must be at least twice as large as the number of algorithms compared. Table A II-2 presents the statistical tests produced over the group of functions. Overall, there is no significant difference between any of the four credit assignment schemes, except that equation A II-4 is dominated by the other approaches over the F1-F14 subset of functions.

Based on these statistical results, equations A II-1 to A II-3 could be selected as the credit assignment scheme for $r_{offspring}$. However, we chose equation A II-3, since it showed the best performance over the unimodal test functions (F1 to F6 in Table A II-1).

Concerning $r_{population}$, no credit assignment scheme comparison is provided here. In fact, even though numerous experiments were conducted in the preliminary design of BNGA to identify the most relevant GDM and PCM, chapters 1 to 4 provide arguments and evidence justifying the choice of D^N_{LN} (equation 5.7) and PCM_{13} (equation 5.8). In terms of their combination, again, various experiments were conducted in the preliminary phase of BNGA development. However, owing to the better overall performance of equation 5.9, this choice was not revisited with the final version of BNGA. Nonetheless, it is important to keep in mind that this choice may have an important effect on the performance of the adaptive system.

Table A II-1 Comparison of the credit assignment schemes proposed for $r_{offspring}$

Table A II-2 Statistical comparison of the $r_{offspring}$ credit assignment schemes by function group

	A II-1	A II-2	A II-3	A II-4
F1-F14	0.306	0.158	0.018	0.006
F15-F25	0.7412	0.591	0.559	0.741
F1-F25	0.742	0.742	0.3771	0.301

APPENDIX III

BNGA PROCESS

This appendix is intended to provide complementary information about BNGA adaptive parameter control approach that was introduced in section 5.3. BNGA is built from a Bayesian network (BN), which is an artificial learning method representing a causal network that exploits the Bayes' rule:

$$BEL(A) = P(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad \text{(A-III.1)}$$

From this theorem, it is possible to update our belief about an event A, given that new information from event B is available. In this formulation, P(A) represents the prior probability of event A, whereas P(A|B) or BEL(A) describes the posterior probability of event A, knowing event B. P(B|A) is the likelihood of A given B, or the conditional probability, and, finally, the marginal probability of B (P(B)) is included for normalization purposes, to ensure that posterior probabilities respect the law of probability $\left(\sum_j P(a_j|B) = 1.0\right)$.

The complete process of BNGA is depicted in Figure A III-1. Overall, BNGA can be divided in three phases. The first phase is related to the initialization of the system through the definition of the CPTs and the prior probabilities of each parameter. As discussed in section 5.3.3 and 5.3.4, both CPTs and prior are set from a uniform distribution to account for the fact that no initial knowledge is available over the problem at hand. The second phase is dedicated to the selection of the state of the parameters for the next generation and is done based on the established prior probabilities. Once the new generation is settled by the GA process, the third phase of BNGA is carried out. This phase involve rewards computation through equations 5.5 and 5.9 and then the CPTs are updated by using equations 5.6 and 5.10 for the parameters set associated to new solutions creation ($r_{offspring}$) and population

governance ($r_{population}$), respectively. Thereafter, the posterior probabilities of the parameters

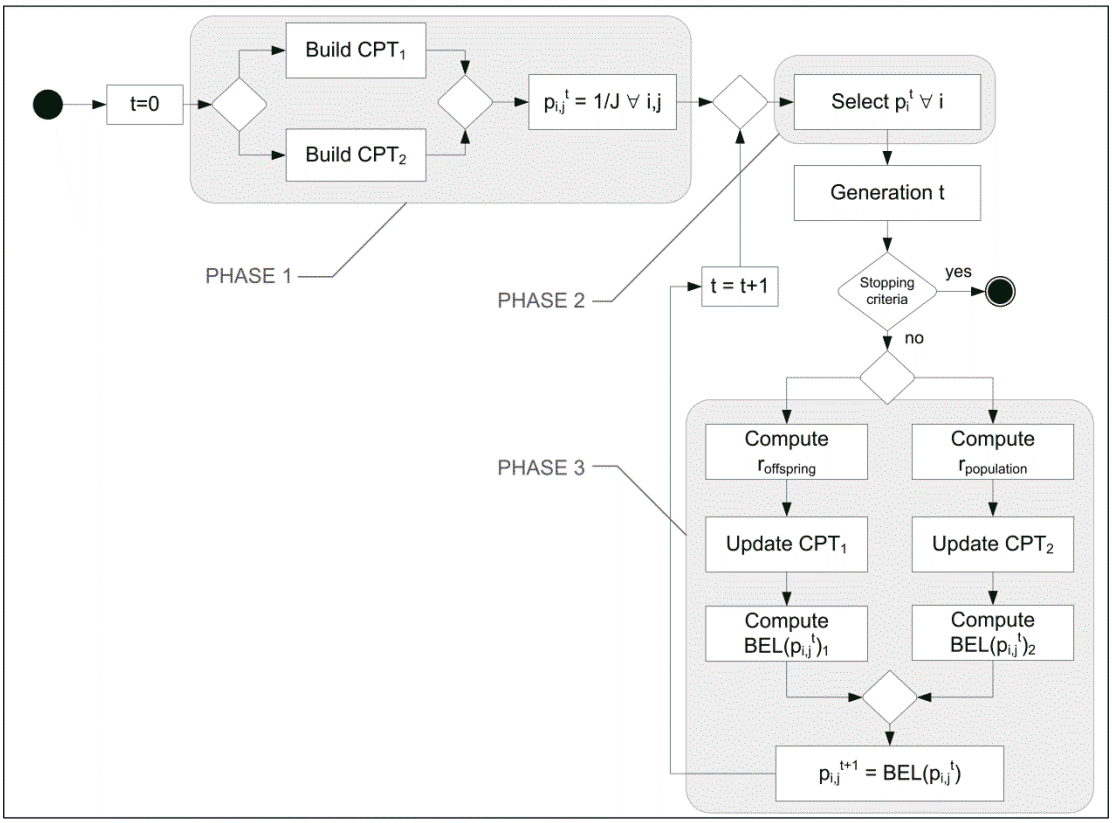


Figure A III-1 Flowchart of the BNGA process

are computed. To keep the adaptive system tractable, we selected a closed form inference strategy for computing them at each generation, which is the message passing algorithm (Pearl, 1988, p. 175) described by equations 5.11 and 5.12, and through pseudocode in Algorithm A III-1. Note that the inference algorithm is called twice at each generation, once for the structure related to $r_{offspring}$ ($p_m = 1$, $p_n = 4$) and again for the structure associated with $r_{population}$ ($p_m = 5$, $p_n = 9$). The last stage of the third phase is to update the prior probabilities for the next generation ($p_{i,j}^{t+1}$) by setting those probabilities equal to the posterior probabilities computed at the current generation. Phases 2 and 3 are thus repeated until the GA stopping criteria are met.

In summary, the accumulated knowledge from the credit assignment schemes is transferred to prior probabilities with the aim to improve the decision making for the parameters

selection of the next generation. Finally, it is worthwhile to mention that the CPTs update involved in the third phase is always made from the initial CPTs. This is critical to avoid double counting the same evidence gathered from one generation (section 5.3.3).

Algorithm A III-1 Inference algorithm used by BNGA

```
Algorithm-A III: BNGA inference (from prior CPT)
  Input: List of the parameters with their states, their prediction lists of
         the non-root states, conditional probability table
  Output: Posterior probabilities of the parameters states
for i = n, n-1, ..., 1 do
    SUM_FREQ(i,1) := 0
    for j = 1, ..., r do
        STATE := parent(i,j)
        for k = 1, ..., size(CPT) do
            IF CPT(k,1) = STATE
                /* save the conditional probabilities
                /* corresponding to the reward
                */
                PROD := CPT(k,2)
                for s = m, ..., 1 do
                    if s ≠ i
                        INSTATE := CPT(k,s)
                        PROD := PROD * pr(s(i),INSTATE)
                    end
                end for
                SUM_PROD(i,j) := SUM_PROD(i,j) + PROD
            end
        end for
    end for
    /* Post-order propagation (message passed by the children)
    /* represent element by element division
    /* normalization
    */
    relevance := SUM_FREQ / Σ_{j=1}^{r} SUM_FREQ

    /* Top-down propagation (message passed by the parent)
    /* directly read the prior, since parents nodes have no evidence
    */
    rip := prior(i,1:r)
    /* posterior probabilities computation
    /* represent element by element multiplication
    */
    POST(i,1:r) := relevance * rip
    /* normalization
    /* represent element by element division
    */
    POST(i,1:r) := POST(i,1:r) / Σ_{j=1}^{r} POST(i,j)

end for
Return (POST)
```

APPENDIX IV

SURVEY OF THE PARAMETER STATES USED BY THE VARIOUS ADAPTIVE APPROACHES OVER THE CEC'05 TEST FUNCTIONS

This appendix contains a survey of the parameter states selected by each adaptive method over each of the CEC'05 test functions, as well as overall consideration of all the test functions. Tables A IV-1 to IV-5 refer to PM, AP, FAUC-RMAB, BNGA or $BNGA_0$, and $BNGA_1$ respectively. In these tables, the most popular state for each parameter is highlighted in bold. As mentioned in section 5.4, the most promising states identified by BNGA (bold values in Table A IV-4) over each test function were selected for the static approach. Note that the naïve approach is not considered here, since the parameter states were selected more or less as they appear at the bottom of Table 5.1.

To summarize, the first state in the selection plan ("Best + random") dominates throughout the test functions, whatever the adaptive method considered. In terms of *offspring*, 2 is the most frequent choice, except for FAUC-RMAB and $BNGA_1$. In the former approach, the ratio is more uniform among the three potential states, with a small bias toward the creation of 10 offspring, while the choice of the number of offspring for the latter appears to be highly problem-dependent. For the genetic operator, PCX is the clear winner for all the approaches, except FAUC-RMAB, where PCX is selected approximately one third of time. Regarding the replacement plan, the *Parents* state dominates in PM, whereas the *1 random* state is selected about half the time by AP, and it stands out as the clear winner for BNGA and $BNGA_1$. FAUC-RMAB seems to allocate no preference to this parameter, as each state is selected more or less uniformly over each test function. For the update plan, PM and AP favor the *Best* state, whereas the choice is near uniform among the three potential states within FAUC-RMAB, BNGA, and $BNGA_1$. The population size is problem-dependent for PM, but, overall, 50 individuals is the state most often selected. For BNGA and $BNGA_1$, 500 individuals seems to be the norm, while this state is selected around 60% of the time by AP. FAUC-RMAB shows a very interesting trend with respect to the population size parameter. In fact, most of the time, 50 individuals are promoted over unimodal test functions, while for

multimodal test functions, the largest population (500 individuals) is the clear winner. This practice is in accordance with the common wisdom about the importance of population size allocation for maximizing search process efficiency. For the addition scheme, all the adaptive methods seem to use both potential states almost equally frequently, with some exceptions over a few test functions. Concerning the removal scheme, the available states are selected almost uniformly throughout each adaptive method, except for PM, where the *Worst* state dominates.

Finally, it is worth noting that this parameter state survey provides only a partial view of the behavior of each adaptive method, as only the problem dependency aspect can be examined. Consequently, it would be unwise to toss out a parameter state based on this knowledge alone, since infrequent states may be responsible for a particular synergy buildup among parameters, or for steering the search path towards a promising outlook.

Table A IV-1 SSGA parameter states selected by the PM adaptive method

Function	Selection pressure					Number of offspring (β_c)		
	Roulette	Group 10	Tournament	Truncation	1.1-5.6	1	5	30
F1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F11	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F12	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F13	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F14	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F15	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F16	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F17	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F18	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F19	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F20	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F21	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F22	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F23	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F24	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F25	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Avg.	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Function	Crossover operator							Mutation rate (p_m)			
	Parent	Random	Uniform	Standard	Blend	BLX	BLX	10	100	500	1000
F1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F11	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F12	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F13	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F14	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F15	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F16	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F17	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F18	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F19	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F20	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F21	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F22	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F23	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F24	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F25	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Avg.	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Function	Inhibition scheme		Removal scheme			
	Random	Child + parent	Worst	Deterministic	Stochastic	Tournament 10
F1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F11	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F12	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F13	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F14	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F15	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F16	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F17	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F18	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F19	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F20	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F21	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F22	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F23	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F24	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F25	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Avg.	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Table A IV-2 SSGA parameter states selected by the AP adaptive method

Function	Selection plan					Number of offspring (λ)			Crossover operator					Replacement plan				Update plan			Population size (%)				Additional scheme					
	Best + random	Group 10	Random	Tournament [illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Parents	1 random	2 random	5 random	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Tournament [illegible]	[illegible]
F1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F11	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F12	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F13	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F14	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F15	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F16	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F17	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F18	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F19	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F20	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F21	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F22	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F23	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F24	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F25	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Table A IV-3 SSGA parameter states selected by the FAUC-RMAB adaptive method

Timetable	Selection plan					Number of offspring (%)			Crossover operator					Replacement plan (%)				Update plan			Population size (%)				Adaptive scheme	Diversity scheme			
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]

Table A IV-4 SSGA parameter states selected by the BNGA adaptive method

[illegible]	Selection plan					Number of offspring (λ)				Genetic operator				Replacement plan				Update plan			Population size (%)				[illegible]		[illegible]			
	[illegible]	[illegible]	Random	Tournament	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Random	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T11	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T12	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T13	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T14	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T15	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T16	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T17	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T18	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T19	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T20	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T21	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T22	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T23	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T24	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T25	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
TOTAL	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Table A IV-5 SSGA parameter states selected by the $BNGA_1$ adaptive method

Function	Selection plan					Number of individuals (N)			Crossover operator					Replacement plan							Population size (N)				Adaptive scheme		Mutation scheme			
	Rec. Tournament	Group Tournament	Random	Tournament 3	LUS	1	5	30	PMX	CMPX	TPX	TPX 1/5	PX-A	Tournament	Tournament	Tournament	Random	Elitist	DTI	ELI	10	100	300	500	Random	Child + random	Wyvern	Tournament 3	Tournament 5	Tournament-LS
F1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F11	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F12	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F13	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F14	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F15	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F16	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F17	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F18	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F19	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F20	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F21	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F22	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F23	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F24	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F25	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
TOTAL	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

APPENDIX V

SUMMARY OF THE EEBs ACHIEVED OVER SOME CEC'05 TEST FUNCTIONS

Since one of the fundamental purposes of BNGA is, like that of any other parameter adaptive setting approach, to steer the search process EEB in the optimal direction, it is highly relevant to illustrate how the adaptive strategies tested really perform in this respect. Figures A V-1 to V-6 present the EEB spectrum achieved by all the parameter setting approaches defined in section 5.4 over test functions F2, F6, F8, F15, F21, and F25 respectively. These test functions were selected to provide a general, but concise picture of the EEB management offered by each parameter setting strategy. On these charts, each color corresponds to a specific run. Accordingly, we decided to exhibit the EEB footprint of the parameter setting strategies, rather than their path history, as provided, for instance, by the EEB diagnostic tool in section 3.8. We recall that the EEB exploration and exploitation axes are computed from the GDM and PCM formulations given by equations 5.7 and 5.8 respectively.

Overall, the EBB of the static parameter setting approach (Figures A V-1 to V-6 a)) represents the conventional EA evolution path well, as the search process is automatically directed towards the best individual found so far. Interestingly, this makes the EEB footprint compatible with its path history. This is not the case for F8, however, but the results from this test function will be discussed separately below for all the parameter setting strategies. Surprisingly, the naïve adaptive approach (Figures A V-1 to V-6 b)) more often than not seems to settle on three specific EEB regions that are well balanced with respect to the exploration axis. In contrast, PM, AP, FAUC-RMAB, BNGA, and $BNGA_1$ (Figures A V-1 to V-6 b) to g)) have a fairly similar EEB footprint. Nevertheless, AP and FAUC-RMAB are shown to favor more extreme EEB zones, with a greater tendency towards more exploration and less exploitation. CMA-ES and G-CMA-ES show very similar behavior regarding their EEB footprint, in that few generations are allocated to extensive exploration, and most of their search processes are rapidly dedicated to exploitation of a narrow landscape region. However, G-CMA-ES provides more exploration capability over high multimodal problems

(F15, F21, and F25) than CMA-ES. This outcome is expected considering the restart feature embedded inside G-CMA-ES. Nevertheless, as shown in Table 5.6, the higher exploration capability of G-CMA-ES is not enough to circumvent the premature convergence issue observed over the high multimodal problems.

Test function F8 (Figure A V-3) is very interesting, as each parameter setting strategy behaves very differently from the other test functions considered. In fact, F8 has a kind of "egg box" landscape, with the global optimum basin located on its bound. Except for occasional FAUC-RMAB generations, where individuals seem to be spread throughout one or more local basins of attraction (Figure A V-3 e)), most adaptive approaches continuously spread their resources across the landscape (GDM → 1). In contrast, the footprints of CMA-ES and especially G-CMA-ES EEB suggest that they concentrate the vast majority of their resources in a very few local basins of attraction. Interestingly, a differentiating feature of these two strategies is highlighted in this test function, which is that they show a coverage of up to approximately 80% of the EEB space. Nevertheless, none of them appears effective, as all the searching approaches tested remain stuck in a local optimum basin (Tables 5.4 and 5.6).

Finally, it is worth mentioning that none of the parameter setting strategies, adaptive or otherwise, was capable of navigating, or, more generally, sustaining useful diversity. We conjecture that this may explain the poor performance of all the approaches tested on multimodal test functions, and suggests an apparent lack of synergies and/or complementarities among the parameters.

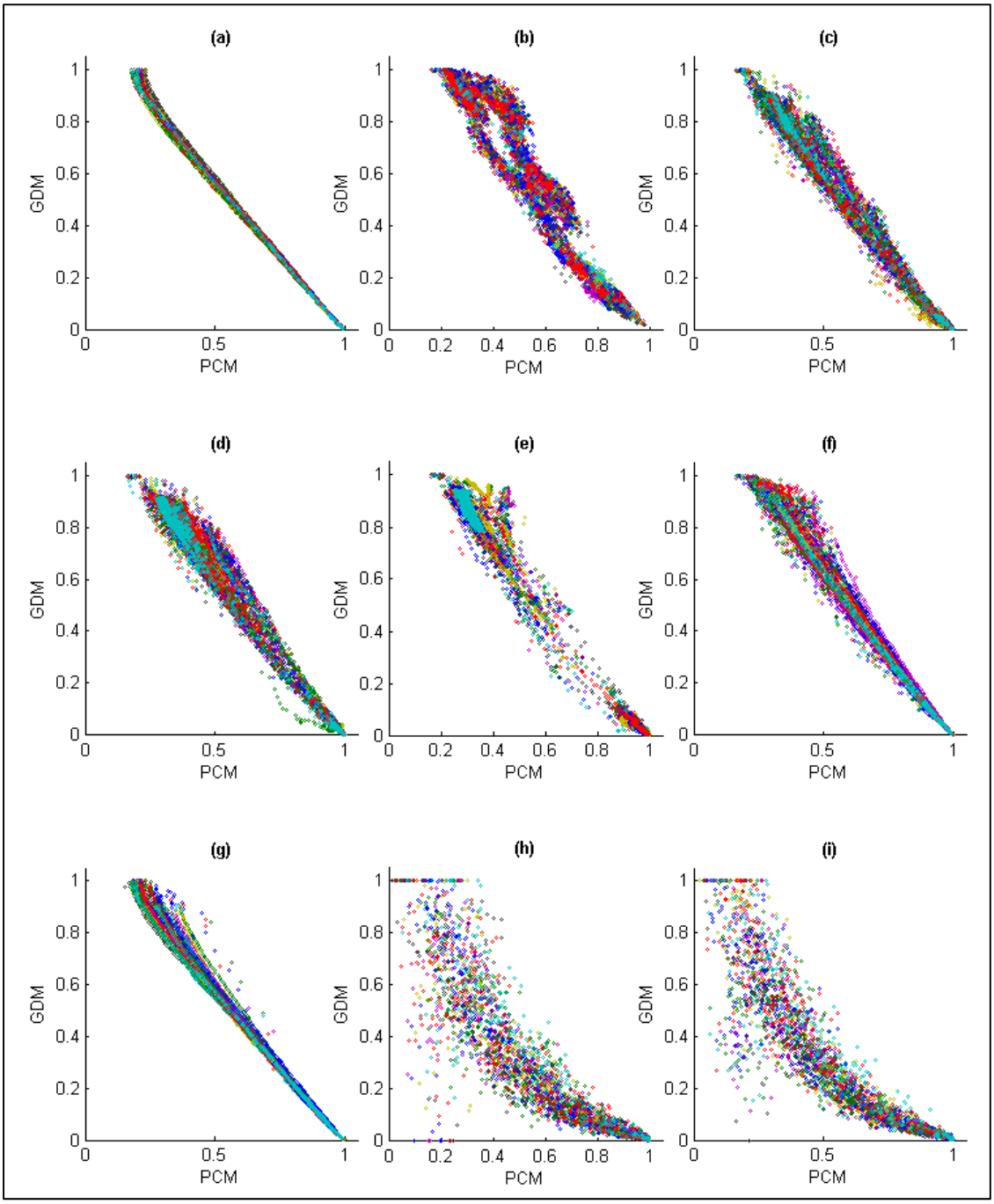


Figure A V-1 EEB history over F2 (Shifted Schwefel's problem 1.2) with: a) Static parameter setting, b) Naïve approach, c) PM, d) AP, e)FAUC-RMAB, f) BNGA, g) $BNGA_1$, h) CMA-ES, i) G-CMA-ES

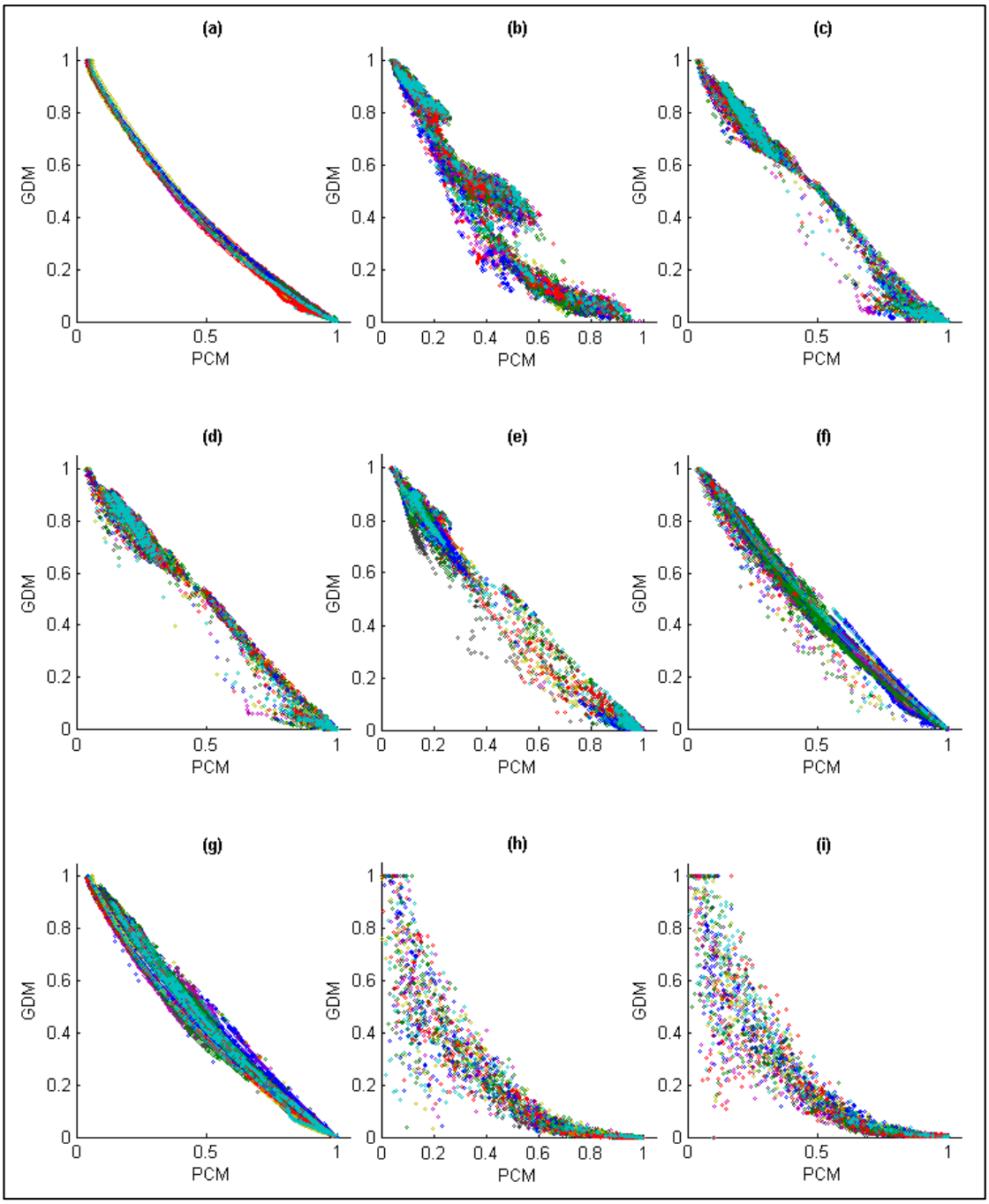


Figure A V-2 EEB history over F6 (Shifted Rosenbrock's function) with: a) Static parameter setting, b) Naïve approach, c) PM, d) AP, e)FAUC-RMAB, f) BNGA, g) $BNGA_1$, h) CMA-ES, i) G-CMA-ES

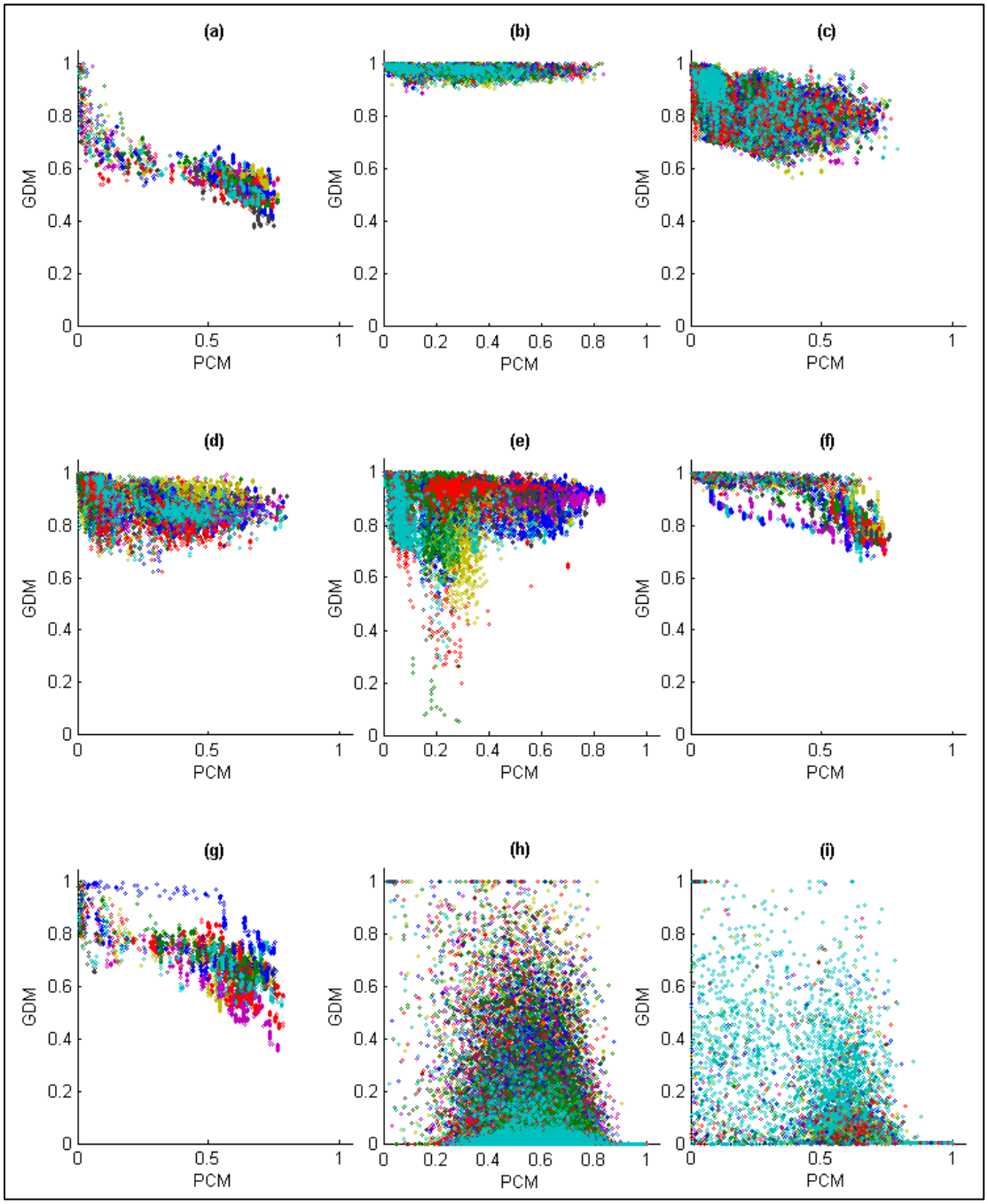


Figure A V-3 EEB history over F8 (Shifted rotated Ackley's function with global optimum on bounds) with: a) Static parameter setting, b) Naïve approach, c) PM, d) AP, e)FAUC-RMAB, f) BNGA, g) $BNGA_1$, h) CMA-ES, i) G-CMA-ES

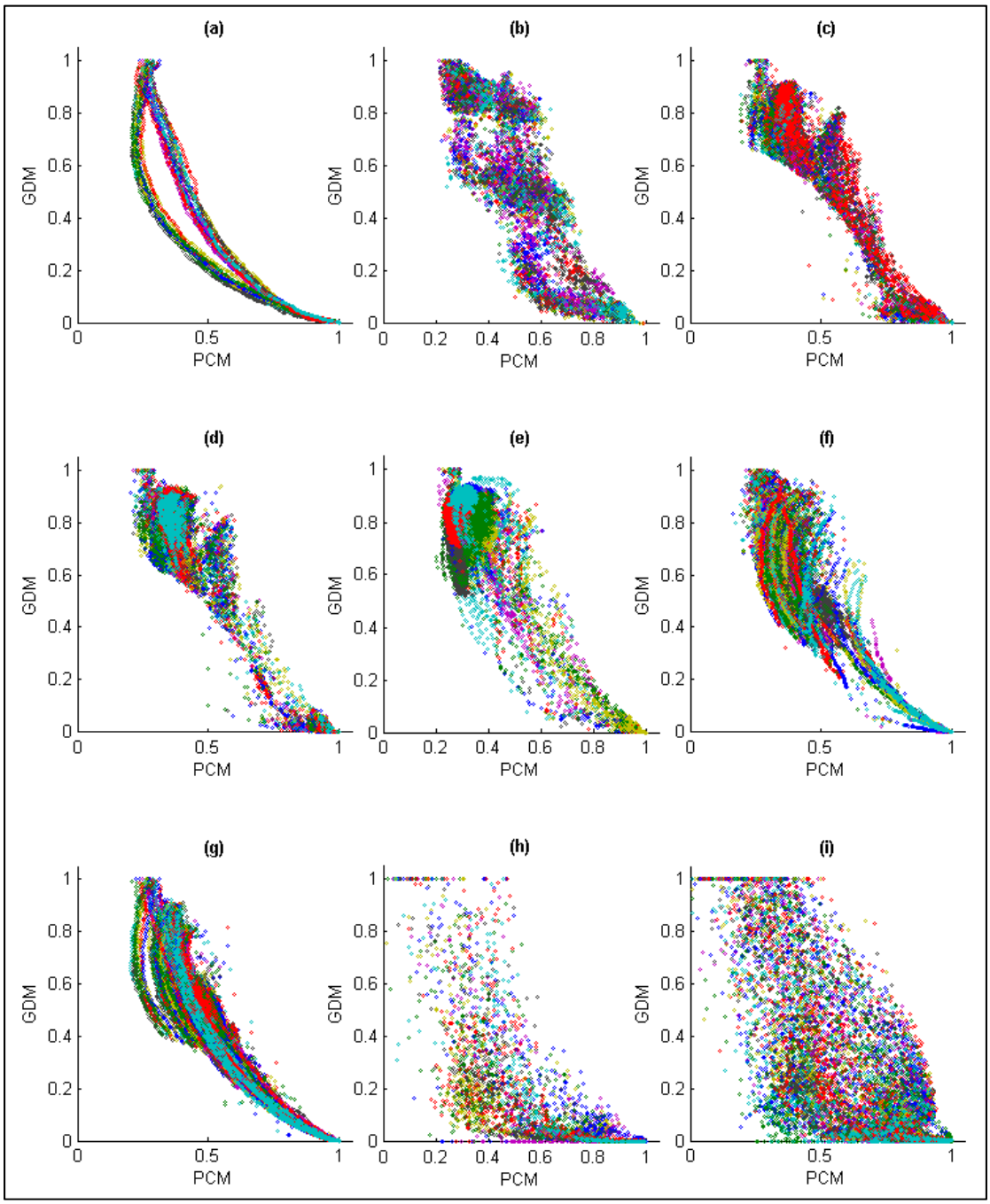


Figure A V-4 EEB history over F15 (Hybrid composition function) with: a) Static parameter setting, b) Naïve approach, c) PM, d) AP, e)FAUC-RMAB, f) BNGA, g) $BNGA_1$, h) CMA-ES, i) G-CMA-ES

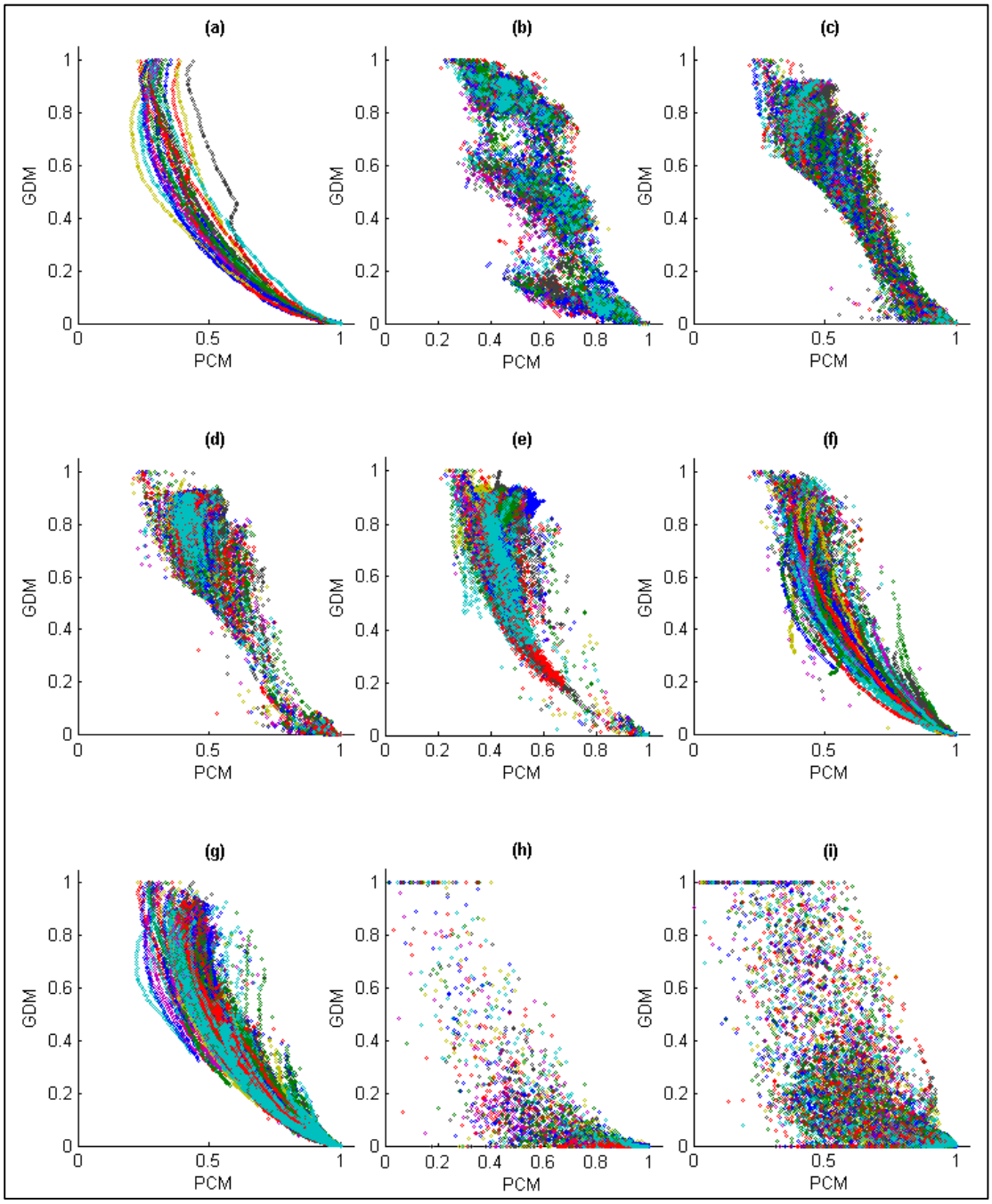


Figure A V-5 EEB history over F21 (Rotated hybrid composition function) with: a) Static parameter setting, b) Naïve approach, c) PM, d) AP, e)FAUC-RMAB, f) BNGA, g) $BNGA_1$, h) CMA-ES, i) G-CMA-ES

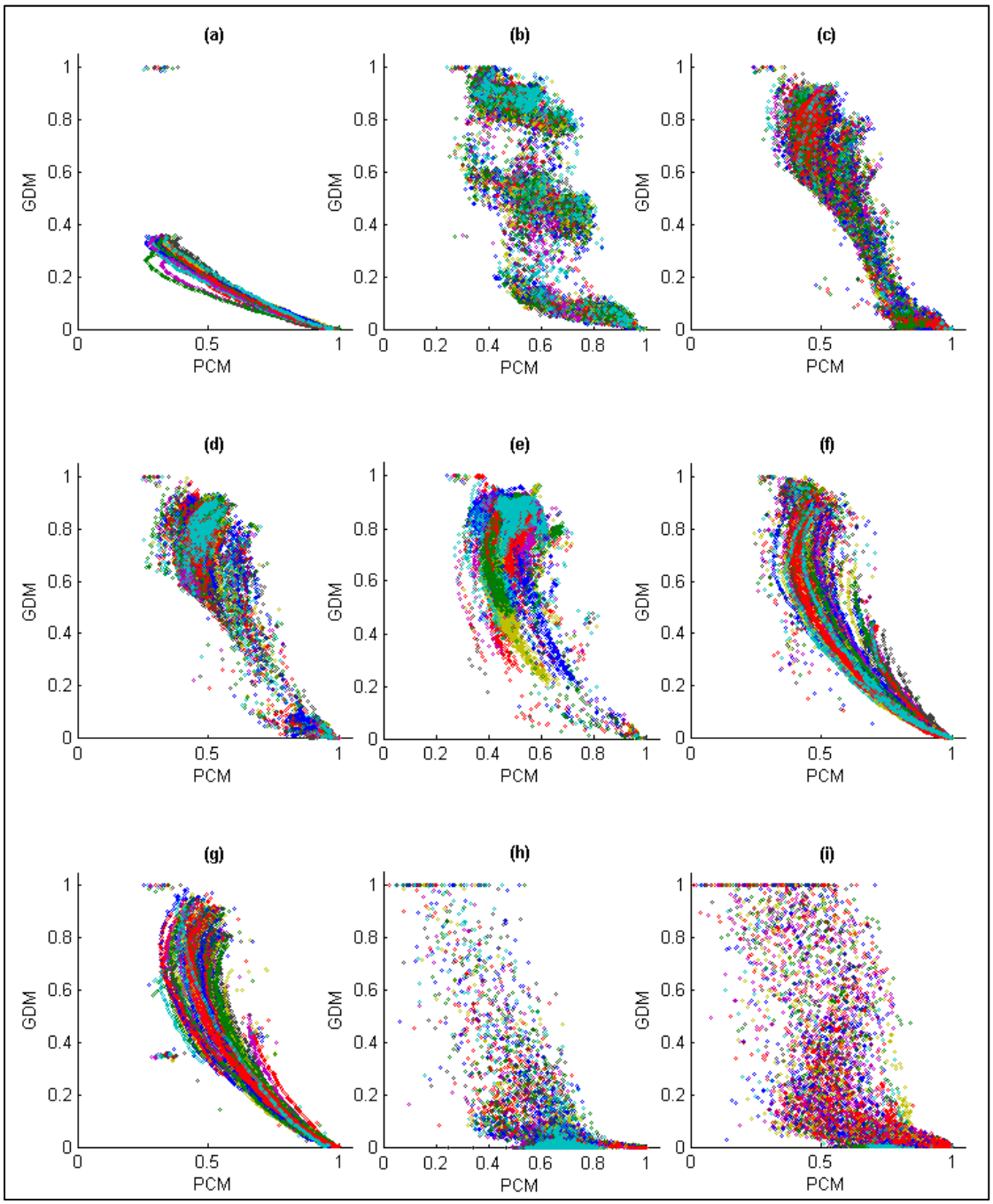


Figure A V-6 EEB history over F25 (Rotated hybrid composition function without bounds) with: a) Static parameter setting, b) Naïve approach, c) PM, d) AP, e) FAUC-RMAB, f) BNGA, g) $BNGA_1$, h) CMA-ES, i) G-CMA-ES

APPENDIX VI

RELEVANCE OF THE KNOWLEDGE INTEGRATION MECHANISM WITHIN BNGA

One of the attractive features of using adaptive parameter setting approaches is that they can provide knowledge about the problem at hand and then reinsert that knowledge into the search algorithm to improve its performance through search component modification. This might be seen as a interesting response to the fundamental assumption behind the NFL theorem (section 0.1), which is that nothing is known a priori about the function to be solved.

We took this line of thought one step further with the BNGA framework by designing knowledge integration mechanisms for it. The basic idea behind this concept is to capitalize on previous runs of a problem by transferring lessons learned to the search algorithm. From a practical perspective, a knowledge integration mechanism is applied by biasing the prior probabilities of each parameter state at the beginning of the search process, instead of using uniform probabilities, as is the case for $BNGA_0$. Intuitively, this kind of mechanism may result in a constant performance improvement of the search as the number of runs increases.

The purpose of this appendix is therefore to test this hypothesis over four different knowledge integration mechanisms designed for BNGA. So, beginning with the second run, the parameter state probabilities at the first generation are defined as follows:

$BNGA_1$: the *mean* of the parameter state probabilities from all the generations of the previous run;

$BNGA_2$: the probabilities of the parameter configuration that provided the *maximum exploration* throughout the previous run;

$BNGA_3$: for p_1 to p_4 – the probabilities of the parameter configuration that created the *highest fitness improvement* throughout the previous run, and

for p_5 to p_9 – the probabilities of the parameter configuration that provided the *maximum exploration* throughout the previous run;

$BNGA_4$: for p_1 to p_4 – the probabilities of the parameter configuration that created the *highest fitness improvement* throughout the previous run, and

for p_5 to p_9 – the probabilities of the parameter configuration that provided the *maximum exploitation* throughout the previous run;

$BNGA_3$ and $BNGA_4$ might be viewed as the two opposite ends of a knowledge integration strategy, while $BNGA_1$ is less aggressive. The difference between $BNGA_2$ and $BNGA_3$ is that the former rate offspring creation parameters (p_1 to p_4) based on their contribution to the population EEB, instead of using their direct impact, as is the case for the latter strategy.

Comparison of those knowledge integration mechanisms is performed over the CEC'05 benchmark following the same methodology as described in section 5.4.1. Table A VI-1 presents the results obtained over the 25 test functions for each strategy. Results are given in the format introduced in section 5.5 (see Table 5.3), and statistical tests are performed as described in the same section.

The results show that $BNGA_3$ is not competitive at all. In contrast, $BNGA_1$ and $BNGA_4$ provide some benefit. In reality, Figure A VI-1 illustrates that $BNGA_1$ and $BNGA_4$ can significantly reduce the number of function evaluations required. Interestingly, this outcome is effective immediately after they are put into operation (run #2). Nonetheless, the advantage of $BNGA_1$ and $BNGA_4$ is recorded only over simple unimodal test functions, where the optimum is always found, whether a knowledge integration mechanism is embedded or not. Regarding $BNGA_2$, no advantage was found over $BNGA_0$. Table A VI-2 confirms statistically that, in terms of solution quality (MBFE), $BNGA_0$ is not dominated by any of the knowledge integration mechanisms that have been designed.

Table A VI-1 Comparison of the proposed knowledge integration mechanisms

When we put into perspective the performance of $BNGA_1$ and $BNGA_4$ that we observed with the time required for assessing the parameter state probabilities at the beginning of each run, we conclude that it is not worth implementing these parameter setting approaches at their current development stage.

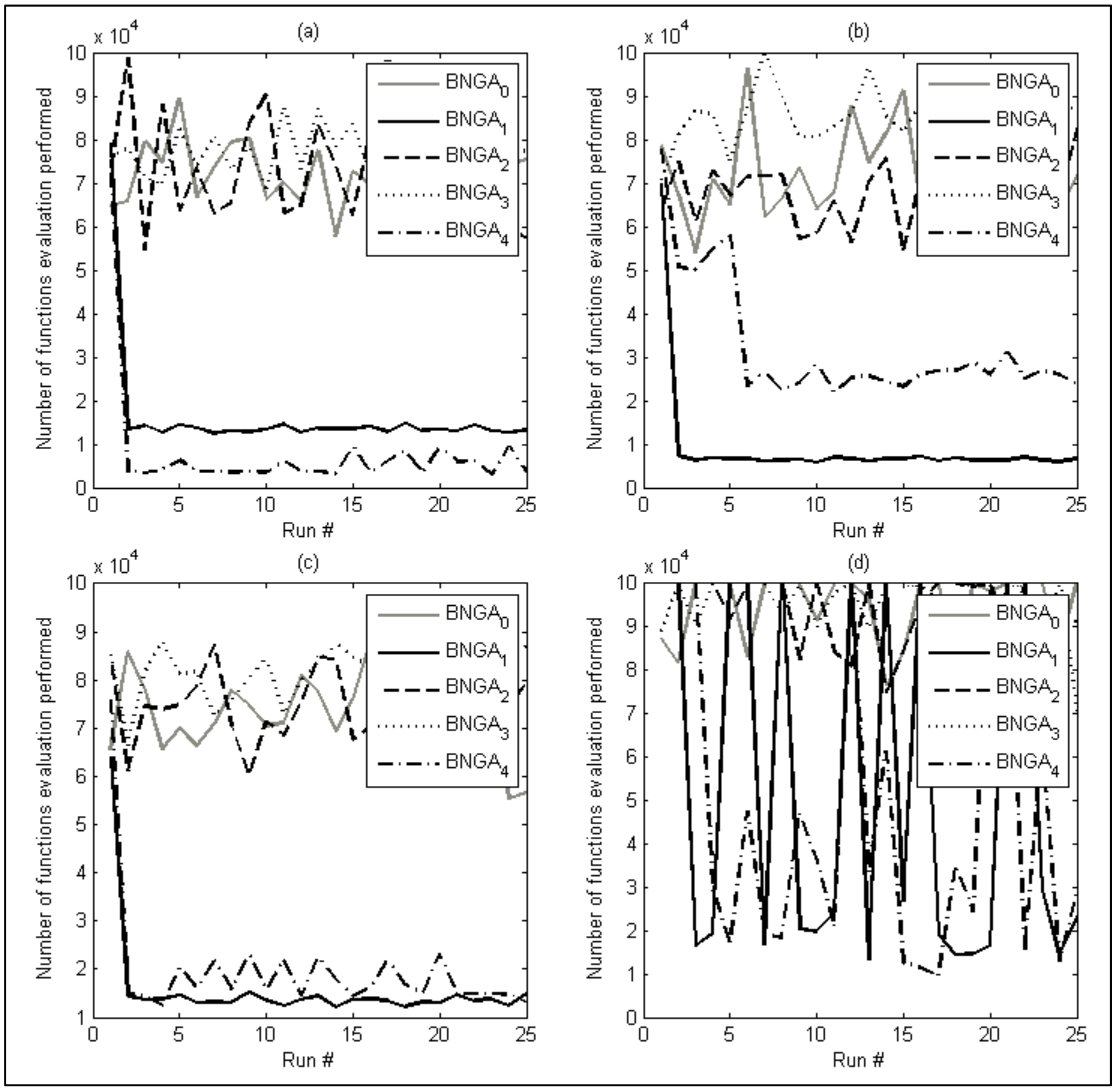


Figure A VI-1 Impact of the proposed knowledge integration mechanisms over the 25 runs history of test function: a) F1, b) F2, c) F4, d) F6

Table A VI-2 Statistical comparison of the proposed knowledge integration mechanism by function group

	$BNGA_1$	$BNGA_1$	$BNGA_2$	$BNGA_3$	$BNGA_4$
F1-F14	0.811	0.069	0.696	0.034	0.811
F15-F25	0.8927	<1E-3	<1E-3	0.006	0.006
F1-F25	0.004	0.147	0.7885	0.003	0.257

LIST OF BIBLIOGRAPHICAL REFERENCES

Abbass, Hussein A. and Kalyanmoy Deb. 2003. "Searching under multi-evolutionary pressures". In *Proceedings of the 2nd International Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*. (Faro, Portugal, April 8-11 2003), p. 391-404. Berlin (Germany): Springer-Verlag.

Alba, Enrique and José M. Troya. 2002. "Improving flexibility and efficiency by adding parallelism to genetic algorithms". *Statistics and Computing*, vol. 12, no° 2, p. 91-114.

Arnone, S., M. Dell'Orto, and A. Tettamanzi. 1994. "Toward a fuzzy government of genetic populations". In *Proceedings of the 6th International Conference on Tools Artificial Intelligence (ICTAI 1994)*. (New Orleans, LA, November 6-9 1994), p. 585-591. Piscataway (NJ): IEEE.

Auger, Anne and Nikolaus Hansen. 2005. "A restart CMA evolution strategy with increasing population size". In *Proceedings of the IEEE Congress on Evolutionary Computation* (*CEC 2005*). (Edinburgh, UK, September 2-4 2005), p. 1769-1776. Piscataway (NJ): IEEE.

Bäck, Thomas. 1992. "The Interaction of Mutation Rate, Selection, and Self-Adaptation Within a Genetic Algorithm". In *Proceedings of the 2nd International Conference on Parallel Problem Solving from Nature* (*PPSN 1992*). (Brussels, Belgium, Sept. 28-30 1992), p. 85-94. Amsterdam (Netherlands): Elsevier Science Publishers.

Ballester, Pedro J. and Jonathan N. Carter. 2004. "An effective real-parameter genetic algorithm with parent centric normal crossover for multimodal optimisation". In *Proceedings of the Genetic and Evolutionary Computation Conference* (*GECCO 04*). (Seattle, WA, June 26-30 2004), p. 901-913. Berlin (Germany): Springer-Verlag.

Barbosa, Helio J.C. and Asla Medeiros e Sá. 2000. "On adaptive operator probabilities in real coded genetic algorithms". In *Workshop on Advances and Trends in Artificial Intelligence for Problem Solving* (*SCCC 2000*). (Santiago, Chile, November 16-18 2000), p. 1-6. Washington DC: IEEE Computer Society.

Barker, Allen L. and W. N. Martin. 2000 "Dynamics of a distance-based population diversity measure". In *Proceedings of the IEEE Congress on Evolutionary Computation (CEC 2000)*. (La Jolla, CA, July 16-19 2000), p. 1002-1009. Piscataway (NJ): IEEE.

Bassett, Jeffrey K., and Kenneth A. De Jong. 2011. "Using multivariate quantitative genetics theory to assist in EA customization". In: *Proceedings of the 11th International Workshop on Foundations of Genetic Algorithms (FOGA 2011)*. (Schwarzenberg, Austria, January 5-8 2011), p. 219-229. New York (NY): ACM Press.

Beume Nicola and Günter Rudolph. 2006. "Faster S-metric calculation by considering dominated hypervolume as Klee's measure problem". In *Proceedings of the Second IASTED International Conference on Computational Intelligence* (*CI 2006*). (San Francisco, CA, November 20-22 2006), p. 233-238. Calgary (AB): IASTED/ACTA Press.

Beyer, Hans-Georg. 1998. "On the "explorative power" of ES/EP-like algorithms". In *Proceedings of the 7th International Conference on Evolutionary Programming VII* (*EP98*). (San Diego, CA, March 1998), p. 323-334. London (UK): Springer-Verlag.

Beyer, Hans-Georg and Kalyanmoy Deb. 2001. "On self-adaptive features in real-parameter evolutionary algorithms". *IEEE Transactions on Evolutionary Computation*, vol. 5, no° 3, p. 250-270.

Boissonnat, Jean-Daniel, Micha Sharir, Boaz Tagansky, and Mariette Yvinec. 1995. "Voronoi diagrams in higher dimensions under certain polyhedral distance functions". In *Proceedings of the Eleventh Anual Symposium on Computational Geometry* (*SCG 95*). (Vancouver, BC, June 5-12 1995), p. 79-88. New York (NY): ACM Press.

Burke, Edmund *et al.* 2003. "Hyper-heuristics: an emerging direction in modern search technology". In *Handbook of metaheuristics*, edited by Glover, Fred and Gary A. Kochenberger, p. 457-474. Norwell (MA): Kluwer Academic Publishers.

Burke, Edmund K, Steven Gustafson, and Graham Kendall. 2004. "Diversity in genetic programming: An analysis of measures and correlation with fitness". *IEEE Transactions on Evolutionary Computation*, vol. 8, no°1, p. 47-62.

Caponio, Andrea, Giuseppe Leonardo Cascella, Ferrante Neri, Nadia Salvatore, and Mark Summer. 2007. "A fast adaptive memetic algorithm for online and offline control design of PMSM drives". *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, vol. 37, no° 1, p. 28-41.

Caponio, Andrea, Ferrante Neri, and Ville Tirronen. 2009. "Super-fit control adaptation in memetic differential evolution frameworks". *Soft Computing – A Fusion of Foundations, Methodologies, and Applications*, vol. 13, no° 8-9, p. 811-831.

Cha, Sung-Hyuk and Sargur N. Srihari. 2002. "On measuring the distance between histograms". *Pattern Recognition*, vol. 35, no° 6, p. 1355-1370.

Collins, R. J. and D. R. Jefferson. 1991. "Selection in massively parallel genetic algorithms". In *Proceedings of the 4th International Conference on Genetic Algorithms (ICGA 1991).* (San Diego, CA, July 13-16 1991), p. 249-256. San Francisco (CA): Morgan Kaufmann Publishers.

Corriveau, Guillaume, Raynald Guilbault, and Antoine Tahan. 2010. “Genetic algorithms and finite element coupling for mechanical optimization”. *Advances in Engineering Software*, vol. 41, no° 3, p. 422-426.

Corriveau, Guillaume, Raynald Guilbault, Antoine Tahan, and Robert Sabourin. 2012a. “Review and study of genotypic diversity measures for real-coded representations”. *IEEE Transactions on Evolutionary Computation*, vol. 16, no° 5, p. 695-710.

Corriveau, Guillaume, Raynald Guilbault, Antoine Tahan, and Robert Sabourin. 2012b. “Evaluation of genotypic diversity measurements exploited in real-coded representation”. Submitted for publication to *Applied Soft Computing*, ASOC-D-12-01592.

Corriveau, Guillaume, Raynald Guilbault, Antoine Tahan, and Robert Sabourin. 2012c. “Bayesian network as an adaptive parameter setting approach for genetic algorithms”. Submitted for publication to *Applied Soft Computing*, ASOC-D-12-00925.

Corriveau, Guillaume, Raynald Guilbault, Antoine Tahan, and Robert Sabourin. 2013. “Review of phenotypic diversity formulations for diagnostic tool”. *Applied Soft Computing*, vol. 13, no° 1, p. 9-26.

DaCosta, Luis, Alvaro Fialho, Marc Schoenauer, and Michèle Sebag. 2008. “Adaptive operator selection with dynamic multi-armed bandits”. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 08)*. (Atlanta, GA, July 12-16 2008), p. 913-920. New York (NY): ACM Press.

Darwen, P. J. 2000. “Black magic: interdependence prevents principled parameter setting, self-adapting cost too much computation”. In *Applied Complexity: from Neural Nets to Managed Landscapes*, (Dunedin, New Zealand, Nov. 19-22 2000), p. 227-237. Christchurch (New Zealand): Institute for Food and Crop Research.

Das, Swagatam, Sayan Maity, Bo-Yang Qu, and P.N. Suganthan. 2011. “Real-parameter evolutionary multimodal optimization – A survey of the state-of-the-art”. *Swarm and Evolutionary Computation*, vol. 1, no° 2, pp.71-88.

Davis Lawrence. 1989. “Adapting operator probabilities in genetic algorithms”. In *Proceedings of the Third International Conference on Genetic Algorithms* (*ICGA 1989*). (Fairfax, VA, June 1989), p. 61-69. San Francisco (CA): Morgan Kaufmann Publishers Inc.

Deb, Kalyanmoy, D. Joshi, and A. Anand. 2002. “Real-Coded Evolutionary Algorithms with Parent-Centric Recombination”. In *Proceedings of the 2002 World Congress on Computational Intelligence (WCCI 2002)*. (Honolulu, HI, May 12-17 2002), p. 61-66. Piscataway (NJ): IEEE.

Deb, Kalyanmoy, A. Anand, and D. Joshi. 2002. “A computationally efficient evolutionary algorithm for real-parameter optimization”. *Evolutionary Computation*, vol. 10, no° 4, p. 371-395.

De Jong, Kenneth A. 1975. “An analysis of the behavior of a class of genetic adaptive systems”. Ph.D. dissertation, Dept. Comput. Commun. Sci., Michigan Univ., Ann Arbor, MI, 256 p.

Dezinger, Jörg, and Jordan Kidney. 2003. “Improving Migration by Diversity”. In *Proceedings of the IEEE Congress on Evolutionary Computation (CEC 2003)*. (Canberra, Australia, December 8-12 2003), p. 700-707. Piscataway (NJ): IEEE.

Diaz-Gomez, Pedro A. and Dean F. Hougen. 2009. “Three interconnected parameters for genetic algorithms”. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 09)*. (Montreal, Canada July 8-12 2009), p. 763-770. New York (NY): ACM Press.

Druzdzel, Marek J. and Linda C. van der Gaag. 2000. “Building Probabilistic Netwroks: “Where do the numbers come from?” Guest Editors’ introduction”. *IEEE Transactions on Knowledge and Data Engineering*, vol. 12, no° 4, p. 481-486.

Eiben, A.E. and C.A. Schippers. 1998. “On evolutionary exploration and exploitation”. *Fundamenta Informaticae*, vol. 35, no° 1-4, p. 1-16.

Eiben, Agoston E., Robert Hinterding, and Zbigniew Michalewicz. 1999. “Parameter control in evolutionary algorithms”. *IEEE Transactions on Evolutionary Computation*, vol. 3, no°2, p. 124-141.

Eiben, Agoston E., and Jim E. Smith. 2003. *Introduction to evolutionary computing*, 1st ed. Berlin (Germany): Springer-Verlag, 299 p.

Eiben, A.E., M.C. Schut, and A.R. de Wilde. 2006. “Boosting genetic algorithms with self-adaptation selection”. In *Proceedings of the IEEE Congress on Evolutionary Computation (CEC 2006)*, (Vancouver, Canada, July 16-21 2006), p. 477-482. Piscataway (NJ): IEEE.

Eshelman Larry J., and J. David Schaffer. 1991. “Preventing premature convergence in genetic algorithms by preventing incest”. In *Proceedings of the 4th International Conference on Genetic Algorithms (ICGA 1991).* (San Diego, CA, July 13-16 1991), p. 115-122. San Francisco (CA): Morgan Kaufmann Publishers.

Eshelman, Larry J. and J. David Schaffer. 1992. “Real-coded genetic algorithms and interval-schemata”. In *Proceedings of the 2nd International Workshop on Foundations of Genetic Algorithms (FOGA 1992)*. (Vail, CO, July 26-29 1992), p. 187-202. San Francisco (CA): Morgan Kaufmann Publishers.

Fialho, Alvaro Roberto Silvestre. 2011. "Adaptive operator selection for optimization". Ph.D. thesis, École Doctorale d'Informatique. Université Paris-Sud XI, Orsay, France, 211p.

Friedrich, Tobias, Nils Hebbinghaus, and Frank Neumann. 2009. "Comparison of simple diversity mechanisms on plateau functions". *Theoretical Computer Science*, vol. 410, no° 26, p. 2455-2462.

Garcia, Salvador, Daniel Molina, Manuel Lozano, and Francisco Herrera. 2009. "A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: a case study on the CEC'2005 special session on real parameter optimization". *Journal of Heuristics*, vol. 15, no° 6, p. 617-644.

Gini, Corrado. 1921. "Measurement of inequality of incomes". *The Economic Journal*, vol. 31, no° 121, p. 124-126.

Goldberg, David E., and Jon Richardson. 1987. "Genetic algorithms with sharing for multimodal function optimization". In: *Proceedings of the 2nd International Conference on Genetic Algorithms* (*ICGA 1987*), (Cambridge, MA, July 28-31 1987), p. 41-49. Hillsdale (NJ): Lawrence Erlbaum Associates.

Goldberg, David E. 1989. *Genetic Algorithms in Search, Optimization, and Machine Learning*, 1st ed. Reading (MA): Addison-Wesley, 412 p.

Goldberg, David E. 1990. "Probability matching, the magnitude of reinforcement, and classifier system bidding". *Machine Learning*, vol. 5, no° 4, p. 407-425.

Goldberg, David E., and Kalyanmoy Deb. 1990. "A comparative analysis of selection schemes used in genetic algorithms". In *Proceedings of the First Workshop on Foundations of Genetic Algorithms (FOGA 1990)*. (Bloomington, IN, July 15-18 1990), p. 69-93. San Francisco (CA): Morgan Kaufmann Publishers.

Good, I. J. 1982. "Diversity as a concept and its measurement: Comment". *Journal of the American Statistical Association*, vol. 77, no° 379, p. 561-563.

Gouvêa Jr., Maury M., and Aluizio F. R. Araújo. 2008. "Diversity control based on population heterozygosity dynamics". In *Proceedings of the IEEE Congress on Evolutionary Computation (CEC 2008)*, (Hong Kong, China, June 1-6 2008), p. 3671-3678. Piscataway (NJ): IEEE.

Grefenstette, John J. 1986. "Optimization of control parameters for genetic algorithms". *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-16, no° 1, p. 122-128.

Gupta, Anil K., Ken G. Smith, and Christina E. Shalley. 2006. “The interplay between exploration and exploitation”. *Academy of Management Journal*, vol. 49, no° 4, p. 693-706.

Hansen, Nikolaus and Andreas Ostermeier. 1996. “Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation”. In *Proceedings of 1996 IEEE International Conference on Evolutionary Computation* (*ICEC’96*). (Nagoya, Japan, May 20-22 1996), p. 312-317. Piscataway (NJ): IEEE.

Hansen, Nikolaus and Stefan Kern. 2004. “Evaluating the CMA evolution strategy on multimodal test functions”. In *Proceedings of the 8th Conference on Parallel Problem Solving from Nature* (*PPSN VIII*). (Birmingham, UK, September 18-22 2004), p. 282-291. Berlin (Germany): Springer-Verlag.

Hansen, Nikolaus. 2006. *Compilation of results on the 2005 CEC benchmark function set*. “CoLab technical report”, Zurich (Switzerland): ETH, 14 p.

Hansen, Nikolaus, Anne Auger, Steffen Finck, and Raymond Ros. 2009. *Real-parameter black-box optimization benchmarking 2009*: *Experimental setup*. “INRIA research report”, RR-6828. Orsay (France): INRIA, 17 p.

Havrda, Jan and Frantisek Charvát. 1967. “Quantification method of classification processes: Concept of structural a-entropy”. *Kybernetika*, vol. 3, no° 1, p. 30-35.

Herrera, Francisco, E. Herrera-Viedma, M. Lozano, and J. L. Verdegay. 1994. “Fuzzy tools to improve genetic algorithms”. In *Proceedings of the 2nd European Congress on Intelligent Techniques and Soft Computing (EUFIT 94).* (Aachen, Germany, September 20-23 1994), p. 1532-1539. Aachen (Germany): ELITE Foundation.

Herrera, Francisco and Manuel Lozano. 1996. “Adaptation of genetic algorithm parameters based on fuzzy logic controllers”. In *Genetic Algorithms and Soft Computing*, edited by Herrera, Francisco and J. L. Verdegay, 1st ed., p. 95-125. “Studies in Fuzziness and Soft Computing”, vol. 8. Heidelberg (Germany): Physica-Verlag.

Herrera, Francisco, Manuel Lozano, and J.L. Verdegay. 1998. “Tackling real-coded genetic algorithms: operators and tools for behavioural analysis”. *Artificial Intelligence Review*, vol. 12, no° 4, p. 265-319.

Hutter, Marcus, and Shane Legg. 2006. “Fitness uniform optimization”. *IEEE Transactions on Evolutionary Computation*, vol. 10, no° 5, p. 568-589.

Ichikawa, Yoshiaki and Yoshikazu Ishii. 1993. “Retaining diversity of genetic algorithms for multivariable optimization and neural network learning”. In *IEEE International Conference on Neural Networks.* (San Francisco, CA, March 28-April 1 1993), p. 1110-1114. Piscataway (NJ): IEEE.

Ishii, Shin, Wako Yoshida, and Junichiro Yoshimoto. 2002. “Control of exploitation-exploration meta-parameter in reinforcement learning”. *Neural Networks*, vol. 15, no° 4, p. 665-687.

Jin, R., W. Chen, and T.W. Simpson. 2001. “Comparative studies of metamodelling techniques under multiple modelling criteria”. *Structural and Multidisciplinary Optimization*, vol. 23, no° 1, p. 1-13.

Jin, Y. 2005. “A comprehensive survey of fitness approximation in evolutionary computation”. *Soft Computing*, vol. 9, no° 1, p. 3-12.

Julstrom, Bryant A. 1995. “What you have done for me lately? Adapting operator probabilities in a steady-state genetic algorithm”. In *Proceedings of the 6th International Conference on Genetic Algorithms (ICGA 1995)*. (Pittsburgh, PA, July 15-19 1995), p. 81-87. San Francisco (CA): Morgan Kaufmann Publishers.

Kennedy, James, Russel C. Eberhart, and Yuhui Shi. 2001. *Swarm Intelligence*, 1st ed. San Francisco (CA): Morgan Kaufmann Publishers, 512 p.

Kita, Hajime, Isao Ono, and Shigenobu Kobayashi. 1998. “Theoretical analysis of the unimodal normal distribution crossover for real-coded genetic algorithms”. In *IEEE World Congress on Evolutionary Computation*. (Anchorage, AK, May 4-9 1998), p. 529-534. Piscataway (NJ): IEEE.

Klee, Victor. 1977. “Can the measure of ∪[ai, bi] be computed in less than O(n log n) steps?”. *The American Mathematical Monthly*, vol. 84, no° 4, p. 284-285.

Konfrst, Zdenek. 2004. “Parallel genetic algorithms: advances, computing trends, applications and perspectives”. In *Proceedings of the 18th International Parallel and Distributed Processing Symposium* (*IPDPS’04*). (Santa Fe, NM, April 26-30 2004). Washington D.C.: IEEE Computer Society.

Kramer, Oliver. 2010. “Evolutionary self-adaptation: a survey of operators and strategy parameters”. *Evolutionary Intelligence*, vol. 3, no° 2, p. 51-65.

Lacevic, Bakir, Samim Konjicija, and Zikrija Avdagic. 2007. “Population diversity measure based on singular values of the distance matrix”. In *Proceedings of the IEEE Congress on Evolutionary Computation (CEC 2007)*, (Singapore, September 25-28 2007), p. 1863-1869. Piscataway (NJ): IEEE.

Lacevic, Bakir and Edoardo Amaldi. 2011. “Ectropy of diversity measures for populations in Euclidean space”. *Information Sciences*, vol. 181, no° 11, p. 2316-2339.

Larranaga Pedro and Jose A. Lozano (Eds.). 2002. *Estimation of distribution algorithms: A new tool for evolutionary computation*. 1st ed. Norwell (MA): Kluwer Academic Publishers, 382 p.

Lee, Michael A. and Hideyuki Takagi. 1993. "Dynamic control of genetic algorithms using fuzzy logic techniques". In *Proceedings of the 5th International Conference on Genetic Algorithms (ICGA 1993)*. (Urbana-Champaign, IL, July 1993), p. 76-83. San Francisco (CA): Morgan Kaufmann Publishers.

Lee, Jeho, Young U. Ryu. 2002. "Exploration, exploitation and adaptive rationality: the neo-Schumpeterian perspective". *Simulation Modelling Practice and Theory*, vol. 10, no° 5-7, p. 297-320.

Lieberson, Stanley. 1969 "Measuring population diversity". American Sociological Review, vol. 34, no° 6, p. 850-862.

Lobo, Fernando G., Claudio F. Lima, and Zbigniew Michalewicz. (Eds.). 2007. *Parameter setting in evolutionary algorithms* (Studies in Computational Intelligence, Vol. 54). Berlin (Germany): Springer-Verlag, 318 p.

Lozano, Manuel, Francisco Herrera, and José Ramón Cano. 2008. "Replacement strategies to preserve useful diversity in steady-state genetic algorithms". *Information Sciences*, vol. 178, no° 23, p. 4421-4433.

Lunacek, Monte and Darrell Whitley. 2006. "The dispersion metric and the CMA evolution strategy". In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 06)*. (Seattle, WA, July 8-12 2006), p. 477-484. New York (NY): ACM Press.

Lyons, N. I. and K. Hutcheson. 1978. "Comparing diversities: Gini's index". *Journal of Statistical Computation and Simulation*, vol. 8, no° 1, p. 75-80.

Mahfoud, Samir W. 1995. *Niching methods for genetic algorithms*. "Technical report from Illinois Genetic Algorithms Laboratory", IlliGAL 95001. Urbana-Champaign (IL): University of Illinois, 251 p.

Matsui, K. 1999. "New selection method to improve the population diversity in genetic algorithms". In *Proceedings 1999 IEEE International Conference on System, Man, and Cybernetics (SMC'99)*. (Tokyo, Japan, October 12-15 1999), p. 625-630. Piscataway (NJ): IEEE.

Mattiussi, Claudio, Markus Waibel, and Dario Floreano. 2004. "Measures of diversity for populations and distances between individuals with highly reorganizable genomes". *Evolutionary Computation*, vol. 12, no° 4, p. 495-515.

Maturana, Jorge and Frédéric Saubion. 2008. “A compass to guide genetic algorithms”. In *Proceedings of the 10th International Conference on Parallel Problem Solving from Nature* (*PPSN X*). (Dortmund, Germany, Sept. 13-17 2008), p. 256-265. Berlin (Germany): Springer-Verlag.

Maturana, Jorge, Frédéric Lardeux, and Frédéric Saubion. 2010. “Autonomous operator management for evolutionary algorithms”. *Journal of Heuristics*, vol. 16, no° 6, p. 881-909.

Mauldin, Michael L. 1984. “Maintaining diversity in genetic search”. In *Proceeding of the National Conference on Artificial Intelligence (AAAI-84)*. (Austin, TX, Aug. 6-10 1984), pp. 247-250. Palo Alto (CA): Association for the Advancement of Artificial Intelligence.

McPhee, Nicholas Freitag and Nicholas J. Hopper. 1999. “Analysis of genetic diversity through population history”. In *Proceedings of the Genetic and Evolutionary Computation Conference* (*GECCO 1999*). (Orlando, FL, July 13-17 1999), p. 1112-1120. San Francisco (CA): Morgan Kaufmann Publishers.

Mei-Yi, Li, Cai Zi-Xing, and Sun Guo-Yun. 2004. “An adaptive genetic algorithm with diversity-guided mutation and its global convergence property”. *Journal of Central South University of Technology*, vol. 11, no° 3, p. 323-327.

Miao, Ai-min, Xin-ling Shi, Jun-hua Zhang, En-yong Wang, and Shu-qing Peng. 2009. “A modified particle swarm optimizer with dynamical inertia weight”. In *Fuzzy Information and Engineering*, Edited by Cao, Bingyuan, Tai-Fu Li, and Cheng-Yi Zhang, p. 767-776. “Advances in Soft Computing”, vol. 62. Berlin (Germany): Springer-Verlag.

Molina, Daniel, Manuel Lozano, Carlos Garcia-Martinez, and Francisco Herrera. 2010. “Memetic algorithms for continuous optimization based on local search chains”. *Evolutionary Computation*, vol. 18, no° 1, p. 27-63.

Monsieurs, Patrick and Eddy Flerackers. 2003. “Reducing population size while maintaining diversity”. In *Proceedings of the 6th European Conference on Genetic Programming* (*EuroGP 2003*) *(Lecture Notes in Computer Science, vol. 2610)*. (Essex, UK, April 14-16 2003), p. 142-152. Heidelberg (Germany): Springer.

Morrison Ronald W. and Kenneth A. De Jong. 2002. “Measurement of population diversity”. In *Select. papers from the 5th European Conference Artificial Evolution* (*AE 2001*), (Le Creusot, France, October 29-31 2001), p. 31-41. London (UK): Springer-Verlag.

Motoki, Tatsuya. 2002. “Calculating the expected loss of diversity of selection schemes”. *Evolutionary Computation*, vol. 10, no° 4, p. 397-422.

Munawar, Asim, Mohamed Wahib, Masaharu Munetomo, and Kiyoshi Akama. 2008. "A survey: genetic algorithms and the fast evolving world of parallel computing". In *Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications* (*HPCC'08*). (Dalian, China, September 25-27 2008), p. 897-902. Piscataway (NJ): IEEE.

Nannen, Volker, S.K. Smit and A.E. Eiben. 2008. "Costs and Benefits of Tuning Parameters of Evolutionary Algorithms". In *Proceedings of the 10th International Conference on Parallel Problem Solving from Nature* (*PPSN X*). (Dortmund, Germany, September 13-17 2008), p. 528-538. Berlin (Germany): Springer-Verlag.

Nayak, Tapan Kumar. 1985. "On diversity measures based on entropy functions". *Communications in Statististics. Theory and Methods*, vol. 14, no° 1, p. 203-215.

Neri, Ferrante, Giuseppe Leonardo Cascella, Nadia Salvatore, Anna V. Kononova, and Giuseppe Acciani. 2006. "Prudent-daring vs. tolerant survivor selection schemes in control design of electric drives". In *Proceedings of the Application of Evolutionary Computation (EvoWorkshops 2006)*. (Budapest, Hungary, April 10-12 2006), p. 805-809. Berlin (Germany): Springer-Verlag.

Neri, Ferrante, Jari Toivanen, Giuseppe Leonardo Cascella, and Yew-Soon Ong. 2007. "An adaptive multimeme algorithm for designing HIV multidrug therapies". *IEEE/ACM Transaction on Computational Biology and Bioinformatics*, vol. 4, no° 2, p. 264-278.

Neri, Ferrante, Jari Toivanen, and Raino A. Mäkinen. 2007. "An adaptive evolutionary algorithm with intelligent mutation local searchers for designing multidrug therapies for HIV". *Applied Intelligence*, vol. 27, no° 3, p. 219-235.

Neri, Ferrante and Raino A. E. Mäkinen. 2007. "Hierarchical evolutionary algorithms and noise compensation via adaptation". In *Evolutionary Computation in Dynamic and Uncertain Environments*, Edited by Yang, Shengxiang, Yew-Soon Ong, and Yaochu Jin, p. 345-369. "Studies in Computational Intelligence", vol. 51. Berlin (Germany): Springer-Verlag

Neri, Ferrante, Niko Kotilainen, and Mikko Vapa. 2008. "A memetic-neural approach to discover resources in P2P networks" In *Recent Advances in Evolutionary Computation for Combinatorial Optimization*, Edited By Cotta, Carlos and Jano van Hemert, p. 113-129. "Studies in Computational Intelligence", vol. 153. Berlin (Germany): Springer-Verlag.

Nowostaski, Mariusz and Riccardo Poli. 1999. "Parallel genetic algorithm taxonomy". In *Proceedings of the Third International Conference on Knowledge-Based Intelligent Information Engineering Systems* (*KES 1999*). (Adelaide, South Australia, August 31 – September 1 1999), p. 88-92. Piscataway (NJ): IEEE.

Odetayo, Michael O. 1997. “Empirical study of the interdependencies of genetic algorithms parameters”. In *Proceedings of the 23rd EUROMICRO Conference*. (Budapest, Hungary, September 1-4 1997), p. 639-643. Piscataway (NJ): IEEE.

Olorunda, Olusegun and Andries P. Engelbrecht, “Measuring exploration/exploitation in particle swarms using swarm diversity”. In *Proceedings of the IEEE Congress on Evolutionary Computation (CEC 2008)*, (Hong Kong, China, June 1-6 2008), p. 1128-1134. Piscataway (NJ): IEEE.

Ono, Isao and Shigenobu Kobayashi. 1997. “A real-coded genetic algorithm for function optimization using unimodal normal distribution crossover”. In *Proceedings of the 7th International Conference on Genetic Algorithms (ICGA 1997)*. (East Lansing, MI, July 19-23 1997), p. 246-253. San Francisco (CA): Morgan Kaufmann Publishers.

Ozcan, Ender, Burak Bilgin, and Emin Erkan Korkmaz. 2008. “A comprehensive analysis of hyper-heuristics”. *Intelligent Data Analysis*, vol. 12, no° 1, p. 3-23.

Patil, G.P., and C. Taillie. 1982 “Diversity as a concept and its measurement”. Journal of the American Statistical Association, vol. 77, no° 379, p. 548-561.

Pavón, Reyes, Fernando Diaz, and Victoria Luzón. 2008. “A model for parameter setting based on Bayesian networks”. *Engineering Applications of Artificial Intelligence*, vol. 21, no° 1, p. 14-25.

Pavón, Reyes, Fernando Diaz, Rosalia Laza, and Victoria Luzón. 2009. “Automatic parameter tuning with a Bayesian case-based reasoning system. A case of study”. *Expert Systems with Applications*, vol. 36, no° 2, p. 3407-3420.

Pearl, Judea. 1988. *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Revised 2nd printing. San Francisco (CA): Morgan Kaufmann Publishers, 552 p.

Pellerin, Eric, Luc Pigeon, and Sylvain Delisle. 2004. “Self-adaptive parameters in genetic algorithms”. In *Proceedings of the SPIE – Data Mining and Knowledge Discovery: Theory, Tools, and Technology VI (vol. 5433)*. (Orlando, FL, April 12 2004), p. 53-64. Bellingham (WA): SPIE.

Peng Fei, Ke Tang, Guoliang Chen, and Xin Yao. 2010. “Population-based algorithm portfolios for numerical optimization”. *IEEE Transactions on Evolutionary Computation*, vol. 14, no° 5, p. 782-800.

Price, Kenneth, Rainer M. Storn and Jouni A. Lampinen. 2005. *Differential Evolution: A Practical Approach to Global Optimization*, 1st ed. Berlin (Germany): Springer-Verlag, 538 p.

Rechenberg, Ingo. 1973. *Evolutionsstrategie: Optimierung technischer systeme nach prinzipien der biologischen evolution*. 1st ed. Stuttgart (Germany): Frommann-Holzboog. 170 p.

Rényi, Alfréd. 1961. "On measures of entropy and information". In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*. (Berkeley, CA, June 20-July 30 1961), p. 547-561. Berkeley (CA): University of California Press.

Riget, Jacques and Jakob S. Vesterstrom. 2002. *A diversity-guided particle swarm optimizer – the ARPSO*. "EVALife Technical Report", 2002-02. Aarhus C (Denmark): University of Aarhus, 13 p.

Rojas, Ignacio, Jesus Gonzalez, Héctor Pomares, J. J. Merelo, P. A. Castillo, and G. Romero. 2002. "Statistical analysis of the main parameters involved in the design of a genetic algorithm". *IEEE Transaction on Systems Man, and Cybernetics – Part C: Applications and Reviews*, vol. 32, no° 1, p. 31-37.

Rosca, Justinian P. 1995. "Entropy-driven adaptive representation". In *Proceedings of the Workshop on Genetic Programming: From Theory to Real-World Applications*. (Tahoe City, CA, July 9-12 1995), p. 23-32. New York: University of Rochester.

Sareni, Bruno and Laurent Krähenbühl. 1998. "Fitness sharing and niching methods revisited". *IEEE Transactions on Evolutionary Computation*, vol. 2, no° 3, p. 97-106.

Satoh, Hiroshi, Massayuki Yamamura, and Shigenobu Kobayashi. 1996. "Minimal generation gap model for Gas considering both exploration and exploitation". In *Proceedings of the Fourth International Conference on Soft Computing* (*IIZUKA'96*). (Iizuka, Fukuoka, Japan, September 30 – October 5 1996), p. 494-497. World Scientific Publishing.

Schaffer, David J., Richard A. Caruana, Larry J. Eshelman, and Rajarshi Das. 1989. "A study of control parameters affecting online performance of genetic algorithms for function optimization". In *Proceedings of the Third International Conference on Genetic Algorithms* (*ICGA 1989*). (Fairfax, VA, June 1989), p. 51-60. San Francisco (CA): Morgan Kaufmann Publishers Inc.

Shannon, Claude E. 1948 "A mathematical theory of communication". *Bell System Technical Journal*, vol. 27, (July, Oct.), p. 379-423, 623-656.

Simpson, E. H. 1949. "Measurement of diversity". *Nature*, vol. 163, p. 688.

Sinha, Ankur, Santosh Tiwari, and Kalyanmoy Deb. 2005. "A population-based, steady-state procedure for real-parameter optimization". In *Proceedings of the IEEE Congress on*

Evolutionary Computation (*CEC 2005*). (Edinburgh, UK, September 2-4 2005), p. 514-521. Piscataway (NJ): IEEE.

Smith, J.E. and T.C. Fogarty. 1997. "Operator and parameter adaptation in genetic algorithms". *Soft Computing – A Fusion of Foundations, Methodologies and Applications*, vol. 1, no° 2, p. 81-87.

Solow, Andrew R., and Stephen Polasky. 1994. "Measuring biological diversity". *Environmental and Ecological Statistics*, vol. 1, no° 2, p. 95-107.

Sprent, Peter and Nigel C. Smeeton. 2000. *Applied nonparametric statistical methods*, 3rd ed. Boca Raton (FL): Chapman & Hall/CRC, 480 p.

Srinivas, M. and L.M. Patnaik. 1994. "Adaptive probabilities of crossover and mutation in genetic algorithms". *IEEE Transactions on Systems, Man and Cybernetics*, vol. 24, no° 4, p. 656-667.

Stephens, C. R., I. Garcia Olmedo, J. Mora Vargas, H. Waelbroeck. 1998. "Self-adaptation in evolving systems". *Artificial Life*, vol. 4, no° 2, p. 183-201.

Subbu, Raj, Arthur C. Sanderson, and Piero P. Bonissone. 1998. "Fuzzy logic controlled genetic algorithms versus tuned genetic algorithms: an agile manufacturing application" In *Proceedings of the 1998 IEEE ISIC/CIRA/ISAS Joint Conference*. (Gaithersburg, MD, September 14-17 1998), p. 434-440. Piscataway (NJ): IEEE.

Suganthan P. N. *et al*. 2005. *Problem definitions and evaluation criteria for the CEC 2005 special session on real parameter optimization*. "KanGAL Technical Report", 2005005. Singapore: Nanyang Technological University, 50 p.

Sywerda, Gilbert. 1989. "Uniform crossover in genetic algorithms". In *Proceedings of the 3rd International Conference on Genetic Algorithms (ICGA 1989)*. (Fairfax, VA, June 1989), p. 2-9. San Francisco (CA): Morgan Kaufmann Publishers Inc.

Thierens, Dirk. 2005. "An adaptive pursuit strategy for allocating operator probabilities". In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 05)*. (Washington DC, June 25-29 2005), p. 1539-1546. New York (NY): ACM Press.

Tirronen, Ville, Ferrante Neri, Tommi Karkkainen, Kirsi Majava, and Tuomo Rossi. 2007. "A memetic differential evolution in filter design for defect detection in paper production" In *Proceedings of the Application of Evolutionary Computation (EvoWorkshops 2007)*. (Valencia, Spain, April 11-13 2007), p. 320-329. Berlin (Germany): Springer-Verlag.

Tirronen, Ville, and Ferrante Neri. 2009. "Differential evolution with fitness diversity self-adaptation". In *Nature-Inspired Algorithms for Optimisation*, edited by Chiong,

Raymond, p. 199-234. “Studies in Computational Intelligence”, Vol. 193. Berlin (Germany): Springer-Verlag.

Turkey, Mikdam, and Riccardo Poli. 2012. “An empirical tool for analysing the collective behaviour of population-based algorithms”. In: *Proceedings of the Application of Evolutionary Computation (EvoApplications 2012)*, (Malaga, Spain, April 11-13 2012), p. 103-113. Heidelberg (Germany): Springer.

Tuson Andrew and Peter Ross. 1998. “Adapting operator settings in genetic algorithms”. *Evolutionary Computation*, vol. 6, no° 2, p. 161-184.

Ulrich, Tamara, Johannes Bader, and Lothar Thiele. 2010. “Defining and optimizing indicator-based diversity measures in multiobjective search”. In *Proceedings of the 11th Conference on Parallel Problem Solving from Nature (PPSN 2010).* (Krakow, Poland, September 11-15 2010), p. 707-717. Berlin (Germany): Springer-Verlag.

Ursem, Rasmus K. 1999. “Multinational evolutionary algorithms”. In *Proceedings of the IEEE Congress on Evolutionary Computation (CEC 1999)*. (Washington DC, July 6-9 1999), p. 1633-1640. Piscataway (NJ): IEEE.

Ursem, Rasmus K. 2002. “Diversity-guided evolutionary algorithms”. In *Proceedings of the 7th Conference on Parallel Problem Solving from Nature (PPSN 2002).* (Granada, Spain, September 7-11 2002), p. 462-471. Berlin (Germany): Springer-Verlag.

Vasconcelos, J.A., J.A. Ramirez, R.H.C. Takahashi, and R.R. Saldanha. 2001. “Improvements in genetic algorithms”. *IEEE Transactions on Magnetics*, vol. 37, no° 5, p. 3414-3417.

Weitzman, Martin L. 1992. “On diversity”. *The Quarterly Journal of Economics*, vol. 107, no° 2, p. 363-405.

Whitacre, James M., Tuan Q. Pham, and Ruhul A. Sarker. 2006. “Use of statistical outlier detection method in adaptive evolutionary algorithms”. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 06)*. (Seattle, WA, July 8-12 2006), p. 1345-1352. New York (NY): ACM Press.

Whitacre, James M. 2011a. “Recent trends indicate rapid growth of nature-inspired optimization in academia and industry”. *Computing*, vol. 93, no° 2-4, p. 121-133.

Whitacre, James M. 2011b. “Survival of the flexible: explaining the recent popularity of nature-inspired optimization within a rapidly evolving world”. *Computing*, vol. 93, no° 2-4, p. 135-146.

Wineberg, Mark and Franz Oppacher. 2003 “The underlying similarity of diversity measures used in evolutionary computation”. In *Proceedings of the Genetic and Evolutionary*

Computation Conference (GECCO 03). (Chicago, IL, July 12-16 2003), p. 1493-1504. New York (NY): ACM Press.

Wineberg, Mark and Steffen Christensen. 2007. "An introduction to statistical analysis for evolutionary computation". In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 07)*. (London, England, July 7-9 2007), p. 3765-3791. New York (NY): ACM Press.

Wolpert, David H. and William G. Macready. 1997. "No free lunch theorems for optimization". *IEEE Transactions on Evolutionary Computation*, vol. 1, no° 1, p. 67-82.

Yoon, Hyun-Sook and Byung-Ro Moon. 2002. "An Empirical Study on the Synergy of Multiple Crossover Operators". *IEEE Transactions on Evolutionary Computation*, vol. 6, no°2, p. 212-223.