Recherche de carte d'idéotypes de sorgho d'après un modèle de culture : optimisation conditionnelle à l'aide d'un métamodèle de krigeage

Diariétou Sambakhé

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Biostatistique

École doctorale Information, Structures, Systèmes

Unité de recherche CIRAD-Agroécologie et intensification durable des cultures annuelles

Recherche de carte d'idéotypes de sorgho d'après un modèle de culture : optimisation conditionnelle à l'aide d'un métamodèle de krigeage

Présentée par Diariétou Sambakhé

Le 25 Avril 2018

Sous la direction de Jean-Noël Bacro

Devant le jury composé de

Costes Evelyne, Directrice de Recherche, INRA, Montpellier	Rapporteur
Faivre Robert, Directeur de Recherche, INRA, Toulouse	Examinateur
Gozé Eric, Cadre de Recherche, CIRAD,Montpellier	Co-encadrant
Iooss Bertrand, Chercheur senior, EDF R&D, Chatou	Rapporteur
Rouan Lauriane, Cadre de Recherche, CIRAD, Montpellier	Co-encadrant
Bacro Jean-Noël, Professeur, IMAG, Montpellier	Directeur

UNIVERSITÉ DE MONTPELLIER

Remerciements

Je tiens à remercier chaleureusement toutes les personnes qui ont contribué à l'aboutissement de cette thèse. Mes premiers remerciements vont vers mon directeur de thèse Jean-Noel Bacro et mes encadrants Eric Gozé et Lauriane Rouan. Merci pour votre confiance, votre disponibilité, votre rigueur scientifique et surtout pour votre gentillesse.

Ce travail n'aurait pas été possible sans le soutien du Programme de Productivité Agricole en Afrique de l'Ouest PPAAO et de ISRA/CERAAS, qui m'ont permis, grâce à une allocation de re-cherches et diverses aides financières, de me consacrer sereinement à l'élaboration de ma thèse.
Je remercie chaleureusement Evelyne Costes et Bertrand Iooss pour avoir accepté d'être les rapporteurs de ma thèse, et qui ont bien voulu me faire part de leurs pertinentes remarques et suggestions. Je remercie par la même occasion Robert Faivre pour avoir accepté d'être le président du jury de ma soutenance.
Je remercie également les membres de mon comité de thèse, Victor Picheny, Catherine Trottier, Benjamin Sultan et Philippe Letourmy , grâce à leurs précieux conseils, j'ai pu re-cadrer la thèse.
Je remercie chaleureusement Michael Dingkuhn, Myriam Adam, et Christian Baron pour avoir mis à notre disposition leurs connaissances et leurs données expérimentales.
Ce travail n'aurait pu être mené à bien sans la disponibilité et l'accueil chaleureux que m'ont témoi-gné mes collègues du CIRAD. Je remercie particulièrement les chefs des unités AIDA et AGAP. Je tiens à remercier spécialement Edward Gérardeaux, Jean-Paul Gourlot, Janine Jane, Sabrina Azouni, Anne-Laure Fruteau de Laclos pour leur gentillesse et leur bienveillance à mon égard.
Merci à Romain Loison pour les nombreuses discussions que nous avons échangées, ainsi que pour ses précieux conseils concernant la rédaction de ma thèse et la préparation de ma soutenance. Merci aussi à Ana Lopez et Oriol pour m'avoir fait découvrir les différents restaurants de Montpellier.
Je voudrais également remercier tous les doctorants de UMR AIDA avec qui j'ai passé de bons moments, et qui d'une manière ou d'une autre ont apporté leur contribution à ce travail.
Je remercie mes amies Diary Ba et Christine Boutavin pour m'avoir soutenu et encouragé.

Je remercie enfin celles et ceux qui me sont chers et que j'ai en quelque sorte négliger

pendant ces trois années de thèse. Leurs attentions et encouragements m'ont accompagnée tout au long de ces années. Je suis particulièrement redevable à ma mère, Fatou Tall , pour son soutien moral et pour sa confiance indéfectible dans mes choix. Enfin, j'ai une pensée toute particulière pour mon défunt père, Mouraby Sambakhé, je te porterai toujours dans mon cœur.

Je dédie ce travail à mon cher époux Ousmane Diop et à mes adorables enfants Mouhamed et Zakaria, pour leurs amours et leurs patiences durant toutes ces années d'absence. Je ne saurais jamais vous exprimer ma profonde gratitude.

Table des matières

1 Introduction **1**

2 Choix variétal pour un ensemble d'environnements : quels outils ? **9**

2.1 Analyse statistique de l'interaction génotype × environnement 10

2.1.1 Modèle d'analyse de la variance 10

2.1.2 Méthodes descriptives 13

2.1.3 Une méthode explicative: la régression factorielle 14

2.2 Modélisation d'une culture 15

2.2.1 Croissance et développement de la plante 15

2.2.2 Modèle de culture 16

2.3 Prédiction avec une météo encore inconnue 21

2.3.1 Générateur stochastique de climat 22

3 Méthodes d'optimisation **25**

3.1 Principales méthodes d'optimisation locales 26

3.1.1 Algorithme de descente 27

3.1.2 Méthode de Newton 28

3.1.3 Méthode du gradient stochastique 29

3.1.4 Les méthodes de recherche directe 30

3.2 Principales méthodes d'optimisation globale 31

3.2.1 Algorithme évolutionnaire 32

3.2.2 Le recuit simulé 32

3.2.3 Optimisation à base de métamodèle 33

3.3 Quelle méthode pour la recherche de génotype optimal ? 34

3.3.1 Notions sur le krigeage 35

3.3.2 Choix du critère d'échantillonnage pour l'optimisation à base de métamodèle 38

3.3.3 Optimisation globale d'une fonction déterministe 41

3.3.4 Optimisation conditionnelle d'une fonction déterministe 45

3.3.5 Optimisation globale d'une fonction bruitée 46

4 Définition d'un nouveau critère pour l'optimisation conditionnelle d'une fonction bruitée 49

4.1 Profil d'amélioration espérée sur les quantiles $(PEQI)$ 50

4.2 Algorithme d'optimisation pour la recherche d'une ligne de crête 52

4.3 Preuve de convergence 53

4.4 Application sur des fonctions tests et comparaison de $PEQI$ avec PEI . 58

4.5 Comparaison avec le profil d'amélioration espérée selon différents facteurs expérimentaux et algorithmiques 60

5 Application sur un modèle de culture 69

5.1 Couplage modèle SAMARA et générateur stochastique de climat Marksim 70

5.1.1 Le modèle SAMARA 70

5.1.2 Générateur stochastique de climat Marksim 71

5.2 Détermination d'une carte d'idéotypes de sorgho 73

5.2.1 Choix des paramètres pour l'optimisation 73

5.2.2 Étude de la sensibilité de la distribution du rendement 74

5.3 Mise en œuvre de l'algorithme d'optimisation à base de $PEQI$ pour la détermination de la carte d'idéotypes 85

5.4 Conclusion 94

6 Conclusion et perspectives 95

7 Annexes **97**

7.1 Quelques notions de topologie 97

7.1.1 Espace de Hilbert à noyau reproduisant 97

7.1.2 Point adhérent 97

7.1.3 Partie dense 98

7.1.4 Suite extraite ou sous suite 98

7.1.5 Inégalité de Cauchy-Schwarz 98

7.2 Quantile de krigeage 98

7.2.1 Propriétés du quantile de krigeage 98

7.2.2 Mise à jour du quantile de krigeage par l'ajout d'un point 99

8 Annexes **101**

8.1 Propriétés du quantile de krigeage 101

8.2 Mise à jour du quantile de krigeage par l'ajout d'un point 102

Table des figures

1.1 Gradient Nord Sud des valeurs de précipitations 2

2.1 (A) pas d'interaction, (B) interaction quantitative, (C) interaction quantitative, (D) interaction qualitative 11

2.2 Illustration d'un modèle de culture 21

3.1 Prédiction par le krigeage et erreur de prédiction 36

3.2 Illustration d'une optimisation avec l'algorithme EGO. La fonction objectif (trait plein bleu), le prédicteur (trait plein vert), les intervalles de confiance à 95% construits à partir de l'erreur type de prédiction (traits en pointillé), les points d'échantillonnage (points rouges) et la position du prochain point à évaluer (ligne verticale pointillée, correspondant au maximum du critère). 44

3.3 Nouveau critère pour l'optimisation conditionnelle d'une fonction bruitée 48

4.1 Échantillonnage de points avec le critère $PEQI$ 51

4.2 Lignes de crête réelles et prédites en utilisant le critère $PEQI$ avec $\beta = 0.1$. Des niveaux de bruits de 20 et 50% de la variance des réponses de la fonction test sont testés. 59

4.3 PEI vs PEQI. RMSE de la ligne de crête en fonction du niveau de bruit et du critère d'échantillonnage. Pour le critère $PEQI$, β est fixé à 0.1. . . 61

4.4 Influence de β sur $PEQI$. RMSE de la ligne de crête en fonction du niveau de bruit et de la taille du plan initial. Pour des plans initiaux de 20 et 40 points, 50 points sont ajoutés séquentiellement à l'aide du critère PEI 63

4.5 Influence de la taille du plan initial sur PEI. RMSE de la ligne de crête en fonction du niveau de bruit et de la taille du plan initial. À des plans initiaux de 20 et 40 points, 50 points sont ajoutés séquentiellement à l'aide du critère PEI 65

4.6 Influence de la taille du plan initial sur $PEQI$. RMSE de la ligne de crête en fonction du niveau de bruit et de la taille du plan initial. À des plans initiaux de 20 et 40 points, 50 points sont ajoutés séquentiellement à l'aide du critère $PEQI$ avec β = 0.1 66

5.1 Disposition des différentes valeurs des différents paramètres 75

5.2 Pluviométrie moyenne annelle en (mm) 76

5.3 Sensibilité de la distribution des rendements à 4 paramètres près de Catio (Guinée Bissao) 77

5.4 Sensibilité de la distribution des rendements à 4 paramètres près de Ziguinchor (Sénégal) 79

5.5 Sensibilité de la distribution des rendements à 4 paramètres près de Nioro du Rip (Sénégal) 81

5.6 Sensibilité de la distribution des rendements à 4 paramètres près de Diourbel (Sénégal) 83

5.7 Sensibilité de la distribution des rendements à 4 paramètres près de Louga (Sénégal) 84

5.8 Couplage du modèle de culture avec le modèle de climat 85

5.9 Carte des valeurs optimales de la longueur du cycle (ndjbvp) pour le rendement moyen 86

5.10 Carte des valeurs optimales de la longueur des racines (rootfrontmax) pour le rendement moyen 87

5.11 Carte des valeurs optimales de la mortalité des feuilles (leaf.death) pour le rendement moyen 87

5.12 Carte des valeurs optimales de la capacité de réserves dans les tiges (stem.reserve) pour le rendement moyen 88

5.13 Sensibilité de la distribution des rendements à 4 paramètres près de Kassaro 89

5.14 Sensibilité de la distribution des rendements à 4 paramètres près de Kobenri 91

5.15 Sensibilité de la distribution des rendements à 4 paramètres près de Timbedra 92

5.16 Sensibilité de la distribution des rendements à 4 paramètres près de près de Niono 93

I

Introduction

Contexte et objectif

La croissance de la population mondiale suscite le besoin perpétuel d'augmenter la production agricole. Au Sahel, où l'incidence de la malnutrition est la plus élevée au monde [103], la saturation des espaces cultivés exclut le recours à la mise en culture de nouvelles terres comme au raccourcissement des jachères : il faut augmenter la production par unité de surface cultivée, ce que les agronomes appellent rendement.

L'augmentation du rendement passe par l'amélioration des plantes et de leur environnement modifié par les pratiques culturales (travail du sol, fertilisation, gestion des mauvaises herbes etc). Les meilleures plantes ne sont pas nécessairement les mêmes pour toutes les situations : il faut que chacune soit adaptée à son environnement. C'est pourquoi la plupart des espèces végétales cultivées existent sous la forme de nombreuses variétés, adaptées tant à l'environnement qu'elles occupent qu'à l'usage auquel on destine leur production.

Dans les pays développés, autant pour des raisons de commodité de commercialisation que de structuration géographique de l'environnement, celui-ci est divisé en méga-environnements, supposés homogènes.
Au Sahel, la réalité est différente : l'environnement est fortement contraint par la disponibilité en eau dans le sol, et celle-ci décrit un gradient continu depuis le désert du Sahara au Nord jusqu'à la zone climatique dite soudanienne au Sud où l'eau n'est plus un facteur limitant (figure 1.1).

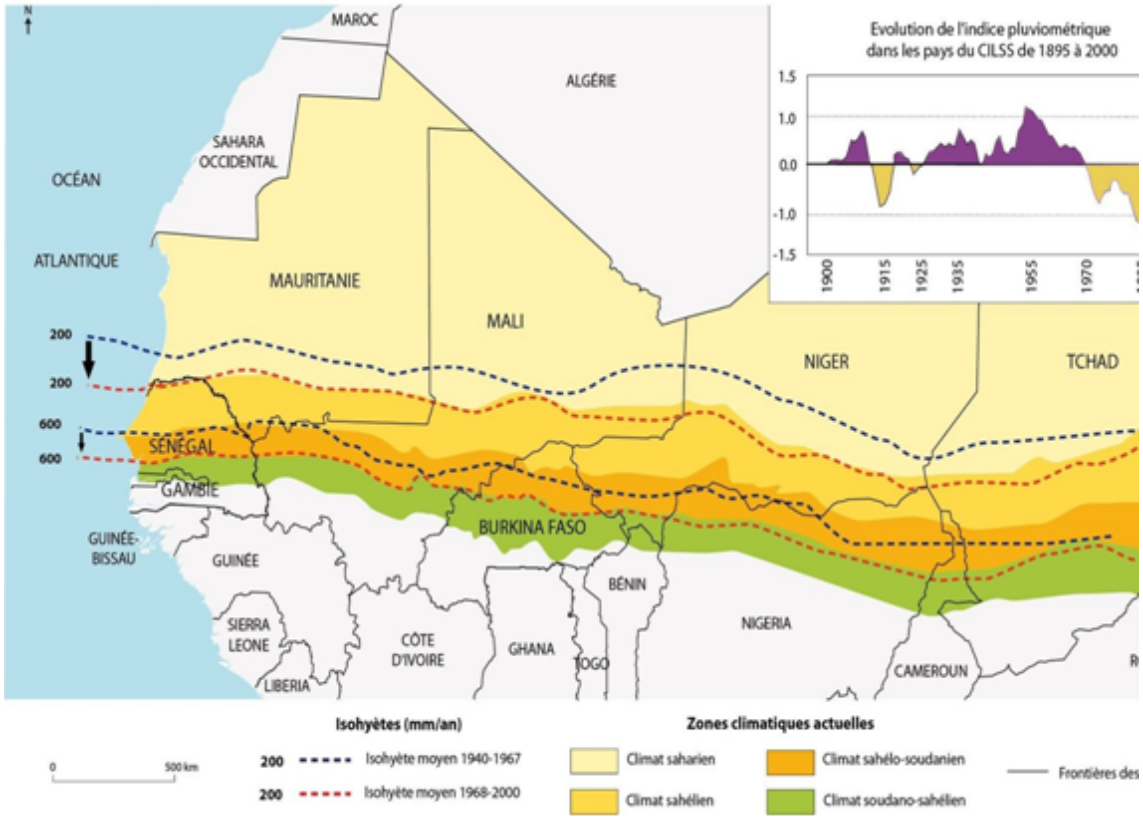


FIGURE 1.1 – *Gradient Nord Sud des valeurs de précipitations*

Traditionnellement, ces sont les paysans eux-mêmes qui pratiquent la sélection végétale en gardant le meilleur de leur production comme semence, et les ressources génétiques s'échangent entre familles au gré des alliances [91]. Depuis les découvertes scientifiques sur l'hérédité [146], des spécialistes de l'amélioration génétique se sont peu à peu substitués aux paysans dans cette tâche, et ont fait progresser considérablement les rendements, conjointement à l'amélioration et l'homogénéisation de l'environnement apportées par les spécialistes du sol, de la protection des plantes et des systèmes de culture.

L'amélioration génétique comprend plusieurs phases : la définition d'un cahier des charges, la recherche de géniteurs présentant des traits d'intérêt, la réalisation de croisements qui vont combiner aléatoirement les différents allèles des gènes déterminant ces traits dans la descendance, et la sélection dans celle-ci des meilleurs spécimens.

La phase d'évaluation des variétés ainsi créées est imprécise, car l'environnement auquel elles doivent s'adapter change d'une année à l'autre : la variété la meilleure une année ne le sera peut-être pas l'année suivante ou les années à venir; on dit qu'il y a interaction entre génotype et année. Si celle-ci pose déjà des problèmes pour l'homologation des variétés dans les pays tempérés, elle en pose d'autant plus pour leur évaluation au Sahel où les années ne se ressemblent pas. En effet, la production est, on l'a dit, conditionnée par la saison des pluies, et celles-ci sont aussi très irrégulières et variables d'une année à l'autre. L'observation d'un rendement moyen est donc peu reproductible, et l'estimation d'un rendement peu précise, sauf à renouveler l'observa-

tion en de nombreuses occasions. Il est donc difficile de choisir entre plusieurs variétés par observation directe.

Une alternative moins empirique serait de repérer sur la plante des caractères faciles à observer qui permettraient de prédire la réponse de la plante aux variations de l'environnement. Par exemple, une variété qui enfouit profondément ses racines est plus susceptible que d'autres de résister à de longs épisodes sans pluie. L'observation directe de cette résistance serait soumise au hasard de l'existence ou non de tels épisodes, alors que la mesure de la longueur des racines est reproductible, pour autant que l'on définisse bien les conditions dans lesquelles on l'opère.

Cette démarche de construction a priori d'une variété en dressant la liste des caractères souhaitables en fonction des connaissances que l'on a de l'environnement et du fonctionnement de la plante est la conception d'un idéotype. Donald [34] a introduit cette notion et construit un idéotype de céréale par pur raisonnement. Cependant, le fonctionnement d'une plante est complexe, si bien que la recherche d'idéotypes par le simple raisonnement est une affaire d'expert, forcément subjective. La formalisation mathématique des phénomènes physiologiques et la simulation des équations qui en découlent à l'aide d'ordinateurs afin de prédire la réponse à l'environnement de la plante fait l'objet de modèles de culture de diverses sortes [92]. Un modèle de culture décrit le fonctionnement d'un couvert végétal cultivé et sa réponse à son environnement. Il repose sur des équations mathématiques faisant intervenir les variables de l'environnement et de la plante. Certains modèles sont assez génériques pour couvrir plusieurs variétés, voire plusieurs espèces [1]: des paramètres dits variétaux permettent de modifier les réponses du modèle à l'environnement.
On peut donc voir dans un modèle de culture le moyen objectif de définir la plante idéale : c'est celle dont les paramètres lui permettent de remplir au mieux les conditions du cahier des charges dans l'environnement-cible auquel on la destine. Si par exemple le cahier des charges est d'obtenir un rendement élevé dans des conditions environnementales données qui dépendent du lieu, on va rechercher les valeurs des paramètres variétaux qui maximisent le rendement pour ces conditions-là.
Pour le même jeu de paramètres variétaux, à chaque fois que l'on change de saison des pluies la sortie du modèle change. Donc pour des conditions à venir, que l'on ne connaît pas encore, mais dont on connaît la distribution de probabilité, on peut chercher à maximiser l'espérance du rendement, à minimiser sa variance, ou à maximiser un compromis qui peut être un quantile de la production, par exemple le rendement dépassé huit années sur dix. Cependant l'estimation de ces moments ou quantiles repose en pratique sur un nombre limité de simulations; elle est donc affectée d'une erreur. On a ainsi un problème d'optimisation d'une fonction bruitée à résoudre.
Comme on l'a énoncé plus haut, il y a un gradient Nord Sud des valeurs de précipitations, la meilleure variété change donc quand on se déplace du Nord au Sud. Si maintenant on veut déterminer un gradient des meilleures variétés sur un gradient d'environnements cibles, on se retrouve devant un problème d'optimisation conditionnellement à l'environnement. Calculer l'optimum indépendamment en chaque lieu u prendrait beaucoup de temps. Pour un sol donné, les variations du climat étant lentes dans l'espace, on peut supposer que l'optimum en un lieu $u + \delta$ peu éloigné de u est proche de l'optimum en u. Une possibilité est alors d'ajuster une surface de réponse de façon à prendre en compte cette dépendance. La surface de réponse est une approximation du modèle de culture, et ce *modèle de modèle* est appelé métamodèle.

L'utilisation de métamodèles facilite l'optimisation des modèles complexes et coûteux à évaluer [75, 4, 54]. A la différence des algorithmes d'optimisation locaux (Nelder-Mead [101], la méthode GPS [66], etc), une optimisation globale reposant sur un métamodèle garde la mémoire de toutes les évaluations du modèle, et en nécessite moins que les algorithmes stochastiques (recuit simulé, génétique,...). Il existe plusieurs types de métamodèle parmi lesquels on peut citer : les interpolateurs locaux, les techniques polynomiales, les splines, les réseaux de neurones et le krigeage [54].
L'optimisation à base de métamodèle de krigeage est utilisée avec succès pour l'optimisation de fonctions coûteuses dans le domaine industriel [10, 111, 71]. Dans le domaine agricole, le métamodèle de krigeage pourrait avoir un intérêt non pas parce que le modèle est coûteux mais parce que l'interaction variété x lieu nécessite de produire un optimum par lieu, et pour cela une optimisation conditionnelle au lieu.

Le sorgho étant l'une des céréales les plus consommées au Sahel, l'objectif de cette thèse est de trouver une carte d'idéotypes de sorgho adaptés au climat sahélien présent et futur en utilisant un modèle de culture. Plus spécifiquement, notre objectif est de trouver pour chaque environnement u fixé, le jeu de paramètres v qui maximise l'espérance ou un quantile d'une fonction f bruitée, autrement dit une carte ou lieu des optimums. Le chapitre 2 met en évidence les limites des outils statistiques pour prédire l'interaction génotype $\times$ environnement, et les outils de modélisation utilisés aussi dans ce but. Dans le chapitre 3, nous faisons un rappel sur le principe de l'optimisation à base de métamodèle et sur le choix du critère d'échantillonnage suivant les différents buts poursuivis: optimisation globale de fonction déterministe, optimisation conditionnelle de fonction déterministe et optimisation globale de fonction bruitée. Nous introduisons en chapitre 4 une variante du critère d'amélioration espérée (EI) pour un problème d'optimisation conditionnelle d'une fonction bruitée. Ce nouveau critère, nommé profil d'amélioration espérée sur les quantiles $(PEQI)$, est une généralisation du critère profil d'amélioration espérée (PEI) défini par Ginsbourger *et al*. [55] pour l'optimisation d'un modèle conditionnellement à une partie de ses paramètres. Nous démontrons sa convergence et différents exemples d'applications de l'algorithme sont présentés sur des fonctions test pour en illustrer l'intérêt. Nous comparons empiriquement la méthode proposée avec la méthode basée sur le critère de profil d'amélioration espérée (PEI). Enfin dans le chapitre 5, nous présentons une application du nouveau critère $PEQI$ pour la recherche d'une carte d'idéotypes de sorgho au Sahel en utilisant le modèle de culture SAMARA [33].

Notations

Symboles

- u coordonnées géographiques (latitude et longitude)
- v paramètres représentant la variété
- $x = (u, v)$
- C_u modèle stochastique du climat
- f fonction à optimiser
- K matrice de covariance
- k fonction de covariance
- γ semi variogramme
- m moyenne du processus gaussien Y
- μ fonction moyenne du modèle de krigeage
- σ^2 variance du processus gaussien Y
- β le fractile
- q_n le quantile de krigeage
- Q_{n+1} quantile de krigeage attendu avec la mise à jour du plan d'expérience avec l'ajout d'un nouveau point
- $\underline{x}_n = (x_1, \cdots, x_n)$
- $\hat{\xi}_{n+1}(x) = \hat{\xi}_{n+1}(x, f, \underline{x}_n)$ l'espérance conditionnelle de $Q_{n+1}(x)$ sachant $\tilde{Y}(x_1), \cdots \tilde{Y}(x_n)$
- Y processus gaussien
- $\tilde{Y}$ somme d'un processus et d'un bruit gaussiens
- τ^2 la variance du bruit
- $m_n(x)$ estimation de la moyenne du modèle de krigeage au point x
- $s_n^2(x)$ estimation de la variance du modèle de krigeage au point x
- $s_{Q_{n+1}}^2(x)$ estimation de la variance du quantile de krigeage au point x
- Φ fonction de répartition de la loi normale centrée réduite
- ϕ densité de la loi normale centrée réduite
- ε terme résiduel dans les modèles statistiques
- ϵ bruit blanc ajouté à la fonction test f
- $y_{max,n}$ la valeur du maximum courant

- $y^*_{max,n}$ maximum observé sur le plan d'expérience initial
- $q^*_{max,n}$ maximum des quantiles observés sur le plan d'expérience initial
- ψ expression analytique du critère $PEQI$

Abréviations

- ACP Analyse en Composantes Principales
- $AMMI$ Additive Main effect and Multiplicative Interaction
- EI Expected Improvement
- EGO Efficient Global Optimization
- AEI Augmented Expected Improvement
- AKG Approximate Knowledge Gradient
- MQ Minimal Quantile
- EQI Expected Quantile Improvement
- RI Reinterpolation procedure
- PEI Profile Expected Improvement
- $PEQI$ Profile Expected Quantile Improvement
- LHS Latin Hypersquare Sampling
- $G \times E$ Interaction génotype $\times$ environnement
- MSE Mean Square Error
- $RMSE$ Root Mean Square Error
- $IMSE$ Integrated Mean Squared Error
- $RRMSE$ Relative RMSE
- $RKHS$ Hilbert space with reproducing kernel
- NEB no-empty-ball

Quelques définitions en agronomie

- Le terme variété permet de différentier les individus au sein d'une même espèce: les variétés d'une même espèce sont des populations distinctes, homogènes et stables.
- Le génotype peut correspondre à une variété, ou bien à du matériel végétal en cours de sélection, qui n'a pas la stabilité d'une variété car il ne se reproduit pas à l'identique.
- Le milieu désigne le milieu naturel, sa fertilité en général, ses aptitudes culturales particulières ou plus précisément certaines contraintes ou conditions pédoclimatiques.
- Lorsque la performance du génotype est fonction du milieu, on parlera d'interaction génotype $\times$ milieu ou d'interaction génotype $\times$ environnement.

II

Choix variétal pour un ensemble d'environnements : quels outils ?

De nombreux facteurs limitants comme l'eau, la température, l'azote du sol ou encore le rayonnement peuvent affecter le rendement des cultures tout au long du cycle de développement des plantes. Dans ce cadre, la sélection variétale apparaît comme un outil important. Elle a pour but la création de variétés qui soient productives et peu sensibles aux contraintes de l'environnement. Pour cela, de nouvelles variétés sont créées en croisant des variétés existantes choisies pour leurs qualités respectives selon différents critères ou traits morphologiques, phénologiques et physiologiques. Ensuite les meilleures plantes issues de ces croisements sont sélectionnées sur la base de leur phénotype (caractéristique observable de la plante) jusqu'à obtenir une nouvelle plante avec des qualités voulues [118]. Mais le phénotype d'une plante est déterminé à la fois par sa constitution génétique (génotype) et les conditions dans lesquelles elle s'est développée (l'environnement), ces deux effets ne sont pas toujours additifs, il y a souvent une interaction génotype $\times$ environnement. Donc pour faire de la sélection, il est important d'analyser le comportement des variétés en fonction des caractéristiques de l'environnement.

2.1 Analyse statistique de l'interaction génotype $\times$ environnement

Pour déterminer les variétés les meilleures en fonction de l'environnement, la méthode classiquement utilisée est l'expérimentation avec des essais multilocaux et pluriannuels. Les résultats de l'expérimentation sont ensuite utilisés dans des modèles statistiques de l'interaction génotype $\times$ environnement ($G \times E$) qui permet de rendre compte de l'interaction, ou de la prédire. Nous présentons dans les sections suivantes différentes méthodes statistiques qui permettent d'identifier, de décrire et d'expliquer l'interaction $G \times E$.

2.1.1 Modèle d'analyse de la variance

Environnement à effet fixé

Les effets du génotype, de l'environnement et de l'interaction sont explicités en ajustant un modèle statistique aux résultats des expérimentations. Le modèle statistique le plus utilisé est l'analyse de la variance. Dans une expérimentation multilocale et/ou pluriannuelle, si l'on considère non pas les données élémentaires, mais les moyennes par variété et par environnement le modèle d'analyse de la variance va s'écrire comme suit:

$$Y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ij} \tag{2.1}$$

où μ représente un terme constant, Y_{ij} représente le phénotype moyen du génotype i dans l'environnement j, α_i représente l'effet du génotype i, β_j l'effet de l'environnement j, $(\alpha\beta)_{ij}$ représente l'effet de l'interaction $G \times E$ et le terme aléatoire ε_{ij} représente l'erreur de prédiction, supposée être normalement distribuée, d'espérance nulle et de

variance constante σ^2, $\varepsilon_{ij} \sim N(0, \sigma^2)$.
Avec les contraintes usuelles sur les paramètres, à savoir $\sum_i \alpha_i = \sum_j \beta_j = 0, \forall j$, $\sum_i (\alpha\beta)_{ij} = 0$ et $\forall i, \sum_j (\alpha\beta)_{ij} = 0$, les estimations des différents effets sont données par:

$$\hat{\alpha}_i = Y_{i.} - Y_{..} \tag{2.2}$$

$$\hat{\beta}_j = Y_{.j} - Y_{..} \tag{2.3}$$

$$\hat{(\alpha\beta)}_{ij} = Y_{ij} - Y_{i.} - Y_{.j} + Y_{..} \tag{2.4}$$

où $Y_{i.}$ désigne le phénotype moyen du génotype i sur tous les environnements, $Y_{.j}$ le phénotype moyen de tous les génotypes dans l'environnement j et $Y_{..}$ le phénotype moyen de tous les génotypes sur tous les environnements.
D'après (2.4) l'interaction $(\alpha\beta)_{ij}$ peut être estimée par la différence entre l'effet particulier du génotype i dans l'environnement j et l'effet moyen de ce même génotype i. La figure 2.1A illustre le cas où il n'y a pas d'interaction, tous les génotypes répondent de la même manière à un changement d'environnement. Les autres graphiques illustrent deux types d'interaction $G \times E$ où les différences phénotypiques entre les génotypes sont donc variables d'un environnement à l'autre. L'interaction est dite quantitative (figure 2.1B et figure 2.1C) si le classement des génotypes est conservé d'un environnement à un autre; dans le cas contraire, on dira que l'interaction est qualitative (figure 2.1D). Une interaction qualitative implique que le choix du meilleur génotype est déterminé par l'environnement.

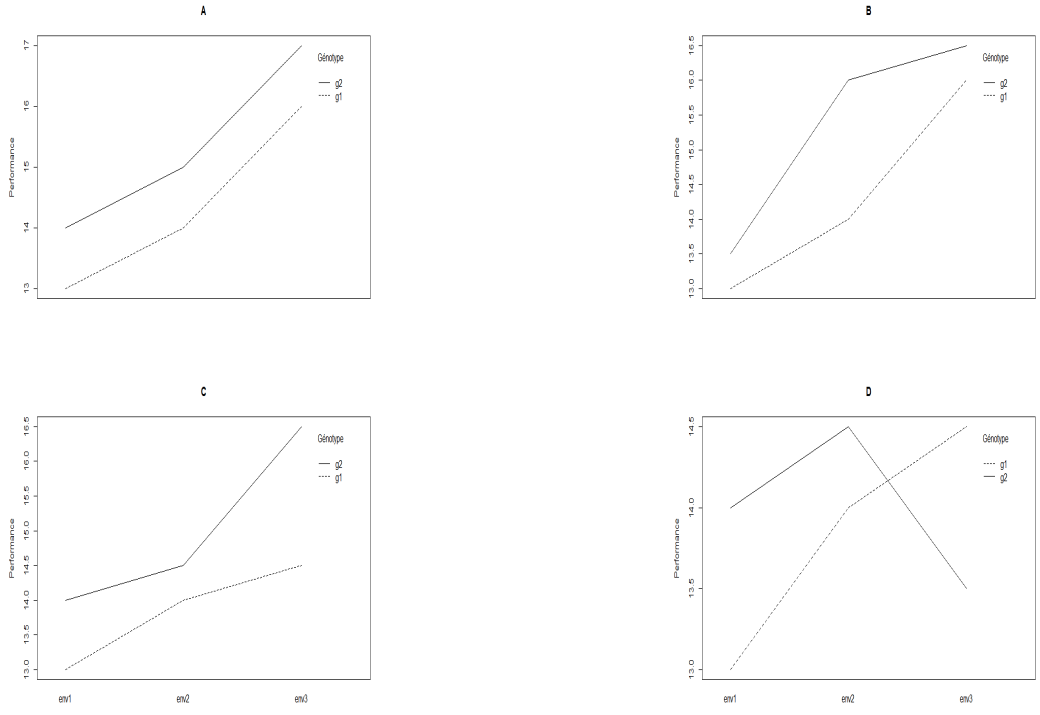


FIGURE 2.1 – *(A) pas d'interaction, (B) interaction quantitative, (C) interaction quantitative, (D) interaction qualitative*

Environnement à effet aléatoire

Même si l'interaction $G \times E$ existe, on ne peut pas imaginer comparer des génotypes dans chaque environnement : on va plutôt définir une population cible d'environnements, dans laquelle les sites d'expérimentation sont choisis au hasard (en principe). Alors l'effet environnement et l'interaction $G \times E$ deviennent aléatoires, on les suppose gaussiens, centrés, de variances σ^2_β et $\sigma^2_{\alpha\beta}$. Pour montrer que le génotype $i^{'}$ est meilleur que le génotype i, on calcule leur différence de production $E(Y_{i^{'}j} - Y_{ij})$, $j = 1, \cdots, J$ les environnements. Cette différence est estimée par: $Y_{i^{'}.} - Y_{i.}$ et la variance de cette estimation est égale à $Var(Y_{i^{'}.} - Y_{i.}) = \frac{2}{J}\left(Var(\alpha\beta) + Var(\varepsilon)\right)$. Pour que l'estimateur soit précis, il faut soit augmenter le nombre d'environnements, soit réduire la variance de l'interaction $G \times E$. L'expérimentation sur le terrain étant souvent une tâche lourde qui limite le nombre d'expérimentations possibles, la solution la plus réaliste serait alors de réduire la variance de l'interaction.
L'interaction gêne donc l'évaluation des génotypes car elle détermine un aléa sur les différences entre génotypes. Une manière de réduire cette gêne serait d'expliquer l'interaction en fonction des génotypes et des environnements.
Le modèle d'analyse de variance permet de déterminer l'existence ou non d'une interaction $G \times E$, mais il ne permet pas une compréhension ni une prédiction de cette interaction.

Sans forcément permettre de comprendre jusqu'aux mécanismes de l'interaction $G \times E$, différents modèles permettent de la décrire ou de la prédire en fonction de caractéristiques des génotypes et des environnements.
Dans la littérature deux types de modèle sont considérés: les modèles descriptifs et les modèles explicatifs.
Les modèles descriptifs permettent de résumer le comportement individuel des génotypes sur un ensemble d'environnements. Cela permet d'obtenir des regroupements de génotypes ayant des comportements similaires et des groupes d'environnements qui sont semblables. Ci-dessous, nous présentons deux modèles descriptifs : la régression conjointe [41] et le modèle $AMMI$ [28, 58, 50].
Les approches explicatives visent à prédire l'interaction $G \times E$ en fonction de variables décrivant d'une part le génotype par des traits morphologiques ou des marqueurs génétiques, et d'autre part l'environnement par des mesures concernant le sol (chimie) ou l'atmosphère (météorologique). Cela permet, par exemple, de prédire le comportement d'un génotype dans des environnements nouveaux. Différents modèles existent et dans la suite nous nous limiterons à la présentation d'un modèle particulier, la régression factorielle [3, 31, 137].

2.1.2 Méthodes descriptives

Régression conjointe

La régression conjointe [24, 41, 131] consiste à expliquer le terme d'interaction $G \times E$ de (2.1) par une simple régression sur l'effet environnement, avec une pente spécifique à chaque génotype:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \rho_i \beta_j + \varepsilon_{ij} \tag{2.5}$$
$$\sum_i \alpha_i = \sum_j \beta_j = 0$$

où μ représente un terme constant, α_i l'effet principal du génotype i, β_j l'effet principal de l'environnement j, ρ_i le coefficient de régression génotypique.
Les paramètres sont estimés en deux étapes: les paramètres μ, α_i et β_j sont d'abord estimés sur un premier modèle additif, le coefficient de régression ρ_i étant ensuite estimé en utilisant les résidus de ce premier modèle.
Les coefficients de régression ρ_i renseignent sur l'adaptabilité des génotypes. Le coefficient de régression sera positif pour les génotypes valorisant bien les environnements favorables, négatif pour les génotypes bien adaptés aux environnements défavorables et nul pour les génotypes stables moyennement adaptés à tous les environnements.
Certains auteurs [108, 38] préconisent aussi d'associer au coefficient de régression d'autres critères qui renseignent sur la qualité d'ajustement à la droite de régression conjointe pour juger de l'adaptabilité d'un génotype.
Contrairement au modèle d'analyse de la variance, la régression conjointe permet de décrire l'interaction $G \times E$ en identifiant les génotypes stables et les génotypes sensibles. Cependant, la notion de stabilité d'un génotype n'est pas universelle: elle dépend des environnements mais aussi des autres génotypes auquel il est comparé [46].
La méthode permet aussi de définir des groupes d'environnements où les génotypes ont des performances similaires et des groupes de génotypes qui ont des comportements similaires dans les différents environnements. Toutefois, la principale limite dans cet exercice vient du fait que la qualité de l'environnement est résumée par une seule variable, communément appelée fertilité. Dans un monde où les aptitudes des environnements ne se résument pas à une seule variable, cela peut laisser une part non expliquée de l'interaction $G \times E$. Dans la section suivante, cette limitation disparait avec la prise en compte de caractéristiques environnementales multidimensionnelles.

Modèle AMMI

Le modèle AMMI (Additive Main effect and Multiplicative Interaction) [28, 58, 50, 138] permet également de décrire l'interaction $G \times E$. La qualité d'un environnement ne s'y résume pas à une moyenne, mais à plusieurs variables latentes estimées par analyse factorielle. Ce modèle s'ajuste en deux étapes. D'abord, les effets principaux sont estimés par moindres carrés à l'aide du modèle additif d'analyse de la variance. Ensuite une décomposition en valeurs singulières (une ACP) est appliquée aux résidus du modèle additif, afin de décomposer l'interaction $G \times E$ en une somme de

termes multiplicatifs. Chaque terme multiplicatif est formé par le produit d'un paramètre caractérisant le génotype (score génotypique) et d'un paramètre caractérisant l'environnement (score environnemental). Le modèle peut être écrit comme suit:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \sum_{k=1}^{K} \lambda_k \gamma_{ik} \delta_{jk} + \varepsilon_{ij} \tag{2.6}$$

où μ représente un terme constant, α_i l'effet principal du génotype i, β_j l'effet principal de l'environnement j, λ_k la valeur singulière caractérisant la part de l'interaction expliquée par le $k^{ième}$ terme, γ_{ik} le score génotypique caractérisant la sensibilité génotypique, δ_{jk} le score environnemental.
Le modèle AMMI comprend les K termes d'interactions issus des K premières composantes principales de l'interaction. Comme dans toute ACP, le premier terme multiplicatif est celui qui explique la plus grande part des variations, suivi des autres termes.
Les scores génotypiques et environnementaux sont utilisés pour faire une représentation graphique appelée biplot [48] qui permet de comprendre l'interaction. Les génotypes et les environnements sont projetés sur les deux axes principaux de l'ACP. Sur le biplot, les génotypes sont représentés par des points et les environnements par des axes qui passent par l'ordonnée à l'origine qui représente le comportement moyen.
Comme l'estimation du score d'un environnement nécessite la mesure des productions de tous génotypes dans cet environnement, aucune prédiction n'est possible pour un nouvel environnement. Le raisonnement symétrique s'applique aux génotypes.

Les méthodes descriptives pour l'analyse de l'interaction $G \times E$ permettent d'identifier les différences d'adaptation entre génotypes et de les expliciter en décomposant l'interaction $G \times E$. Elles ne permettent pas de prédire le comportement des génotypes dans de nouveaux environnements, les paramètres du modèle étant liés aux données utilisées.

2.1.3 Une méthode explicative: la régression factorielle

Contrairement aux méthodes précédentes, qui ne tiennent compte d'aucune variable génotypique ni environnementale, la régression factorielle [3, 31, 137], consiste à prendre les produits des variables génotypiques par les variables environnementales comme régresseur de l'interaction $G \times E$. Cela permet de fournir des explications biologiques des résultats observés. Les covariables génotypiques telles que la précocité, la profondeur maximale des racines etc, renseignent sur les origines génotypiques de l'interaction $G \times E$. Tandis que les covariables environnementales telles que la température, les précipitations, le sol etc, renseignent sur les contraintes environnementales ayant un impact sur le comportement des génotypes.
En considérant une seule covariable environnementale z, le modèle de régression factorielle peut être écrit comme suit:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \rho_i z_j + \varepsilon_{ij} \tag{2.7}$$

Les modèles de régression factorielle peuvent être étendus pour inclure plusieurs covariables environnementale et plusieurs covariables génotypiques [137], par exemple

comme suit:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \sum_{kh} \gamma_{kh} x_{ik} z_{jh} + \sum_{h} \rho_{ih} z_{jh} + \sum_{k} \zeta_{jk} x_{ik} + \varepsilon_{ij} \tag{2.8}$$

où μ représente un terme constant, α_i l'effet principal du génotype i, β_j l'effet principal de l'environnement j, ρ_{ih} le coefficient de régression du génotype i pour la h^e covariable de l'environnement, ζ_{jk} le coefficient de régression de l'environnement j pour la k^e covariables du génotype, $\gamma_{kh} x_{ik} z_{jh}$ est la contribution du produit des covariables x_k et z_h à l'explication de l'interaction, $\rho_{ih} z_{jh}$ la contribution du produit des indicatrices des génotypes et de la covariable environnementale z_h et $\zeta_{jk} x_{ik}$ la contribution du produit de la covariable x_k et des indicatrices des environnements.

Des études [62, 13, 6] ont montré que la régression factorielle permet d'expliquer une grande part de l'interaction $G \times E$ avec une prise en compte des données climatiques caractéristiques des environnements. Toutefois, le nombre de variables climatiques pris en compte par la régression factorielle est limité au risque d'avoir un nombre de variables explicatives supérieur au nombre d'observations. D'où la nécessité de choisir, parmi toutes les variables environnementales mesurées, celles qui influent réellement sur la performance des génotypes donc expliquent l'interaction $G \times E$. Cette sélection de covariables environnementales constitue la principale difficulté dans l'utilisation des modèles de régression factorielle [142].

L'évaluation de nouveaux génotypes par simulation à travers un modèle de culture semble être une alternative intéressante. Cette approche donne la possibilité d'évaluer une gamme plus vaste de génotypes et d'environnements dans un intervalle de temps restreint. Elle permet ainsi de prédire l'interaction $G \times E$ [19, 21, 5, 63, 115, 32].
Dans la section suivante, les principales caractérisations de ce que l'on appelle un modèle de culture sont présentées.

2.2 Modélisation d'une culture

En agronomie, de nombreuses études sont consacrées à la modélisation des cultures et plus spécifiquement à la modélisation de la croissance et du développement de la plante en fonction des conditions environnementales. Ces modèles, appelés modèles de culture, permettent de reproduire sous forme virtuelle le fonctionnement de la plante en prenant en compte différents facteurs et processus interagissants.

2.2.1 Croissance et développement de la plante

La croissance est le changement quantitatif irréversible de la taille, du poids et de la surface de la plante ou de ses parties.
Le développement de la plante, quant à lui, correspond à divers changements qualitatifs qui surviennent dans la vie d'une plante. Ces principaux changements sont : la

germination des semences, le développement végétatif (formation des entre nœuds, des feuilles des nouvelles racines etc), la floraison (formation des fleurs) et la fructification (formation des fruits).
Ces changements dépendent des caractéristiques génétiques de la plante mais aussi des facteurs externes. Cependant, indépendamment du milieu dans lequel la plante évolue, la croissance et les échanges suivent les mêmes règles. L'appareil racinaire (les racines) permet l'absorption de l'eau et des sels minéraux pour alimenter les différents organes de la plante. Tandis qu'une partie de l'appareil caulinaire (tiges, feuilles, etc) assure la photosynthèse qui permet la production du sucre (glucose) et de l'oxygène. Pour pouvoir effectuer la photosynthèse, la plante a besoin : de l'eau puisée par ses racines, du dioxyde de carbone (CO2) et de la lumière (solaire) pour obtenir de l'énergie. La plupart des modèles de culture se basent sur ces mécanismes pour reproduire la croissance d'une plante, en ignorant les éléments minéraux considérés comme disponibles en quantité suffisante.

2.2.2 Modèle de culture

Les exemples d'utilisation des modèles de simulation des systèmes biologiques complexes se font de plus en plus nombreux, que cela soit pour la prévision annuelle des rendements agricoles [22, 95, 94, 49, 72, 16, 17] ou dans le cadre de programmes d'amélioration variétale [12].

Les modèles de culture décrivent la croissance et le développement d'un système sol-plante en interaction avec le climat et les itinéraires techniques [9]. Les sorties de ces modèles permettent d'estimer certaines grandeurs agronomiques comme par exemple le rendement. Les modèles de culture sont représentés le plus souvent par un ensemble d'expressions mathématiques constituées de variables reliées dans des fonctions. Ces expressions décrivent l'état du système par des variables d'état et représentent le fonctionnement des composants du système. Selon la nature des relations mathématiques, on distingue deux catégories de modèles [61]:

- Les modèles mécanistes qui décrivent et quantifient la performance de la plante en se basant sur la connaissance plus ou moins précise des processus physique, chimique et biologique du système intervenant dans la croissance et le développement de la plante.

- Les modèles empiriques, parfois appelés modèles statistiques, dont la structure est obtenue par des ajustements à des données expérimentales. Ces relations n'expliquent pas nécessairement le fonctionnement du système sol plante.

Les modèles de culture sont en général un mélange de ces deux types de modèle.
Après une bonne connaissance du système, les différentes étapes de construction d'un modèle de culture sont la conceptualisation, le paramétrage et la validation.

La conceptualisation du modèle

Elle consiste à la mise en place de relations mathématiques qui traduisent le fonctionnement du système. Ces relations mathématiques sont constituées d'équations aux différences qui sont de la forme [19]:

$$U_1(t+\Delta t) = U_1(t) + g_1\left(U(t), X(t); \theta\right)$$

$$\vdots$$

$$U_s(t+\Delta t) = U_s(t) + g_s\left(U(t), X(t); \theta\right)$$

où t est le temps, Δt représente un pas de temps qui est le plus souvent journalier, $U(t) = [U_1(t), \cdots, [U_s(t)]^T$ représente le vecteur des variables d'état au temps t, $X(t)$ le vecteur des variables d'entrée au temps t, θ le vecteur des paramètres des équations mathématiques et g la fonction qui permet de relier les variables d'état aux variables d'entrée.
Les variables d'entrée sont des variables quantitatives ou qualitatives, elles peuvent changer selon la situation étudiée. Elles sont supposées connues et décrivent les caractéristiques du sol (réserve d'eau utile, teneur en matière organique etc), les variables climatiques (température minimale et maximale journalière, rayonnement, cumul des précipitations journalières, demande évaporative etc) et l'itinéraire technique (dates et densités de semis, irrigation, rotation, niveau de fertilisation, la variété qui est représentée par une combinaison de paramètres etc).
Les variables d'état décrivent l'état du système étudié à différents pas de temps (par exemple la quantité d'azote disponible dans le sol, la biomasse aérienne de la culture, le rendement en grain, la profondeur des racines).
La conceptualisation d'un modèle simple qui permet de simuler la croissance du maïs est présenté dans Soltani *et al.* [129]. Le modèle est défini à l'aide de trois variables d'état: la somme de degrés-jour (TT), la surface foliaire par unité de sol (LAI) et la biomasse totale (B). Le modèle utilise comme variables d'entrée la température maximale $(TMAX)$, la température minimale $(TMIN)$ et le rayonnement (I) et comme paramètres à estimer la température de base (T_{base}), au dessous de laquelle la culture ne se développe pas, le coefficient d'extinction de la lumière (K), utilisé pour calculer le rayonnement intercepté, l'efficacité de l'utilisation du rayonnement (RUE), le taux relatif d'augmentation de l'indice de surface foliaire $(alpha)$, l'indice maximal de surface foliaire (LAI_{max}), le seuil de temps thermique à partir duquel la plante est mûre (TTM) et le seuil de temps thermique à partir duquel s'achève la phase végétative de croissance (TTL).
Les équations qui décrivent le modèle sont les suivantes:

$$TT_{t+1} = TT_t + \Delta TT_t \tag{2.9}$$

$$B_{t+1} = B_t + \Delta B_t \tag{2.10}$$

$$LAI_{t+1} = LAI_t + \Delta LAI_t \tag{2.11}$$

où

$$\Delta TT_t = \max\left[\frac{TMIN_t + TMAX_t}{2} - T_{base}, 0\right] \tag{2.12}$$

$$\Delta B_t = RUE \times (1 - \exp(-K \times LAI_t))^{I_t} \tag{2.13}$$

$$\Delta LAI_t = alpha \times \Delta TT_t \times LAI_t \times \max\left[LAI_{max} - LAI_t, 0\right] \tag{2.14}$$

La première étape de la construction du modèle consiste à compter l'avancement dans le temps physiologique comme la moyenne des températures minimale et maximale de la journée pour autant qu'elle excède un seuil appelé température de base T_{base}. Ce temps physiologique est appelé somme de degrés jour (TT) et s'exprime en degrés jour Cj.
Dans un second temps, la variation de surface foliaire est calculée en fonction de la somme de degrés et du LAI existant. Cette variation dépend aussi du seuil de temps thermique de la culture, dans la mesure où les feuilles ne se développent que pendant la phase végétative de croissance, égale dans ce modèle au seuil de temps thermique TTL.
L'hypothèse finale émise est que le taux d'accumulation de biomasse dépend du LAI et de I. L'accroissement de la biomasse se poursuit jusqu'à l'obtention d'un temps thermique maximal TTM.
Les modèles de culture combinent généralement de nombreuses équations aux différences faisant intervenir des seuils, construites suivant le même formalisme (variables explicatives, variables d'état et paramètres). Les modèles de culture sont des modèles à compartiments reliés entre eux. Chaque compartiment est un modèle en lui même et la plupart du temps, on ne s'intéresse qu'à la valeur finale d'un d'entre eux qui peut être la production de biomasse ou de grain. Dans ce cas, nous pouvons intégrer pas à pas les équations aux différences afin d'éliminer les valeurs intermédiaires des variables d'état. Une variable d'état à un moment donné n'est plus alors fonction que de variables explicatives.
Désignons par y la sortie à laquelle on s'intéresse; le modèle peut alors s'écrire

$$y = f(X; \theta) \tag{2.15}$$

où X représente le vecteur des variables explicatives aux différents pas de temps et θ le vecteur des paramètres. A la différence des variables d'entrée qui peuvent varier au cours du temps, une partie des paramètres θ reste constante pour toutes les situations étudiées tandis que l'autre partie, que nous noterons v, change en fonction des caractéristiques des génotypes.

Estimation des paramètres

Après la conceptualisation et l'implémentation du modèle, l'étape suivante pour la construction d'un modèle de culture consiste à estimer les paramètres θ du modèle à partir d'un échantillon de données. Plusieurs méthodes statistiques comme les moindres carrés et le maximum de vraisemblance existent pour l'estimation des paramètres d'un modèle. Cependant ces méthodes sont des approches fréquentistes qui se basent sur un échantillon de données pour trouver la vraie valeur des paramètres en considérant que ceux-ci ne sont pas aléatoires. Cette approche est le plus souvent inappropriée pour l'estimation des paramètres des modèles de culture. En effet le nombre

important de paramètres dans les modèles de culture et la structure complexe des données utilisées pour l'estimation (grande variabilité spatiale et temporelle des données) conduisent souvent à un surajustement ou à une convergence vers un minimum local du critère à optimiser. Toutefois, il existe plusieurs solutions utilisant l'approche fréquentiste. La première consiste à sélectionner un sous-ensemble de paramètres à estimer à partir des données et à fixer les autres à partir d'information sur le système étudié [19]. La deuxième solution serait d'utiliser la régression séquentielle pour choisir successivement les paramètres qui conduisent au meilleur ajustement [143]. Une troisième solution consiste à faire une analyse de sensibilité pour déterminer les paramètres qui influent le plus sur le modèle [121].
Une alternative à l'approche fréquentiste serait l'approche bayésienne. Les paramètres sont alors aléatoires et leur estimation s'obtient au travers un échantillon de données et d'informations a priori sur la distribution des valeurs possibles des paramètres. L'estimation se fait en deux étapes. Dans un premier temps, on définit une distribution de probabilité a priori des paramètres en se basant sur la littérature ou des connaissances d'experts. Dans un second temps, à partir de la distribution a priori, on détermine la distribution a postériori à partir de la distribution a priori et des données en utilisant le théorème de Bayes. Les valeurs des paramètres seront alors déterminés par l'espérance ou le mode de la distribution a postériori.

Validation

La validation du modèle est une étape importante dans l'utilisation des modèles de culture. Elle permettra d'estimer le niveau de fiabilité à accorder au modèle mais également d'identifier les éventuels problèmes afin de pouvoir améliorer le modèle. La validation consiste à comparer les données simulées par le modèle à celles observées selon différentes conditions de sol, de climat et de culture ou de variété. Les critères mathématiques les plus utilisés en modélisation des cultures sont les suivants.
Le critère le plus simple pour mesurer les performances d'un modèle consiste à calculer la différence entre la valeur mesurée Y_i d'une variable et celle estimée par le modèle $\hat{Y}_i$ pour chaque situation i:

$$D_i = Y_i - \hat{Y}_i \tag{2.16}$$

Une manière de résumer les différences D_i consiste à considérer leur espérance, appelée biais d'estimation et estimé par:

$$Biais = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i) \tag{2.17}$$

où N est le nombre de situations étudiées. Un biais négatif implique que le modèle sous-estime la moyenne alors qu'un biais positif implique une sur-estimation de la moyenne. Cependant un faible biais ne présuppose pas forcément une bonne qualité de prédiction. Cela peut être la conséquence d'une compensation entre sous-estimation et sur-estimation de la moyenne. Un critère pour contourner ce problème est l'erreur quadratique moyenne (MSE) et ses variantes $RMSE$ et $RRMSE$.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2 \tag{2.18}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2} \tag{2.19}$$

$$RRMSE = \frac{RMSE}{\bar{Y}} \tag{2.20}$$

L'efficience de modèle et le coefficient de corrélation sont très souvent utilisés. Ils présentent une borne inférieure et/ou supérieure, ce qui facilite leur interprétation et les rend pertinents pour comparer des modèles. L'efficience du modèle est donnée par :

$$EF = 1 - \frac{\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{N}(Y_i - \bar{Y}_i)^2} = 1 - R^2 \tag{2.21}$$

où R^2 est le coefficient de détermination.
Le coefficient de corrélation de Pearson est donné par:

$$r = \frac{\hat{\sigma}_{Y\hat{Y}}}{\sqrt{\hat{\sigma}^2_Y \hat{\sigma}^2_{\hat{Y}}}} \tag{2.22}$$

où $\hat{\sigma}^2_Y$, $\hat{\sigma}^2_{\hat{Y}}$ et $\hat{\sigma}_{Y\hat{Y}}$ représentent respectivement les estimations des variances de Y, de $\hat{Y}$ et de la covariance de Y et $\hat{Y}$.

Une fois calibrés et évalués à partir de résultats sur quelques années d'expérimentations en quelques lieux et en tenant en compte du sol, de la météo et des pratiques culturales, les modèles de culture permettent de simuler la production d'une culture dans le temps et dans l'espace. Ils permettent aussi de prédire le comportement de variétés virtuelles.
On peut ainsi élaborer, in silico, des variétés adaptées à des contraintes pédo-climatiques diverses en vue de guider les sélectionneurs vers des plantes-cibles à atteindre ou idéotypes [34].
Les modèles utilisés sont de type "boîte-noire" c'est-à-dire que l'utilisateur n'a pas un accès direct au code numérique. L'utilisateur fournit simplement des entrées sous forme de séries climatiques, d'informations sur le sol, de pratiques culturales et de paramètres caractérisant la variété devant être simulée; il obtient le résultat de la simulation qui peut être, par exemple, un rendement. On peut donc représenter le modèle de culture selon le schéma donné à la figure 2.2.

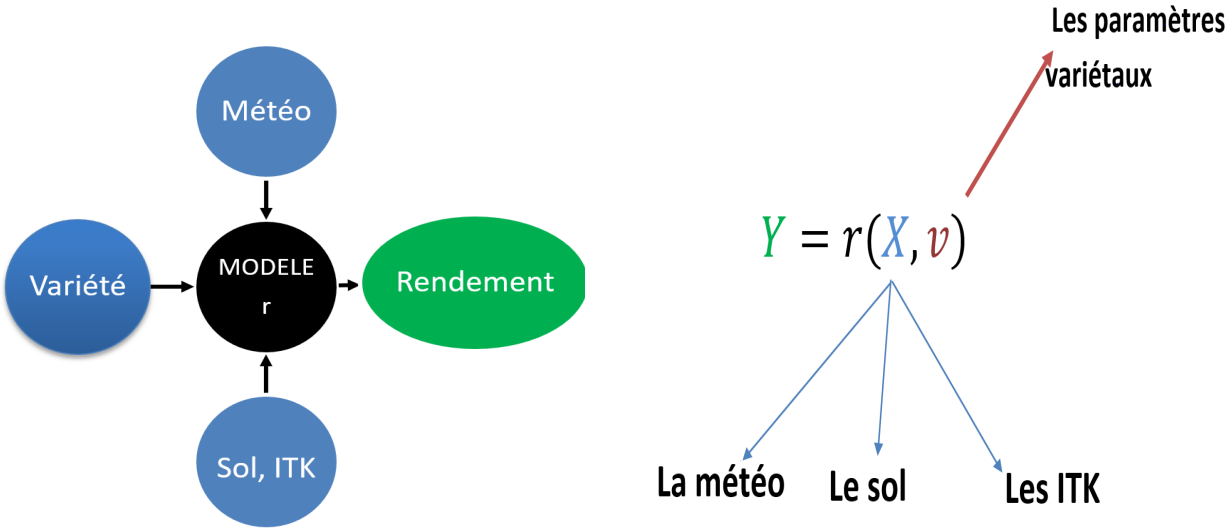


FIGURE 2.2 – *Illustration d'un modèle de culture*

Pour la suite, on considérera les entrées déterministes, à savoir les pratiques culturales (ITK) et le sol, comme étant fixées à des valeurs particulières. Il reste les entrées aléatoires de la météo et les paramètres qui définissent la variété. On considère que la distribution conjointe des variables météo (pluie, vitesse du vent, rayonnement,...) est décrite par un modèle stochastique du climat C_u avec des paramètres qui dépendent in fine de la latitude et la longitude du lieu u. Dans la suite, pour mettre en évidence le fait que la météo dépend du lieu u, on notera notre modèle $f(C_u, v)$ ou pour simplifier $f(u, v)$.

2.3 Prédiction avec une météo encore inconnue

La prédiction de l'interaction $G \times E$ peut être rendue difficile parce que,

- d'une part sur chaque station, on ne dispose pas toujours de suffisamment de données pour bien estimer les incertitudes de la production dues au climat;
- d'autre part parce qu'il y a des endroits sans station météo.

Pour contourner le problème de données météo insuffisantes pour servir de variables d'entrée au modèle de culture, l'utilisation d'un générateur stochastique de climat permettant d'effectuer des simulations du climat de quelques années à plusieurs millénaires est une solution de plus en plus adoptée [15]. Les séries de données générées auront les mêmes caractéristiques statistiques que les données observées. Pour les endroits ne disposant pas de station météo, les paramètres statistiques d'un générateur météorologique doivent être calculés à partir des données des stations météorologiques les plus proches. Des méthodes géostatistiques sont ensuite utilisées pour obtenir une interpolation spatiale des paramètres du modèle à partir des données ponctuelles de la zone. Néanmoins, un certain nombre d'erreurs sont liées intrinsèquement

à ces méthodes: l'incohérence des données météo entre les lieux, le nombre important de données manquantes peuvent avoir des conséquences sur l'interpolation. Il est montré dans [59] que le fait de caler le modèle sur des combinaisons d'années assez disparates d'une station à l'autre entraine une incohérence totale entre les probabilités de transition obtenues par interpolation et les fréquences calculées. Une solution possible serait d'éliminer les années présentant des données manquantes. Pour notre étude, on va utiliser la méthode du krigeage comme interpolateur spatial, en supposant notre plan d'expérience dense et régulier. Il est en effet montré dans Taupin *et al.* [132] que la densité ainsi que le positionnement des stations météo ont une influence importante sur les résultats de l'interpolation.

2.3.1 Générateur stochastique de climat

Un générateur stochastique de climat [113, 112] est un modèle mathématique basé sur une approche statistique et probabiliste de la météorologie. Il permet de produire une série chronologique de variables climatiques comme les températures minimale et maximale, les précipitations, le rayonnement solaire etc. La construction d'un modèle de simulation pour ces variables nécessite l'élaboration d'un concept de relations stochastiques qui existent entre les variables météorologiques. En effet, les données observées étant des séries temporelles, il existe une auto-corrélation temporelle pour chacune des variables étudiées. Il existe également une corrélation entre les variables: le rayonnement et la température par exemple sont susceptibles d'être plus faibles pendant les jours humides que pendant les jours secs.
La première étape de construction consiste à simuler l'occurrence des précipitations. Il existe deux grandes approches pour simuler ces occurrences. La première, dite markovienne et utilisée par des générateurs de climats tels que WGEN [113, 114], USCLIMATE [73] et CLIGEN [145], applique une chaine de Markov pour décrire l'occurrence des jours humides ou secs. La seconde, quant à elle, est basée sur la distribution de la longueur des séries de jours humides et secs. Le principal générateur de climat utilisant cette approche est Lars-WG [112].
La deuxième étape consiste à construire des distributions de probabilité conditionnelles pour les variables climatiques telles que la température moyenne, la quantité de précipitations, le rayonnement solaire etc, en considérant les précipitations comme la variable principale, et en conditionnant les autres variables d'un jour donné selon que la journée soit humide ou sèche.
Soit on veut simuler des données météorologiques simultanées dans plusieurs sites alors il faut prendre en compte à la fois les corrélations spatiales et temporelles. Soit on se contente de simuler des données météorologiques séparément pour chaque site, sans considérer les corrélations spatiales des événements météorologiques.
En ce qui nous concerne, nous cherchons à optimiser une variété pour chaque site, le fait que les données météo soient interdépendantes n'a pas d'importance.

Les modèles mathématiques sont calibrés en estimant les paramètres à partir de séries mesurées, afin de reproduire des propriétés statistiques semblables aux séries mesurées. En particulier les valeurs moyennes, les niveaux de variabilité, les corrélations temporelles, les corrélations entre les variables et la persistance d'un phénomène

doivent être bien reproduits. Une recommandation classique est d'utiliser au moins vingt ans de données quotidiennes pour estimer de façon précise les paramètres considérés [126].

III

Méthodes d'optimisation

Lorsque l'on souhaite trouver une variété adaptée à un lieu donné on se retrouve devant un problème d'optimisation des paramètres caractérisant la variété sous des contraintes pédo-climatiques fixées. Si l'on souhaite obtenir une cartographie des variétés optimales sur une région aux contraintes climatiques contrastées, on se retrouve à devoir réaliser de très nombreuses optimisations compte tenu de la grande diversité des situations pédo-climatiques. La production est aléatoire, car elle dépend fortement entre autres des pluies qui vont tomber pendant la saison de culture. C'est donc la maximisation de l'espérance ou d'un quantile de la production que l'on devra rechercher.

Supposons que le modèle à optimiser soit représenté par une fonction f définie par:

$$\begin{array}{rccc} f & : & \mathbb{E} & \rightarrow & \mathbb{R} \\ & & (u,v) & \rightarrow & f(u,v) \end{array}$$

L'ensemble $\mathbb{E}$ est un compact de $R^d, d \geq 1$, pris comme le produit cartésien de deux ensembles U et V, donc $\mathbb{E} = U \times V$. Le premier ensemble, noté U, contient les entrées u que l'on appellera variables conditionnantes et le deuxième ensemble, noté V, contient les paramètres à optimiser v ou variables conditionnées.
Notre objectif est de trouver pour chaque u fixé, le jeu de paramètres v qui maximise l'espérance ou un quantile de la fonction f, autrement dit une carte ou lieu des optimums.
Pour chaque u, nous devons résoudre le problème $v^* = \text{argmax}_{v \in V} g(v)$ avec $g(v) = E(f(u,v))$ par exemple si l'on cherche à maximiser l'espérance de f.
Dans ce chapitre, nous nous intéressons aux méthodes pour la recherche de suite de points qui converge vers la solution du problème $v^* = \text{argmax}_{v \in V} g(v)$ pour chaque u fixé.
L'expression analytique des modèles de cultures est complexe, leurs sorties souvent discontinues par rapport aux paramètres, l'optimisation se fera donc numériquement. Parmi les méthodes d'optimisation numériques, on distingue deux grandes catégories: les méthodes d'optimisation dites locales et celles dites globales.

3.1 Principales méthodes d'optimisation locales

L'optimisation locale consiste à chercher l'optimum global dans un ensemble d'optima locaux. A partir d'un point initial, l'algorithme cherchera à améliorer la situation initiale en se dirigeant vers l'optimum le plus proche. Pour cela, l'algorithme va exploiter au mieux les informations relatives à l'évolution de la fonction au voisinage du point en cours.
Dans cette partie, nous allons présenter brièvement les méthodes d'optimisation locales classiquement employées. On s'intéressera à un problème de minimisation, sachant que toute recherche de maxima peut se ramener à une recherche de minima par passage à l'opposé.

3.1.1 Algorithme de descente

Les problèmes d'optimisation $v^* = \text{argmin}_{v \in V} g(v)$ sont généralement résolus par les algorithmes de descente de gradient. L'algorithme de descente s'applique lorsque l'on cherche le minimum d'une fonction dont on connait la valeur et celle de sa dérivée pour un jeu de paramètres donnés. Initialement introduits par Cauchy en 1847 [23], les algorithmes de descente sont des algorithmes itératifs basés sur l'amélioration d'une solution approchée pour converger vers la solution du problème v^*. Cela consiste à chercher à l'itération $k+1$, le jeu de paramètre v_{k+1} tel que $g(v_{k+1}) \leq g(v_k)$.
Il s'agit de trouver à chaque itération k, un nouveau jeu de paramètres v_{k+1} dans une direction d_k obtenue à partir de la valeur de la fonction et de son gradient. Une description générale d'un algorithme de descente est la suivante :

1. Choisir arbitrairement un point initial x_0.
2. Tant que le test de convergence est non satisfait,
    3. Calculer la direction de descente d_k telle que: $\nabla g(v_k)^T d_k < 0$
    4. Choisir un pas $s_k > 0$ tel que: $g(v_k + s_k d_k) < g(v_k)$
    5. Calculer le nouveau point: $v_{k+1} = v_k + s_k d_k; k = k+1$

On distingue différentes méthodes d'optimisation basées sur l'algorithme de descente, caractérisées par le choix des directions de descente et du pas effectué dans chaque direction.
Parmi ces algorithmes, on peut citer l'algorithme de gradient à pas fixe, l'algorithme de gradient à pas optimal, l'algorithme de gradients conjugués qui sont rapidement présentés ci-après.

L'algorithme de gradient à pas fixé

La direction de descente est choisie opposée à la direction du gradient, soit $d_k = -\nabla g(v_k)$. Le pas de la descente est fixé au départ. Il doit être choisi de manière judicieuse: trop grand, il rendra une convergence vers le minimum impossible; trop petit il impliquera une convergente très lente.
Les itérations 4 et 5 de l'algorithme de gradient sont alors remplacées par : $v_{k+1} = v_k - s\nabla g(v_k)$

L'algorithme de gradient à pas optimal

Une fois la direction $d_k = -\nabla g(v_k)$ choisie, on s'intéresse à ce qui se passe si on choisit le pas s_k de façon optimale, c'est à dire le pas qui minimise la fonction $f(x_k + sd_k)$. L'étape 4 de l'algorithme de descente est alors remplacée par : calculer un pas optimal s_k solution de $\min_s f(x_k + sd_k)$. La recherche du pas optimal à chaque itération peut s'avérer très coûteuse.

L'algorithme de gradients conjugués

Initialement introduit par Hestenes et al. [64], l'algorithme de gradients conjugués consiste à considérer, à l'itération k, la direction de descente $d_k = -\nabla g(v_k) - \beta_k d_{k-1}$ si $k > 0$ et $d_0 = -\nabla g(v_0)$ sinon. On détermine ensuite le pas optimal dans cette direction. Il existe plusieurs méthodes pour déterminer le terme β_k parmi lesquelles on peut citer: la méthode de Fletcher-Reeves [44], la méthode de Polack-Ribière [51] et la méthode de Hestenes-Stiefel [64].

3.1.2 Méthode de Newton

Cette méthode cherche les points critiques de la fonction, c'est à dire les points qui annulent le gradient, ce dernier étant nul à l'optimum.

L'algorithme de Newton

L'algorithme de Newton [42, 52] repose sur une approximation linéaire du gradient de la fonction, ce qui revient à une approximation quadratique de la fonction elle-même.
Supposons maintenant que g est de classe $C2$ avec une matrice hessienne H. L'idée est de remplacer g au voisinage de l'itéré courant x_k par son développement de Taylor de second ordre :

$$g(v_k + d) = g(v_k) + < \nabla g(v_k), d > + \frac{1}{2} < H(v_k)d, d >$$

dont le gradient s'annule pour $d = -[H(v_k)]^{-1}\nabla g(v_k)$.
Tant que $\|\nabla g(v_k)\| > \epsilon$, $\epsilon > 0$ précision demandée, l'algorithme cherche itérativement le point critique par : $v_{k+1} = v_k - [H(v_k)]^{-1}\nabla g(v_k)$.
La principale limite de cet algorithme est qu'il faut faire à chaque itération le calcul et l'inversion de la matrice hessienne. Cette méthode n'est donc praticable que si à chaque itération, la matrice hessienne $H(v_k)$ est définie positive.

L'algorithme de quasi-Newton

Le calcul et l'inversion de la matrice hessienne sont souvent très difficiles à évaluer dans le cas où l'expression analytique de la fonction est inconnue. L'algorithme de quasi-Newton consiste à approcher l'inverse de la matrice hessienne H selon les méthodes suivantes:

- La méthode BFGS (Broyden-Fletcher-Goldfarb-Shanno)[18, 57, 127]

$$H_{k+1}^{-1} = H_k^{-1} + \frac{\delta_k \delta_k^T}{\delta_k^T s_k} - \frac{H_k^{-1} s_k s_k^T H_k^{-1}}{s_k^T H_k^{-1} s_k}$$

- La méthode DFP (Davidon-Fletcher-Powell)[43, 30]

$$H_{k+1}^{-1} = H_k^{-1} + \frac{s_k s_k^T}{s_k^T \delta_k} - \frac{H_k^{-1}\delta_k \delta_k^T H_k^{-1}}{\delta_k^T H_k^{-1} \delta_k}$$

où $s_k = v_{k+1} - v_k$ et $\delta_k = \nabla g(v_{k+1}) - \nabla g(v_k)$.
On choisit une matrice H_0^{-1} définie positive; en général on prend H_0^{-1} égale à la matrice identité.

3.1.3 Méthode du gradient stochastique

Pour une fonction bruitée f, si le coût de calcul d'une espérance ou d'un quantile peut être assumé, on retombe exactement sur le cas de l'optimisation déterministe que l'on peut résoudre avec les méthodes citées précédemment. Le problème est que cette approche peut être extrêmement coûteuse.
Dans ce cadre, Robbins *et al.* [117] d'une part, Kiefer *et al.* [86] d'autre part, ont mis en place un algorithme dit algorithme de gradient stochastique. Cet algorithme est une approximation de l'algorithme du gradient déterministe. La méthode consiste à mettre en œuvre un algorithme au cours duquel la variable à optimiser v évolue en fonction du gradient partiel de la fonction f par rapport à v évalué pour des réalisations successives de la variable aléatoire C_u et non en fonction du gradient de g.
L'algorithme du gradient stochastique est le suivant :

1. Choisir un point initial v_0, et une suite $(s_k, k \in N)$ de réels positifs.
2. A l'itération k, effectuer un tirage C_u^{k+1} de la variable C_u indépendamment des tirages précédents $(C_u^1, \cdots, C_u^k)$.
3. Calculer le gradient partiel de f en (C_u^{k+1}, v_k) et mettre à jour v: $v_{k+1} = v_k - s_k \nabla f(C_u^{k+1}, v_k)$
4. Incrémenter l'indice k de 1 et retourner à l'étape 2.

Les conditions nécessaires sur les pas pour une convergence presque-sûre vers la solution v^* sont : $\sum_{k\in N} s_k = +\infty$ et $\sum_{k\in N} s_k^2 < +\infty$.
En ce qui concerne les tests d'arrêt, en pratique, on se contentera souvent de fixer un nombre d'itération suffisamment grand.

Les méthodes d'optimisations avec calcul du gradient sont simples à mettre en œuvre. Cependant, dans les cas où nous ne disposons de moyens pour calculer ce gradient (trop coûteux à calculer ou non calculables), l'utilisation des méthodes d'optimisation sans calcul de dérivées devient utile. Ces méthodes utilisent uniquement les évaluations de la fonction à optimiser pour guider l'échantillonnage des nouveaux points. Dans les sections suivantes, nous présentons tout d'abord des méthodes de recherche directe,la méthode du Simplexe de Nelder-Mead et la méthode de Pattern Search, qui comparent toutes les deux les valeurs de la fonction aux points évalués

pour se déplacer suivant des critères géométriques. Ensuite, nous présentons les méthodes d'optimisation globales (les algorithmes évolutionnaires, le recuit simulé). Enfin, l'optimisation à base de métamodèle est présentée. Cette méthode est particulièrement intéressante dans le cas de problème d'optimisation de fonction coûteuse.

3.1.4 Les méthodes de recherche directe

Les méthodes de recherche directe utilisent uniquement les réponses de la fonction pour localiser le minimum. Ces méthodes ne requièrent donc pas d'informations sur les dérivées. Les méthodes de recherche directe peuvent être classées en trois catégories: les méthodes de recherche par motifs généralisés (generalized pattern search, GPS) [134, 136, 135], les méthodes des directions conjuguées (algorithme de Powell et ses variantes) [119, 109, 110, 47], et les méthodes basées sur la figure géométrique d'un simplexe (méthode de Nelder-Mead et ses variantes) [144, 101].

Recherche par motifs généralisés

Les generalized pattern search methods (GPS) sont une généralisation de la méthode de Hooke et Jeeves [66]. Les méthodes GPS sont basées sur un concept simple. Elles utilisent un ensemble de directions prédéfinies $\mathbf{D}$ dans l'espace des paramètres afin de faire évoluer l'algorithme vers la zone où se trouve le minimum. Ces déplacements exploratoires autour du point courant forment des motifs (pattern) qui présentent une disposition invariable. A chaque itération la fonction objectif est évaluée sur les points du motif $M_k = \{v_k + s_k d, d \in \mathbf{D}\}$ (v_k est le point courant et s_k la valeur courante pour le pas d'échantillonnage). Si une amélioration est trouvée, le point associé est accepté comme nouveau point courant, et le pas du prochain motif est conservé ou augmenté. Sinon, le pas du nouveau motif, généré autour de l'ancien point courant, est réduit.
L'algorithme peut être résumé comme suit:

1. Choisir un point initial v_0, le pas initial s_0 et l'ensemble des directions $\mathbf{D}$.
2. Évaluer la fonction sur les points du motif $M_k = \{v_k + s_k d, d \in \mathbf{D}\}$.
   S'il existe un point $v_k + s_k d$ qui vérifie $g(v_k + s_k d) < g(v_k)$, alors $v_{k+1} = v_k + s_k d$, sinon $v_{k+1} = v_k$.
3. Mettre à jour les motifs, s'il existe un point qui apporte une amélioration, le pas du motif est maintenu ou augmenté. Sinon le pas sera réduit.

Méthodes des directions conjuguées

Les méthodes de directions conjuguées ont été introduites par Rosenbrock [119] et Powell [109]. Ces méthodes définissent les directions de recherche au fur et à mesure des itérations. Afin d'accélérer la convergence de l'algorithme, elles prennent en

compte les informations sur la courbure de la fonction à optimiser pour la construction des directions.

Méthode de Nelder-Mead

La méthode de Nelder-Mead [101] est une méthode d'optimisation qui utilise une figure géométrique, particulièrement un simplexe, pour localiser le minimum.
Soit d la dimension de l'espace de définition de la fonction à optimiser. L'algorithme définit un simplexe de dimension d et lui fait subir, au cours des itérations, des transformations (expansion, contraction, rétrécissement) afin que les sommets du simplexe se rapprochent du minimum de la fonction.
Plus précisément, les différentes étapes de l'algorithme peuvent être résumées comme suit:

1. Choisir $p+1$ points $\{v_1, \cdots, v_{p+1}\}$ qui forment un simplexe dans l'espace de définition à d dimensions.
   Calculer les valeurs de la fonction en ces points et réindexer les points de façon à avoir $g(v_1) \leq g(v_2) \leq \cdots \leq g(v_{p+1})$

2. Calculer le centre de gravité v_0 des points $\{x_1, \cdots, x_{p+1}\}$ et de la réflexion v_r de v_{p+1} par rapport à v_0: $v_0 = \sum_{i=1}^{p} v_i/n$ et $v_r = v_0 + (v_0 - v_{p+1})$.

3. Si $g(v_r) < g(v_p)$, calculer $v_e = v_0 + 2(v_0 - v_{p+1})$ (étirement du simplexe), et si $g(v_e) < g(v_r)$, remplacer le point v_{p+1} par le point v_e.
   Sinon remplacer v_{p+1} par v_r.

4. Si $g(v_p) < g(v_r)$, calculer $v_c = v_{p+1} + \frac{1}{2}(v_0 - v_{p+1})$ (contraction du simplexe) et si $g(v_c) \leq g(v_p)$, remplacer v_{p+1} par v_c.
   Sinon remplacer v_i par $v_1 + \frac{1}{2}(v_i - v_1)$ pour $i \geq 2$

3.2 Principales méthodes d'optimisation globale

Les méthodes d'optimisation locale permettent une bonne recherche locale en exploitant les points déjà évalués, mais au prix d'une exploration moindre. Elles ne garantissent donc pas la convergence vers un optimum global et peuvent être piégées dans un optimum local. Contrairement aux méthodes locales, les méthodes globales permettent d'explorer l'espace de recherche de manière plus exhaustive. Ce qui augmente le nombre de chance de tomber sur l'optimum global. A titre d'exemple, trois approches d'optimisation globale s'appuyant sur des concepts différents sont brièvement présentées ci-dessous.

3.2.1 Algorithme évolutionnaire

Les algorithmes évolutionnaires sont des méthodes d'optimisation stochastiques dans lesquelles un ensemble de points dans $\mathbb{E}$ est produit à chaque itération au moyen d'opérations stochastiques inspirées de la théorie de l'évolution selon Darwin [29]. Par analogie avec l'évolution naturelle, un algorithme évolutionnaire est un processus itératif faisant évoluer un ensemble de solutions candidates, appelé population, composé d'individus représentant des solutions possibles du problème d'optimisation. Au cours de chaque itération, appelée génération, la performance des individus d'une population appelée population parent est évaluée à l'aide de la fonction à optimiser. Les meilleurs individus sélectionnés en fonction de leurs valeurs, subissent des opérations génétiques (croisement, mutation et sélection) pour créer de nouveaux individus. Ces nouveaux individus créés forment une nouvelle population appelée population enfant. L'algorithme se poursuit jusqu'à ce qu'un critère d'arrêt soit satisfait.
Pour un problème de minimisation, le principe de fonctionnement d'un algorithme évolutionnaire peut être résumé comme suit:

1. Générer aléatoirement un ensemble initial d'individus nommé *population parent*.
2. Évaluer chaque individu de la population parent.
3. Appliquer les trois opérations génétiques "sélection, croisement et mutation"pour créer un ensemble d'individus appelé *population enfant*.
4. Évaluer chaque individu de la population enfant.
5. Sélectionner les meilleurs individus de la population enfant pour créer une population parent de la nouvelle génération.

Répéter les étapes 3, 4 et 5 jusqu'à ce que l'algorithme converge.
On distingue plusieurs familles de méthodes évolutionnaires: les algorithmes génétiques [65, 37], les stratégies d'évolution [141], la programmation évolutionnaire [45].

3.2.2 Le recuit simulé

Le recuit simulé [87, 88] se base sur un phénomène observé en métallurgie. Dans le recuit simulé, chaque paramètre x de l'espace de recherche est considéré comme l'état d'un système physique et sa réponse $f(x)$, la fonction à optimiser, comme l'énergie interne de ce système. On introduit également un paramètre fictif, la température T du système, dans le but de régler la probabilité d'acceptation des nouveaux points. L'objectif est d'amener progressivement le système d'un état initial arbitraire vers un état correspondant au minimum global d'énergie.
Étant donné un état initial x_0 arbitrairement choisi et une température initiale T_0 élevée, l'algorithme génère une suite de points x_n comme suit:
A chaque itération i de l'algorithme une modification élémentaire de l'état du système

est effectuée en générant aléatoirement un point x dans un voisinage de x_i. Cette modification va entrainer une variation de l'énergie du système. Si cette variation correspond à une baisse de l'énergie du système, c'est à dire si $f(x) < f(x_i)$ alors le jeu de paramètre y est retenu; sinon il est accepté avec une probabilité égale à $\exp\left(\frac{f(x_i) - f(y)}{T_i}\right)$.
En résumé le nouveau point x_{i+1} est choisi comme suit:

$$\begin{cases} x_{i+1} = y \ si \ f(y) \leq f(x_i) \\ x_{i+1} = y \ si \ \exp\left(\frac{f(x_i) - f(x)}{T_i}\right) > u \\ x_{i+1} = x_i \ sinon \end{cases}$$

où u est un nombre choisi aléatoirement compris entre 0 et 1, (T_n) une suite de paramètre positif qui tend vers 0 au cours de l'algorithme. L'algorithme peut sélectionner des points qui n'apportent aucune amélioration. Cependant, la décroissance de la suite (T_n) fera que la probabilité d'accepter une mauvaise solution décroit au cours des itérations.
Le recuit simulé est une méthode d'optimisation qui s'adapte à plusieurs types de problèmes. Cependant la convergence est relativement lente et le réglage de la suite (T_n) délicat.

Il existe d'autres algorithmes d'optimisation globale tels que les méthodes basées sur l'intelligence en essaims (les algorithmes des colonies de fourmis [35], les algorithmes par essaims particulaires [85]) et la recherche tabou [56].
Les algorithmes d'optimisation globale sont le plus souvent inspirés de phénomènes naturels tirés de la biologie, la physique etc. Contrairement aux méthodes d'optimisation avec calcul de gradient, les algorithmes d'optimisation globale n'imposent pas d'hypothèses de régularité sur la fonction à optimiser. Ils ont une bonne capacité d'exploration de l'espace de recherche, ce qui leur donne plus de chance de converger vers l'optimum globale. Cependant ils sont très coûteux en terme de temps.

3.2.3 Optimisation à base de métamodèle

L'utilisation de métamodèles est une solution fréquemment utilisée pour faciliter l'optimisation des modèles complexes et coûteux à évaluer [11, 70, 7, 39, 60, 93]. L'objectif de la métamodélisation est de remplacer un modèle coûteux en temps par un modèle à coût négligeable, afin de pouvoir exploiter le modèle plus facilement. Le métamodèle est une approximation des réponses du modèle étudié, obtenue à partir d'un nombre limité d'évaluation du modèle dans un domaine appelé plan d'expérience. On évalue le modèle pour un certain nombre n de valeurs $D = \{x_1, \cdots, x_n\}$ et on observe les n réponses associées $Y = \{y(x_1), \cdots, y(x_n)\}$. Le gradient de la fonction à optimiser étant inconnu en général, le choix des prochains points à évaluer se fait à l'aide d'un critère d'échantillonnage défini à l'aide du métamodèle. Il existe plusieurs types de métamodèle parmi lesquels on peut citer : les interpolateurs locaux, les techniques polynomiales [75], les splines [124], l'interpolation à noyaux [123] et les réseaux de neurones [130]. Certains métamodèles comme le krigeage et la régression linéaire per-

mettent de modéliser l'incertitude de prédiction. De tels métamodèles sont particulièrement adaptés pour l'optimisation globale. L'estimation de l'incertitude peut guider la phase exploratoire de l'optimisation globale où nous cherchons dans des régions à forte incertitude. Après avoir testé des métamodèles de manière empirique sur différents exemples, Simpson *et al.* [128] proposent d'utiliser un métamodèle de régression linéaire si la fonction f est assez régulière et un métamodèle de krigeage dans le cas ou elle est fortement non linéaire et de dimension raisonnable.

3.3 Quelle méthode pour la recherche de génotype optimal ?

Les méthodes directes, telles que la méthode de Nelder-Mead, n'utilisent pas les informations obtenues lors des itérations précédentes. Donc pour converger vers l'optimum, l'algorithme nécessitera un nombre important d'évaluation de la fonction à optimiser, ce qui peut être limitant pour un problème d'optimisation coûteux.
Les méthodes d'optimisation avec calcul de gradient disposent de nombreuses qualités; en effet, elles ne nécessitent pas un nombre trop important d'itération pour converger vers l'optimum, elles n'imposent pas de limites par rapport à la dimension du problème. Cependant, elles sont souvent écartées pour la résolution de problèmes d'optimisation complexes comme le nôtre. En effet: d'une part, les modèles de culture sont constitués d'une suite d'équations imbriquées les une dans les autres et le calcul des dérivées ou de leurs estimations à l'aide de modèle de substitution s'avère très complexe, voire impossible; d'autre part, les sorties du modèle sont bruitées, le calcul numérique des gradients va vers des résultats complètement aléatoires et mène l'optimisation à l'échec.
Une alternative serait l'utilisation d'un algorithme stochastique. Contrairement aux méthodes avec calcul de gradient, les algorithmes stochastiques ne font pas d'hypothèses particulières sur les propriétés de la fonction f, et ont par conséquent une large gamme d'application. Les algorithmes stochastiques sont utilisés en raison de leur capacité à converger vers un optimum global. Cependant, comme nous voulons trouver l'optimum pour chaque lieu fixé, l'utilisation d'un algorithme stochastique coûtera très cher. En effet, ces algorithmes nécessitent en général un nombre important d'évaluations afin d'obtenir une optimisation de qualité [68]. Nous avons un modèle aléatoire, on va optimiser l'espérance ou son quantile, ce qui nécessitera plus de simulations. Le problème du bruit peut être traité par l'utilisation d'un optimiseur acceptant du bruit dans les critères à optimiser (par exemple un algorithme évolutionnaire [8]). Ces méthodes sont très coûteuses, car on devra estimer les incertitudes sur C_u et optimiser sur v. Une simulation sur 10 ans dure, disons, une seconde. Pour estimer correctement une moyenne simulée, on a coutume de dire qu'il faut 30 ans de simulations, soit 3 secondes. Pour un quantile 0,20 ce sera plus long. Une optimisation par une méthode stochastique comme l'algorithme génétique demande un millier d'appels au modèle, soit environ 1 heure. Si l'on veut une carte des optimums avec un point tous les 10 degrés sur une étendue de L degrés de longitude et l degrés de la latitude, et que l'on choisit d'optimiser en chaque point séparément, il faudra plusieurs jours. On a donc besoin d'une méthode plus économe en nombre d'appels que les algorithmes

stochastiques. Une solution possible serait d'optimiser sur la base d'un métamodèle qui va réduire de façon considérable le temps de calcul. Un algorithme d'optimisation local est inadapté aux modèles stochastiques, un algorithme évolutionnaire nécessite un nombre important d'évaluation pour converger ver l'optimum global, donc coûte trop cher. Pour notre problème, d'après nos calculs, il faudrait 40 jours pour déterminer une carte d'idéotypes de 1000 points. Il faudrait donc réussir à limiter le nombre d'évaluation en vue de faire l'optimisation. Une solution serait de modéliser la surface de réponse en sachant qu'elle est bruitée. C'est une des possibilités des métamodèles. Une autre approche serait donc de faire l'optimisation à base d'un métamodèle de krigeage. Le krigeage permet de modéliser l'incertitude de prédiction, et il est donc particulièrement adapté pour l'optimisation globale. En effet, une recherche dans des régions à forte incertitude peut guider la phase exploratoire de l'optimisation globale.

3.3.1 Notions sur le krigeage

Le krigeage est une technique d'interpolation d'une fonction f à partir de ses réalisations observées en les points d'un plan d'expérience $(x_1, \cdots, x_n)$. Initialement utilisée en géostatistique avec les travaux de l'ingénieur minier sud-africain Krige [89], elle a été développée plus tard par le mathématicien français Matheron [98]. Le krigeage consiste à prédire une fonction f aux points non observés conditionnellement aux valeurs observées aux autres points, en considérant la fonction f comme la réalisation d'un processus gaussien Y

$$Y(x) = m(x) + \delta(x)$$

où x représente le vecteur des variables d'entrée, $m(.)$ est la structure déterministe pour l'espérance de $Y(.)$ et δ un processus gaussien d'espérance nulle et de fonction de covariance donnée par

$$cov(\delta(x), \delta(x+h)) = k(x, x+h) = \sigma^2 R(x, x+h)$$

avec σ^2 la variance et $R(x, x+h)$ une fonction de corrélation. Selon la forme de la tendance $m(.)$, il existe trois types de krigeage qui sont:

- Le krigeage simple lorsque $m(x)$ est une constante connue.
- Le krigeage ordinaire lorsque $m(x)$ est une constante inconnue.
- Le krigeage universel lorsque $m(x)$ est une combinaison linéaire de fonctions $m(x) = \sum_{j=1}^{p} \beta_j \mu_j(x)$

Dans ce manuscrit, seul un rappel des principales notions sur le krigeage universel est donné (pour plus de détails, voir Cressie [26]).
Conditionnellement aux observations, la loi de $Y(x)$ est gaussienne, de moyenne $m_n(x)$ et de variance $s_n^2(x)$.

$$[Y(x)|Y(x_1) = f(x_1), \cdots, Y(x_n) = f(x_n)] \sim N(m_n(x), s_n^2(x)).$$

Pour une fonction de covariance k donnée, les expressions analytiques de la moyenne de prédiction $m_n(x)$ et de la variance de prédiction $s_n^2(x)$ sont données par [26]:

$$m_n(x) = \left(\gamma(x) + \mu(\mu^t K^{-1}\mu)^{-1}(\mu(x) - \mu^t K^{-1}\gamma(x)\right)^t K^{-1}Y \tag{3.1}$$

$$s_n^2(x) = \gamma(x)^t K^{-1}\gamma(x) - \left(\mu(x) - \mu^t K^{-1}\gamma(x)\right)^t \left(\mu^t K^{-1}\mu\right)^{-1} \left(\mu(x) - \mu^t K^{-1}\gamma(x)\right) \tag{3.2}$$

où $\mu(x) = (\mu_1(x), \cdots, \mu_p(x))^t$ est le vecteur de fonctions connues, $\mu = (\mu_j(x_i))_{1\leq i\leq n, 1\leq j\leq p}$ la matrice $n \times p$ d'expérience, $K = (k(x_i, x_j))_{1\leq i\leq n, 1\leq j\leq n}$ la matrice $n \times n$ de covariance, $Y = (f(x_1), \cdots, f(x_n))$ les valeurs de la fonction aux points d'observation, $k(x) = (k(x_1, x), \cdots, k(x_n, x))^t$ le vecteur de covariance entre un point x et les points du plan d'expérience.La moyenne $m_n(x)$ sera utilisée pour approcher $f(x)$ et $s_n^2(x)$ décrira l'incertitude associée à cette prédiction. Un exemple de prédiction d'une fonction de dimension 1 par krigeage est donné par la figure 3.1. Les résultats des évaluations (points rouges) sont interpolés par la courbe verte en gras. Les courbes en pointillés représentent les intervalles de confiance à 95%.
Le krigeage est un interpolateur exact des points du plan d'expérience. C'est le meilleur prédicteur linéaire sans biais. En se basant sur les observations réparties sur le plan d'expérience, le krigeage permet de trouver une approximation $m_n(.)$ de la réponse $f(.)$ du modèle en dehors du plan d'expérience. En ce sens, le choix du plan d'expérience devient important.

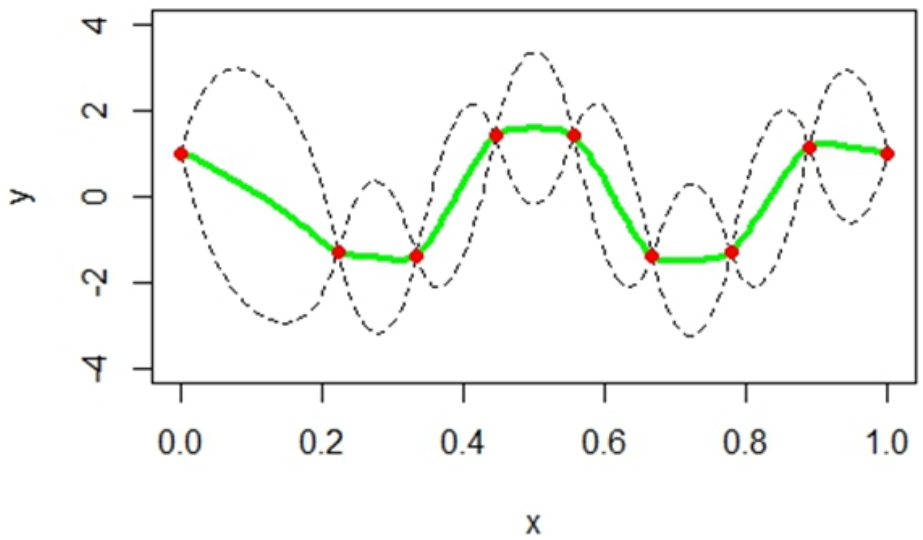


FIGURE 3.1 – *Prédiction par le krigeage et erreur de prédiction*

Par exemple les plans d'expériences où les points sont répartis selon une grille régulière dans l'espace sont classiquement utilisés dans la construction des métamodèles déterministes mais ne sont pas adaptés pour le krigeage. En effet, dans ces plans la majeure partie des points sont placés au bords du domaine expérimental, ce qui ne permet

pas de détecter les éventuelles irrégularités dans les réponses, avec pour conséquence une estimation par maximum de vraisemblance des paramètres de la fonction de covariance de mauvaise qualité. La caractérisation de la dépendance à distance en dessous du pas de grille peut également se poser.
Deux techniques de planification d'expérience sont utilisées pour le krigeage. La première est géométrique, elle consiste à choisir les points du plan d'expérience de sorte qu'ils explorent au mieux le domaine expérimental. Les techniques géométriques les plus utilisées sont les hypercubes latins [99] et les suites à faible discrépance [102].
La deuxième technique de planification consiste à sélectionner le plan qui maximise la qualité des prédictions en optimisant un critère statistique comme l'erreur quadratique moyenne (MSE) [123] ou sa racine $RMSE$. La construction de ce type de plan d'expérience passe par celle d'un premier métamodèle à partir d'un plan d'expérience initial bien choisi (par exemple un hypercube latin) que l'on enrichira de manière séquentielle en utilisant le modèle de krigeage.
Dans la suite, nous nous restreindrons au cas où le processus gaussien Y est soit stationnaire au second ordre, c'est à dire que l'espérance du processus est constante et la covariance entre $Y(x)$ et $Y(x+h)$ ne dépend que de h, soit stationnaire intrinsèque, c'est à dire l'espérance de tout accroissement $Y(x+h)-Y(x)$ est nulle et sa variance existe et dépend uniquement de h. Pour caractériser la dépendance entre $Y(x)$ et $Y(x+h)$, on introduit le semi-variogramme $\gamma(h)$ défini par:

$$\gamma(h) = \frac{1}{2}E[(Y(x) - Y(x+h))^2]$$

Lorsque la fonction de covariance k existe, on peut montrer que le semi variogramme, fonction de la distance, ou plus généralement du vecteur qui sépare deux points, est égal à la variance totale moins la fonction de covariance:

$$\gamma(h) = k(0) - k(h)$$

Le choix de la fonction de covariance $k(h)$, ou du variogramme $2\gamma(h)$, et l'estimation des paramètres associés ont une influence importante dans les prédictions de krigeage. Les différents types de krigeage présupposent la connaissance de la fonction de covariance k ou du variogramme 2γ, ce qui n'est pas réaliste en pratique.
Une estimation du semi-variogramme peut être obtenue à partir du variogramme expérimental défini via l'estimation des moments [27]:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left(Y(x_i + h) - Y(x_i)\right)^2$$

où $N(h)$ est le nombre de paires de points espacés de h.

A partir de la représentation graphique de $\hat{\gamma}(h)$ en fonction de h, il suffit d'ajuster une fonction analytique choisie parmi les fonctions autorisées, par exemple à l'aide de la méthode des moindres carrés. Nous obtenons alors une fonction continue caractérisant le variogramme en fonction de la distance entre les points et aussi de la direction du vecteur si le variogramme est anisotrope. On dira que le variogramme est isotrope si $2\gamma(h)$ ne dépend que du module $r = |h|$. S'il dépend aussi de la direction du vecteur, on dira que le variogramme est anisotrope. Avec un semis irrégulier de points, chaque

vecteur h est représenté par peu voire un seul couple de points. Aussi en pratique on regroupe les couples par classes de distance et d'orientation du vecteur h, puis on calcule les moyennes par classe pour estimer γ. En pratique il n'est donc pas toujours aisé d'estimer le variogramme. A noter que quand une tendance existe on ne peut pas estimer le variogramme directement à partir des observations.
Une autre approche serait de choisir le variogramme ou la fonction de covariance par validation croisée, en comparant les performances en prédiction de plusieurs métamodèles construits avec différentes fonctions de covariance (noyaux de type exponentiel, noyaux cubiques, noyaux de Matérn). Les paramètres de k peuvent être estimés soit par maximum de vraisemblance [96], soit encore une fois par une minimisation de l'erreur de validation croisée [36, 139].

Dans le cas où la fonction f est bruitée, chaque sortie $f(x_i) = \tilde{y}_i$ de f est considérée comme une réalisation d'une variable aléatoire somme des réalisations d'un processus $y(x)$ et d'un bruit blanc gaussien ϵ_i:

$$\tilde{y}_i = y(x_i) + \epsilon_i, 1 \leq i \leq n. \tag{3.3}$$

On supposera que les bruits ϵ_i sont indépendants, identiquement distribués, indépendants du processus Y et que $\epsilon_i \sim N(0, \tau^2)$.
Le principe est le même que pour le krigeage d'une fonction déterministe. A partir des évaluations de la fonction f aux points d'un plan initial $D = \{x_1, \cdots, x_n\}$, le krigeage consiste à considérer la fonction f comme la réalisation d'un processus $\tilde{Y}(x_i) = Y(x_i) + \epsilon_i$.
Il est montré dans [53] que si le processus Y et le bruit gaussien $(\epsilon_i, 1 \leq i \leq n)$ sont stochastiquement indépendants, la distribution du processus Y conditionnellement aux événements

$$\tilde{E}_n = \{Y(x_1) + \epsilon_1 = \tilde{y}_1, \cdots, Y(x_n) + \epsilon_n = \tilde{y}_n\}$$

est normale.
Conditionnellement à n observations, les expressions de la moyenne $m_n(x)$ et de la variance de krigeage $s_n^2(x)$ sont les mêmes que celles données par le krigeage d'un modèle déterministe [106]. La seule différence réside dans le fait que la matrice de covariance $K = (k(x_i, x_j))_{1 \leq i \leq n, 1 \leq j \leq n}$ est remplacée par $\tilde{K} = K + \Gamma$ où Γ est la matrice diagonale de la variance du bruit aléatoire. Dans ce cas, le modèle de krigeage n'est plus une interpolation exacte des réponses du modèle.

3.3.2 Choix du critère d'échantillonnage pour l'optimisation à base de métamodèle

Sachant que nous avons un budget de simulations limité, plutôt que de calculer l'optimum indépendamment en chaque point u, nous envisageons d'ajuster une surface de réponse qui approche la fonction f.
Le principe de l'optimisation à base de métamodèle, brièvement décrit en section 2.2.3, consiste tout d'abord à construire le métamodèle qui donne une approximation du modèle à partir d'un plan initial d'expérimentation; ensuite à définir un critère d'échantillonnage qui permet d'enrichir de manière "intelligente" le plan initial par l'ajout de

nouveaux points. Les nouveaux points sont ceux qui optimisent le critère d'échantillonnage. L'intérêt de l'optimisation à base de métamodèle est donc de remplacer un problème d'optimisation complexe ou coûteux par une série d'optimisation simple à résoudre.
L'optimisation à base de métamodèle de krigeage peut être décrite selon les différentes étapes suivantes:

1. Réalisation d'un plan initial.
2. Construction du métamodèle à partir du plan d'expérience.
3. Recherche du point ou des points qui optimise(nt) un critère d'échantillonnage donné.
4. Calcul de la fonction en ce(s) point(s).
5. Ajout du point ou des points au plan d'expérience et mise à jour du métamodèle.

Les étapes 3, 4 et 5 sont répétées n fois. Le nombre n d'itérations est fixé à l'avance en fonction du budget de l'utilisateur.

Pour déterminer le critère d'échantillonnage pour disons la maximisation d'une fonction, une première idée serait de choisir l'espérance prédite (3.1) en retenant pour prochain point celui où cette espérance prédite (3.1) est maximale. Cependant, dans le cas le plus aisé du krigeage simple où $\mu(x)$ est réduit à la valeur 1, ce point est tout simplement le point du plan d'expérience où la valeur observée est maximale. Choisir comme critère l'espérance prédite conduirait alors à exploiter sans fin le premier minimum observé.
Une autre idée naturelle serait de choisir comme critère la variance de prédiction (3.2) en retenant à chaque étape le point qui la maximise, comblant ainsi séquentiellement les zones mal couvertes par le plan d'expérience. Ce critère permet une bonne exploration de l'espace mais ne tient pas du tout compte des valeurs observées $\mathbf{y}$. En effet, la variance de prédiction (3.2) ne dépend que du plan d'expérience; il en est de même des critères MSE [122] et $IMSE$ [74], dont la minimisation se ramène à la minimisation des distances entre points d'observations.
Il est donc important de choisir un critère d'échantillonnage qui fait un compromis entre exploitation (recherche locale autour des maxima courants) et exploration (recherche dans les régions les plus inexplorées). Comme le montrent les exemples agronomiques présentés ci-dessous, le critère doit également prendre en compte le type de problème posé.
Du modèle de culture, nous regroupons les entrées environnementales dans un vecteur u, et les paramètres variétaux dans un vecteur v. Dans le vocabulaire de la métamodèlisation, u comme v sont des entrées.
L'idée est de déterminer par simulation quel est le meilleur v pour un u donné et de proposer ce v optimal comme but pour l'amélioration variétale, autrement dit un idéotype [34].
Selon que u est fixé ou pas, déterministe ou stochastique, plusieurs types de problème

d'optimisation découlent de notre objectif.

Un premier cas d'intérêt concerne le cadre d'une optimisation simple. Quand l'environnement u est connu à l'avance, ce qui est peut être le cas par exemple dans une serre où toutes les conditions environnementales sont contrôlées, le problème est de trouver v assurant un optimum global du rendement prédit par le modèle de culture. Une méthode d'optimisation appelée Efficient Global Optimisation (EGO) [76], basée sur un métamodèle de krigeage permet de résoudre ce problème. Elle repose sur l'enrichissement séquentiel du plan d'expérience initial par l'ajout de nouveaux points qui maximisent un critère appelé Expected Improvement (EI) [76]. Le critère est défini comme l'espérance du dépassement du maximum des réponses déjà observées sur le plan d'expérience, appelé maximum courant.

Un deuxième cas d'intérêt serait de considérer une approche d'optimisation conditionnelle. Typiquement, on peut contrôler parfaitement les entrées, mais en les faisant varier avec la saison: en hiver où le chauffage coûte cher, on peut choisir une température plus basse qu'en été où c'est la ventilation qui coûte cher. Faire varier le vecteur u décrivant les conditions environnementales va déplacer l'optimum sur v. Le problème est alors de déterminer un v optimal conditionnellement à un vecteur u variable d'une serre à l'autre ou d'une saison à l'autre. Il s'agit alors de résoudre un problème d'optimisation d'un modèle déterministe conditionnellement à une partie de ses entrées. Plutôt que de se ramener à la recherche pour chaque valeur de u d'un optimum sur v à l'aide d'un métamodèle fonction de v, une alternative est de construire un métamodèle unique fonction de u et v et d'en déduire le lieu des optimums conditionnels ou ligne de crête. La mémoire est conservée d'une valeur de u à l'autre et on espère ainsi une meilleure estimation de la carte des optimums qu'avec des optimisations séparées.
Une adaptation de la méthode EGO a été développée à cet effet [55]. Cette méthode utilise les informations de l'optimum conditionnellement à une valeur u pour la recherche de l'optimum dans les autres endroits.

Un troisième cas d'intérêt serait de considérer une approche d'optimisation d'un modèle bruité. Typiquement, en dehors d'une serre, une culture subit des aléas climatiques: les entrées météorologiques du modèle $f(u, v)$ deviennent stochastiques. On ne peut pas optimiser une fonction dont une partie des entrées est aléatoire, mais on peut optimiser l'espérance ou un quantile de cette fonction. Toutefois si son expression analytique n'est pas connue, il faut estimer l'espérance ou le quantile par simulation. On n'a alors accès qu'à des estimations bruitées, et il faut optimiser une fonction dont la sortie est bruitée. Or, pour l'optimisation globale d'un modèle bruité, la valeur du maximum courant est inconnue et le critère EI utilisé dans l'approche EGO n'est plus valide. Plusieurs extensions de la méthode EGO ont donc été développées sur la base de nouveaux critères à maximiser. Parmi ceux-ci, on peut citer le critère d'amélioration espérée modifié (AEI) [67], le gradient de connaissance approximatif (AKG) [125], le critère de quantile minimal (MQ) [25], l'amélioration espérée sur les quantiles de krigeage (EQI) [106] et la procédure de réinterpolation (RI) [69].

Un quatrième cas d'intérêt serait celui de l'optimisation conditionnelle d'un modèle bruité. Toujours en environnement aléatoire, la distribution de probabilité des entrées météorologiques varie d'un lieu à l'autre, et avec cette variation l'optimum sur v est déplacé. Il faudrait donc envisager l'optimisation d'un modèle bruité conditionnellement aux paramètres du climat. Mais à la différence des autres situations, aucun critère

d'échantillonnage spécifique n'existe à notre connaissance pour l'optimisation conditionnelle d'un modèle bruité.
Dans cette section, nous présentons les principaux critères existants pour différents cadres d'optimisation globale: optimisation d'une fonction déterministe, optimisation conditionnelle d'une fonction déterministe et optimisation d'une fonction bruitée.

3.3.3 Optimisation globale d'une fonction déterministe

Nous considérons que f est la réalisation d'un processus gaussien Y de moyenne et de fonction covariance connue. Nous supposons également que n évaluations de la fonction à optimiser sont faites.
Notons $y_{max,n} = \max\{Y(x_1), \cdots, Y(x_n)\} = \max\{f(x_1), \cdots, f(x_n)\}$ le maximum courant et respectivement $m(x)$ et $s(x)$ la prédiction et l'écart-type de l'erreur de prédiction, obtenus par krigeage au point x.

Probabilité d'amélioration espérée

La probabilité d'amélioration espérée [90] est le premier critère proposé pour l'optimisation à base de krigeage. Elle permet de déterminer le point de l'ensemble $\mathbb{E}$ où la probabilité d'être supérieure à $y_{max,n}$ est maximale. Ce point constituera le nouveau point à évaluer. Conditionnellement à $y = \{Y(x_1), \cdots, Y(x_n)\}$, chaque $Y(x) \sim N(m(x), s^2(x))$. La probabilité d'amélioration en un point x s'écrit :

$$\begin{aligned} PI(x) &= Pr(Y(x) \geq y_{max,n}) \\ &= \Phi\left(\frac{m(x) - y_{max,n}}{s(x)}\right) \end{aligned}$$

où Φ représente la fonction de répartition de la loi normale centrée réduite. Le prochain point à évaluer est celui qui maximise la fonction PI. Le problème de ce critère est qu'il favorise l'exploitation au profit d'une exploration. Il y aura ainsi beaucoup plus d'évaluations au voisinage du maximum courant. Pour pallier ce problème, Jones [75] a défini une variante de ce critère en considérant la probabilité d'amélioration par rapport à des seuils $T_n = y_{max,n} + \varepsilon_n$ où (ε_n) est une suite positive tendant vers zéro lorsque n augmente. A l'étape n de l'optimisation, cette variante est définie par:

$$\begin{aligned} PI(x) &= Pr(Y(x) \geq T_n) \\ &= \Phi\left(\frac{m(x) - T_n}{s(x)}\right) \end{aligned}$$

Une petite valeur de ε_n entrainera une recherche autour du maximum courant, tandis qu'une valeur élévée de ε_n favorisera une exploration plus forte.

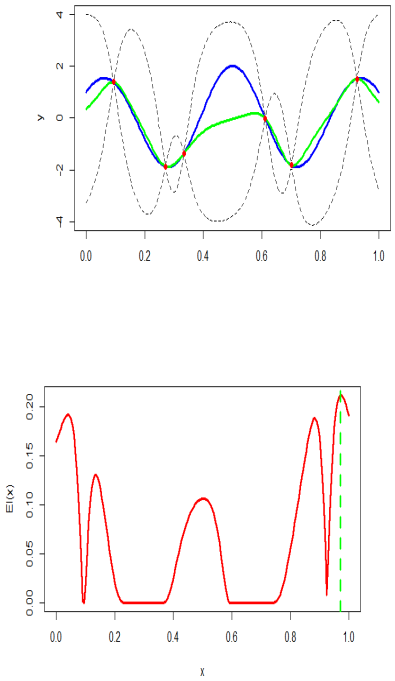
Amélioration espérée pour l'optimisation d'un modèle déterministe

Un critère qui permet de contourner les problèmes liés à l'utilisation de la probabilité d'amélioration est le critère d'amélioration espérée EI. Initialement introduit par Mockus *et al* [100], puis utilisé dans l'article de Jones et *et al* [76] pour la méthode EGO, il est défini comme suit:
si f est une fonction à maximiser, l'amélioration est le dépassement d'un seuil égal à la valeur maximale observée sur le plan d'expérience.
Soit à nouveau $y_{max,n} = \max_{1 \leq i \leq n} f(x_i)$ ce maximum observé, l'amélioration à un nouveau point x est:

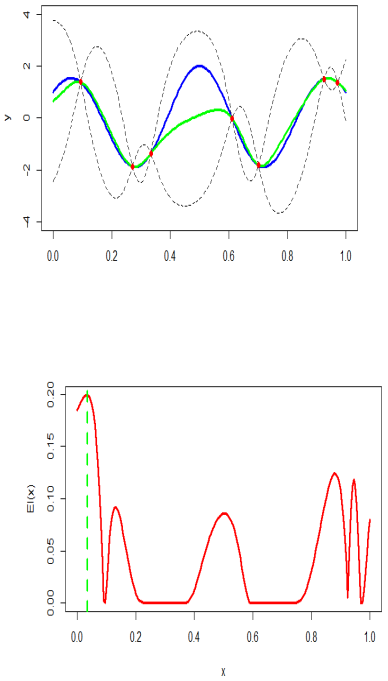
$$I(x) = [f(x) - y_{max,n}]_+ \tag{3.4}$$

où $[z]_+ = \max(0, z)$.
Cette amélioration ne peut pas être connue à l'avance car la fonction f n'est pas encore évaluée au point x, mais on suppose qu'elle est la réalisation d'un processus gaussien Y de moyenne et de covariance connues. L'amélioration est alors définie comme suit:

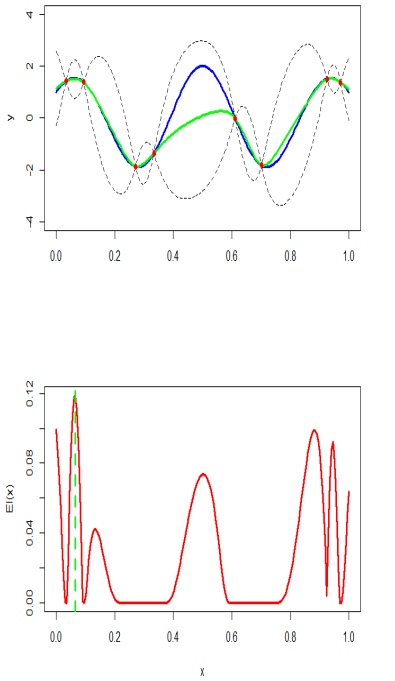
$$I(x) = [Y(x) - y_{max,n}]_+ \tag{3.5}$$



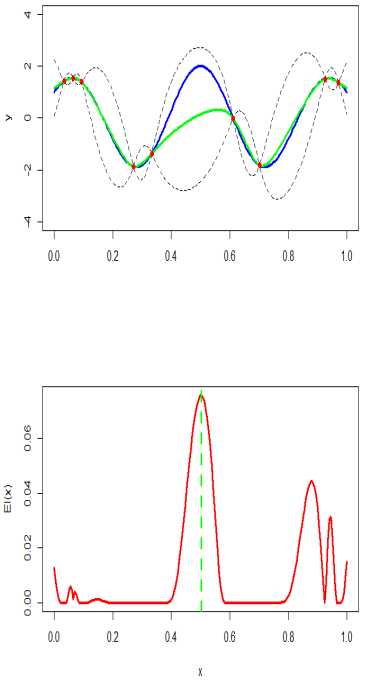
(a) Itération 1



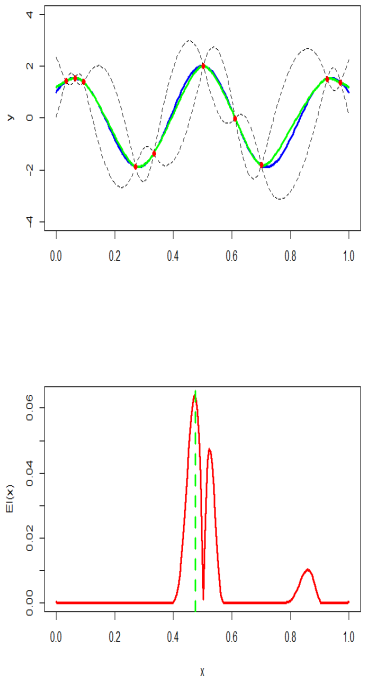
(b) itération 2



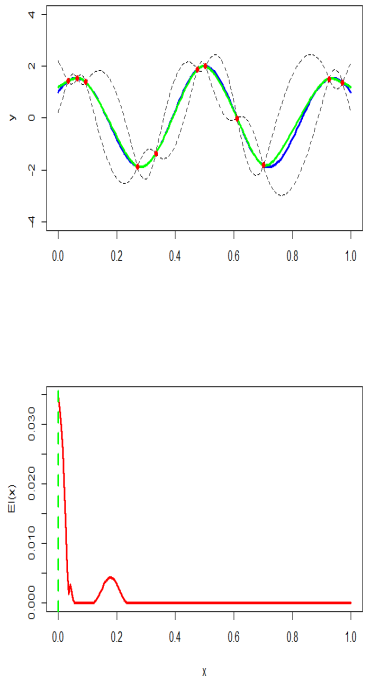
(a) Itération 3



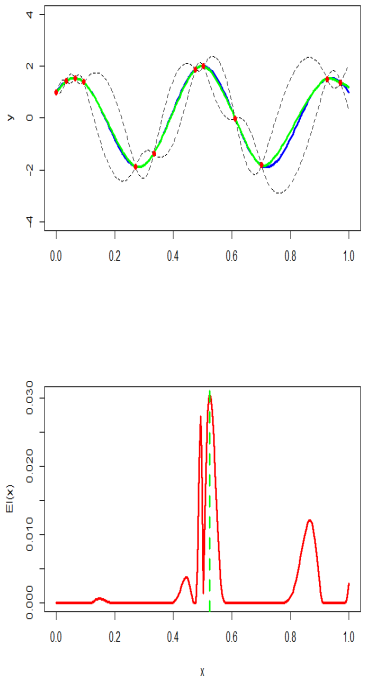
(b) itération 4



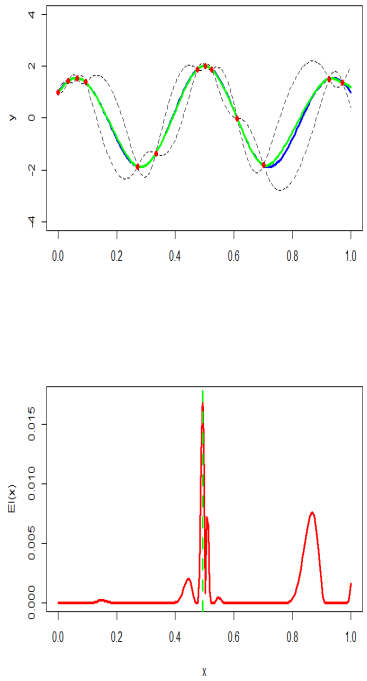
(a) Itération 5



(b) itération 6



(a) Itération 7



(b) itération 8

FIGURE 3.2 – *Illustration d'une optimisation avec l'algorithme EGO. La fonction objectif (trait plein bleu), le prédicteur (trait plein vert), les intervalles de confiance à 95% construits à partir de l'erreur type de prédiction (traits en pointillé), les points d'échantillonnage (points rouges) et la position du prochain point à évaluer (ligne verticale pointillée, correspondant au maximum du critère).*

L'amélioration au point x est alors une variable aléatoire dont on connait la loi conditionnellement aux n observations déjà réalisées puisque l'on connait la loi conditionnelle de $Y(x)$. On peut alors calculer l'espérance de cette amélioration que l'on appelle amélioration espérée (EI):

$$EI(x) = E\left[(Y(x) - y_{max,n})_+ \,|Y(x_1), \cdots, Y(x_n)\right] \tag{3.6}$$

Jones *et al.* [76] ont publié une expression analytique de l'amélioration espérée:

$$EI(x) = (m_n(x) - y_{max,n}) \times \Phi\left(\frac{m_n(x) - y_{max,n}}{s_n(x)}\right) + s(x) \times \phi\left(\frac{m_n(x) - y_{max,n}}{s_n(x)}\right). \tag{3.7}$$

où ϕ et Φ sont respectivement la densité et la fonction de répartition de la loi normale standard, $m_n(.)$ et $s_n(.)$ donnés par (3.1) et (3.2) représentent respectivement la moyenne et l'écart type de krigeage.
Le nouveau point d'évaluation pour améliorer la qualité du métamodèle (étape 3 de l'algorithme d'optimisation) est choisi de manière à maximiser le critère d'amélioration espérée. Contrairement à la probabilité d'amélioration PI qui favorise une exploitation au profit d'une exploration si on ne règle pas les seuils de façon optimale, le critère d'amélioration espérée fait un compromis satisfaisant entre l'exploitation et l'exploration sans ajout de paramètres. En effet, comme le montre la figure 3.2, l'amélioration espérée peut être élevée soit parce que l'erreur de prédiction est élevée (voir itération 4 de la figure 3.2) soit parce que la prédiction donnée par le krigeage est élevée donc proche de l'optimum courant (voir itération 1 à 3 de la 3.2).
C'est le critère d'amélioration espérée EI qui en pratique a donné de meilleurs résultats et fait davantage l'objet de développement méthodologiques.

3.3.4 Optimisation conditionnelle d'une fonction déterministe

Profil d'amélioration espérée pour l'optimisation conditionnelle

L'objectif ici est de trouver l'optimum d'une fonction conditionnellement à une partie de ses paramètres. Il s'agit de trouver une ligne de crête qui est de la forme

$$L = \{\left(u_0, \text{argmax}_{v\in V} f(u,v) \,|\, u = u_0\right), u_0 \in U\}.$$

On rappelle que l'ensemble des entrées X de la fonction f est alors divisé en deux sous-ensembles: un premier, noté U, contenant les variables conditionnantes et un second, noté V, de variables conditionnées. Il faut un critère permettant de cibler les meilleures entrées $v \in V$ pour chaque valeur de $u \in U$ fixée. Comme le maximum courant à u fixé n'est généralement pas connu, le critère EI qui permet la recherche de points apportant une amélioration par rapport à un maximum courant ne peut convenir. Aussi, Ginsbourger *et al.* [55] ont développé une variante du critère d'amélioration espérée appelée profil d'amélioration espérée (PEI): pour chaque valeur possible de u, le maximum courant, généralement inconnu, est remplacé par un maximum prédit $\max_{w\in V} m(u,w)$. Le critère est défini par:

$$PEI(u,v) = E\left[(Y(u,v) - \min\left(\max_{w\in V} m(u,w), \max_{1\le i\le n} Y(u_i,v_i)\right))_+ \,|Y(u_1,v_1), \cdots, Y(u_n,v_n)\right] \tag{3.8}$$

A noter que le terme $\max_{1\leq i\leq n} Y(u_i, v_i)$ est nécessaire pour la preuve de convergence donnée dans [55].

3.3.5 Optimisation globale d'une fonction bruitée

Lorsque le modèle est bruité, les vraies réponses aux points du plan d'expérience sont inconnues et le critère EI n'est plus valide. En effet

- la valeur du maximum courant $y_{max,n}$ n'est plus la valeur maximale de la fonction $f(x)$ sur le plan d'expérience, puisqu'elles différent maintenant par le bruit;
- les incertitudes de prédictions ne sont pas correctement prises en compte.

Amélioration espérée augmentée (AEI)

Lorsque la variance du bruit est homogène et égale à τ^2, Huang *et al.* [67] proposent de remplacer dans l'expression (3.6) le maximum observé $y_{max,n}$ par le maximum du quantile prédit: $\hat{T} = m_n(x^*)$ où

$$x^* = \text{argmax}_{1\leq i\leq n} \left[m_n(x_i) + \Phi^{-1}(\beta) \times s_n(x_i)\right]$$

A noter que la valeur $\beta = 0.84$ correspondant à $\Phi^{-1}(\beta) = 1$ est recommandée dans [67].
Pour que l'amélioration espérée obtenue soit moins sensible au bruit, elle est multipliée par $\left(1 - \frac{\tau}{\sqrt{s_n^2 + \tau^2}}\right)$. Ce facteur amplifie l'importance de la variance de krigeage, améliorant ainsi l'exploration. Huang *et al.* [67] ont ainsi proposé un nouveau critère d'échantillonnage appelé amélioration espérée augmentée (AEI), défini comme suit:

$$AEI(x) = E\left[(Y(x) - m_n(x^*))_+ |\tilde{Y}(x_1), \cdots, \tilde{Y}(x_n)\right] \times \left(1 - \frac{\tau}{\sqrt{s_n^2 + \tau^2}}\right)$$

Amélioration espérée sur les quantiles de krigeage (EQI)

Un autre critère d'échantillonnage fondé sur l'amélioration du nouveau quantile de krigeage par rapport à l'ancien a été proposé par Picheny *et al.* [106, 107, 105]. Considérant le quantile de krigeage $q_n(x) = m_n(x) + \Phi^{-1}(\beta)s_n(x)$ comme mesure de référence, Picheny *et al.* [106] ont défini l'amélioration comme le dépassement du meilleur β quantile, entre l'étape courante (n) et l'étape suivante $(n+1)$:

$$I_n(x) = [q_{n+1}(x_{n+1}) - \min_{1\leq i\leq n} q_n(x_i)]_+ \tag{3.9}$$

Cette amélioration ne peut pas être connue de façon exacte à l'étape n, car $q_{n+1}(x_{n+1})$ dépend de l'évaluation de f au point x_{n+1}. Cependant, en utilisant le modèle de krigeage, q_{n+1} peut être prédit par Q_{n+1} où Q_{n+1} est la fonction quantile du modèle de

krigeage que l'on attend après la mise à jour avec un nouveau point.
Picheny *et al.* [106] ont alors défini le critère d'amélioration sur les quantiles de krigeage par:

$$I(x) = [Q_{n+1}(x) - \max_{1\leq i\leq n} q_n(x_i)]_+$$

où

$$Q_{n+1}(x) = M_{n+1}(x) + \Phi^{-1}(\beta)S_n(x),$$
$$M_{n+1}(x) = E\left[Y(x)|\tilde{Y}(x_1),\cdots,\tilde{Y}(x_n),\tilde{Y}_{n+1}\right],$$
$$S^2_{n+1}(x) = Var\left[Y(x)|\tilde{Y}(x_1),\cdots,\tilde{Y}(x_n),\tilde{Y}_{n+1}\right].$$

A noter que Q_{n+1} est une fonction aléatoire.

En notant $q_{max,n} = \max_{1\leq i\leq n} q_n(x_i)$, l'amélioration espérée sur les quantiles (EQI) est définie par:

$$EQI_n(x) = E\left[(Q_{n+1}(x) - q_{max,n})_+ |\tilde{Y}(x_1),\cdots,\tilde{Y}(x_n)\right] \tag{3.10}$$

En se basant sur le fait que, conditionnellement à $\tilde{E}_n = \{Y(x_1)+\epsilon_1 = \tilde{y}_1,\cdots,Y(x_n)+\epsilon_n = \tilde{y}_n\}$ le processus $\tilde{Y}_{n+1}$ est gaussien de moyenne et de variance connues, Picheny *et al.* [106] ont montré que conditionnellement aux n observations bruitées, la distribution de Q_{n+1} est gaussienne, ce qui conduit à une formule analytique pour EQI:

$$EQI_n(x) = (m_Q(x) - q_{max,n}) \times \Phi(\frac{m_Q(x) - q_{max,n}}{s_Q(x)}) + s_Q(x) \times \phi(\frac{m_Q(x) - q_{max,n}}{s_Q(x)}) \tag{3.11}$$

où m_Q et s_Q indiquent respectivement la moyenne et l'écart type de Q_{n+1}.
La nouvelle observation $\tilde{y}(x)$ étant une réalisation de la variable aléatoire gaussienne de moyenne $m_{n+1}(x)$ et de variance $s^2_{n+1}(x) + \tau^2$, où τ^2 désigne la variance du bruit au nouveau point x, Picheny *et al.* [106] ont montré que:

$$m_Q(x) = m_n(x) + \Phi^{-1}(\beta)\sqrt{\frac{\tau^2 s^2_n(x)}{\tau^2 + s^2_n(x)}} \;;\; s^2_Q(x) = \frac{[s^2_n(x)]^2}{\tau^2 + s^2_n(x)}.$$

Dans ce chapitre, différentes méthodes d'optimisation classiques ont été présentées ainsi que leurs limites pour la résolution de problèmes d'optimisation complexe ou coûteuse. Ces limites ont justifié le développement de méthodes d'optimisation à l'aide de surfaces de réponses ou métamodèle. Un métamodèle construit à partir de quelques évaluations de la fonction va être utilisé pour définir un critère d'échantillonnage. Ce critère permettra de quantifier l'intérêt d'une nouvelle évaluation et différents critères ont été proposés selon des types de problème posé.
Dans le chapitre suivant, nous allons proposer un critère qui correspond au cadre de l'optimisation conditionnelle d'une fonction bruitée; ce qui correspond à notre problème de recherche de carte d'idéotypes. Comme l'illustre la figure (3.3), de façon naturelle, on va chercher à généraliser les critères EI, EQI et PEI.

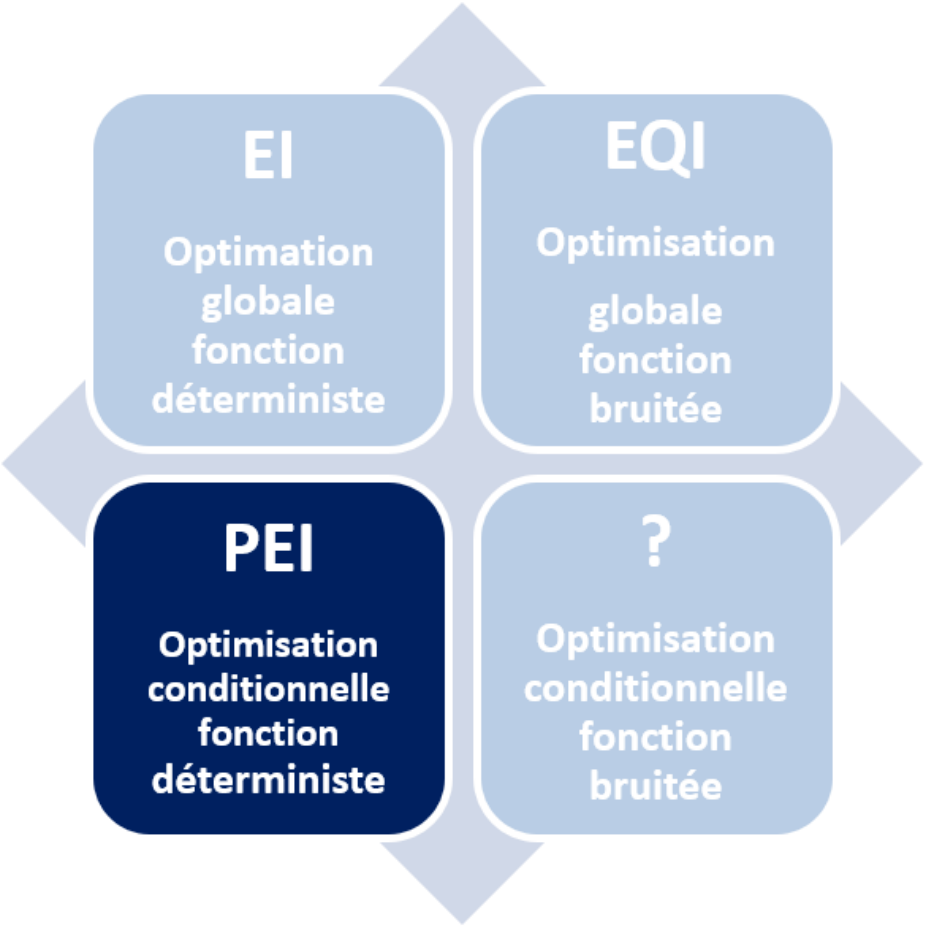


FIGURE 3.3 – *Nouveau critère pour l'optimisation conditionnelle d'une fonction bruitée*

IV

Définition d'un nouveau critère pour l'optimisation conditionnelle d'une fonction bruitée

4.1 Profil d'amélioration espérée sur les quantiles $(PEQI)$

Nous avons vu dans la section 2 que des critères d'échantillonnage ont été développés pour l'optimisation de fonctions déterministes ou bruitées et l'optimisation conditionnelle de fonctions déterministes. Dans cette section, nous présentons un nouveau critère pour l'optimisation conditionnelle d'une fonction bruitée.
Sous les mêmes hypothèses que dans les sections 2.3 et 2.4, c'est à dire:

- d'une part l'ensemble des entrées $\mathbb{E}$ est un ensemble produit dont chaque élément est un couple (u, v) où $u \in U$ est une variable conditionnante et $v \in V$ est une variable conditionnée,
- d'autre part, les mesures de la fonction contiennent un bruit aléatoire que l'on suppose représenté par une suite de variables aléatoires indépendantes, identiquement distribuées,

pour chaque u fixé, nous voulons déterminer le jeu de paramètres v^* qui maximise la fonction f:

$$\max_{v \in V} f(u, v).$$

Le modèle étant bruité dans le sens où une expérience répétée donne des résultats différents, pour chaque paramètre u fixé, nous allons déterminer le jeu de paramètres v^* qui maximise l'espérance de la fonction. On va alors rechercher pour chaque u_0 fixé, l'optimum de $E\left(f(u, v)\right)$ conditionnellement à $u = u_0$. Cela revient à trouver une ligne de crête (L) qui est de la forme :

$$L = \{\big(u_0, \operatorname{argmax}_{v \in V} E\left[f(u, v) \,|\, u = u_0\right], u_0 \in U\big)\}.$$

La méthode de Monte Carlo est utilisée pour estimer l'espérance de la fonction f en un point (u, v). En supposant que cette estimation est déjà faite sur l'ensemble des points d'un plan d'expérience initial $D = \{x_1, \cdots, x_n\} = \{(u_1, v_1), \cdots, (u_n, v_n)\}$ et que f est la réalisation d'un processus gaussien Y, on n'a accès qu'à des mesures de la forme $\tilde{Y}(u_i, v_i) = Y(u_i, v_i) + \epsilon_i$ où ϵ_i est supposé être une réalisation d'une variable aléatoire ϵ et la valeur exacte du maximum courant pour un u fixé n'est pas connue. Notre approche consiste donc à comparer $Q_{n+1}(u, v)$ avec $\max_{w \in V} q_n(u, w)$ pour chaque u fixé. Nous redéfinissons alors l'amélioration sur les quantiles de krigeage (7) sous une forme conditionnelle:

$$I_n(u, v) = [Q_{n+1}(u, v) - \min(\max_{w \in V} q_n(u, w), \max_{1 \leq i \leq n} f(u_i, v_i))]_+$$

Pour chaque u fixé, l'amélioration est définie comme le dépassement d'un seuil égal à la valeur maximale des quantiles observée sur le plan d'expérience.
On rappelle que Q_{n+1} est la fonction quantile du modèle de krigeage que l'on attend après la mise à jour avec un nouveau point, $q_n(u, v) = m_n(u, v) + \Phi^{-1}(\beta) s_n(u, v)$ où $m_n(u, v) = E[Y(u, v) | \tilde{Y}(u_1, v_1), \cdots, \tilde{Y}(u_n, v_n)]$, $s_n^2(u, v) = \mathrm{Var}[Y(u, v) | \tilde{Y}(u_1, v_1), \cdots, \tilde{Y}(u_n, v_n)]$.
Le terme $\max_{1 \leq i \leq n} f(u_i, v_i))$ permet la preuve de convergence que nous verrons plus loin.

L'amélioration ainsi définie est une variable aléatoire dont à chaque étape n on connaît la loi conditionnellement aux observations. On peut alors calculer son espérance que l'on appellera amélioration conditionnelle sur les quantiles $(PEQI)$

$$PEQI_n(u,v) = E\left[I_n(u,v)|\tilde{Y}(u_1,v_1),\cdots,\tilde{Y}(u_n,v_n)\right] \tag{4.1}$$

Conditionnellement aux observations, la distribution du processus Q_{n+1} étant gaussienne [106], en utilisant les propriétés d'un processus gaussien, on peut déduire l'expression analytique de l'amélioration conditionnelle sur les quantiles comme suit:

$$PEQI_n(x) = \left(m_Q(u,v) - y^*_{max,n}\right)\times\Phi\left(\frac{m_Q(u,v) - y^*_{max,n}}{s_Q(u,v)}\right) + s_Q(u,v)\times\phi\left(\frac{m_Q(u,v) - y^*_{max,n}}{s_Q(u,v)}\right) \tag{4.2}$$

où $y^*_{max,n} = \min(\max_{w\in V} q_n(u,w), \max_{1\leq i\leq n} f(u_i,v_i))$ est le meilleur quantile courant, m_Q et s_Q indiquent respectivement la moyenne et l'écart type de Q_{n+1}.
D'après l'écriture du critère, il est à noter que le critère $PEQI$ est une généralisation des critères PEI, EQI et EI. En effet, il vérifie les propriétés suivantes:

- En absence de bruit, le critère $PEQI$ est équivalent au critère PEI.
- Si on s'intéresse à une seule valeur de u, le critère $PEQI$ est égal au critère EQI.
- Si on s'intéresse à une seule valeur de u et que le bruit est nul, alors $PEQI$ se résume à EI.

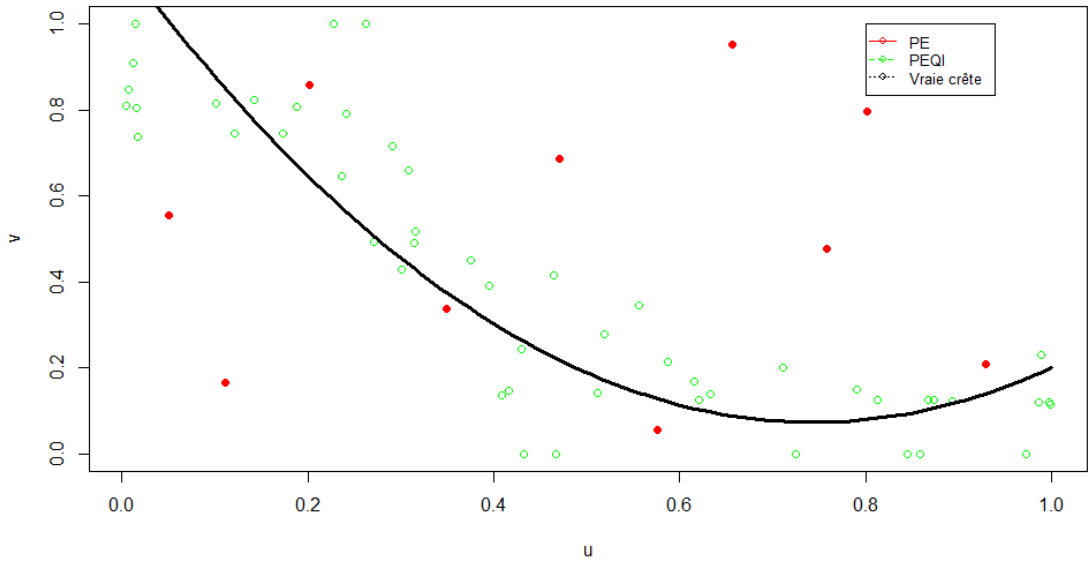


FIGURE 4.1 – *Échantillonnage de points avec le critère* $PEQI$

Le critère prend en compte la stochasticité dans le modèle puisqu'il permet l'échantillonnage des points qui améliore un seuil égal au meilleur quantile observé. Il permet également un échantillonnage des meilleurs points pour chaque valeur de u fixé. Notre

objectif est d'optimiser la fonction f conditionnellement aux valeurs de u. On n'a donc pas intérêt à utiliser un critère qui permet de couvrir l'ensemble de l'espace de recherche. Ce qu'il nous faut, c'est un critère qui permet une recherche dans des endroits spécifiques, ce qui va réduire le coût de calcul. Comme le montre la figure 4.1 les points ajoutés (points verts) sont pour la plupart autour de la vraie ligne de crête (courbe noire). Nous aurons ainsi une bonne estimation de la vraie ligne de crête sans explorer l'ensemble de l'espace.

4.2 Algorithme d'optimisation pour la recherche d'une ligne de crête

Maximiser le critère $PEQI$ sur l'ensemble $U \times V$ suppose pour chaque $u \in U$ fixé deux maximisations. Tout d'abord celle de $q_n(u, w)$ sur $w \in V$ et ensuite celle de $PEQI(u, v)$ pour $v \in V$. De même que pour le critère EI, chacune de ces maximisations se fera à l'aide d'une méthode d'optimisation globale, car les fonctions $q_n(u, .)$ et $PEQI(u, .)$ présentent souvent de multiples maxima locaux. Un algorithme génétique sera utilisé pour déterminer w^* et v^* de manière la plus précise possible. Cette démarche, qui doit être faite pour chaque $u \in U$ afin de déterminer l'optimum de $PEQI$ sur l'ensemble $U \times V$, peut être très consommatrice de ressources. Or l'imprécision de l'optimum en u n'est pas gênante car on n'est pas intéressé par la recherche de l'optimum global du modèle f sur l'ensemble $U \times V$ mais plutôt par la recherche de l'optimum du modèle f sur V pour chaque u fixé.
Nous allons alors adopter la stratégie de Ginsbourger *et al.* [55]: à chaque itération, on choisit une grille G de points u et on optimise le critère sur l'ensemble $G \times V$.
Posons $PEQI_n^*(u) = \max_{v \in V} PEQI_n(u, v)$.
Soit $u^* = \text{argmax}_{u \in G} PEQI_n^*(u)$, le prochain point à évaluer sera le point

$$x^* = (u^*, \text{argmax}_{w \in V} PEQI_n(u^*, w))$$

Pour un u fixé, $PEQI_n^*(u)$ est calculé en optimisant le critère $PEQI_n$ sur l'ensemble V à l'aide d'un algorithme génétique.

Nous définissons l'algorithme d'optimisation comme suit:

1. Choix d'un plan initial $D = \{(u_1, v_1), \cdots, (u_n, v_n)\}$.
2. Construction du métamodèle de krigeage à partir du plan d'expérience.
3. Sélection d'une grille G de points u.
4. Pour chaque point u de la grille, évaluer $\max_{w \in V} q_n(u, w)$ à l'aide d'un algorithme génétique.
5. Pour chaque point u de la grille, trouver le point v_j^* qui maximise $PEQI(u, v)$.
6. Recherche du point (u^*, v^*) solution de $\max_{u \in G} \max_{v \in V} PEQI_n(u, v)$.
7. Calcul du modèle bruité en ce nouveau point.

8. Ajout du point au plan initial et mise à jour du métamodèle.

Les étapes de 3 à 8 sont répétées N fois. Le nombre N d'itérations est fixé à l'avance en fonction du budget de simulations

4.3 Preuve de convergence

Dans cette partie, nous traitons la convergence de l'algorithme d'optimisation défini précédemment en nous inspirant de [140] où la convergence de l'algorithme d'amélioration espérée est démontrée sous l'hypothèse que la fonction déterministe f à optimiser est la réalisation d'un processus gaussien Y de moyenne nulle et de covariance connue.
La démonstration est faite sous une hypothèse particulière vérifiée par la fonction de covariance, à savoir la propriété de no-empty-ball (NEB).

Définition (propriété de no-empty-ball NEB [140])

Nous dirons que le processus gaussien Y ou, de façon équivalente, la fonction de covariance k possède la propriété NEB si, pour toutes suites $(x_n)_{n\geq 1}$ dans $\mathbb{E}$ et tout $y \in X$, les assertions suivantes sont équivalentes :

(i) y est un point adhérent de l'ensemble $\{x_n, n \geq 1\}$.

(ii) $s_n^2(y,(x_1,\cdots,x_n)) \to 0$ quand $n \to +\infty$.

Il est démontré dans [140] que la fonction de covariance Matérn vérifie la propriété de no-empty-ball.
Puisque la fonction de covariance k est supposée continue, (i) implique toujours (ii) dans la définition. La propriété NEB est donc équivalente à l'affirmation selon laquelle, si l'erreur de prédiction en y tend vers zéro, toute boule centrée en y possède un point dans $\{x_n, n \geq 1\}$ (ie, pour tout $\epsilon > 0$, il existe au moins un $n \geq 1$ tel que $|y - x_n| < \epsilon$).
Nous allons montrer que la suite de points générée par notre algorithme d'optimisation est dense dans le domaine de recherche donc converge. Ce qui se traduit par le théorème suivant:

Théorème 1

Soit H un espace de Hilbert à noyau reproduisant k. Supposons que la fonction de covariance k vérifie la propriété de no-empty-ball (NEB) définie ci-dessus. Alors,

pour tout $x_{init} \in X$ et tout $f \in H$ la suite $(X_n(f))_{n\geq 1}$ générée par l'algorithme

$$\begin{cases} X_1 = x_{init} \in \mathbb{E} \\ X_{n+1} = \text{argmax}_{x\in X} PEQI_n(x) \end{cases}$$

est dense dans $\mathbb{E}$.

Notations

Pour tout $n \geq 1$, posons $\underline{x}_n = (x_1, \cdots, x_n) = ((u_1, v_1), \cdots, (u_n, v_n)) \in X^n$, et pour $x \in X$, désignons par $\hat{\xi}_{n+1}(x, f, \underline{x}_n)$ l'espérance conditionnelle de $Q_{n+1}(x)$ sachant $\tilde{Y}(x_1), \cdots \tilde{Y}(x_n)$, par $s^2_{Q_{n+1}}(x, \underline{x}_n, f)$ la variance du quantile de krigeage, par $s^2_n(x, \underline{x}_n, f)$ la variance de prédiction de krigeage et par $PEQI_n(x)$ le profil d'amélioration espérée sur les quantiles au point x.
Pour la suite, sauf si c'est nécessaire, nous noterons $\hat{\xi}_{n+1}(x)$ l'espérance conditionnelle de $Q_{n+1}(x)$ sachant $\tilde{Y}(x_1), \cdots \tilde{Y}(x_n)$, $s^2_{Q_{n+1}}(x)$ la variance du quantile krigeage et $s^2_n(x)$ la variance de krigeage.

Comme dans [140], nous allons considérer une généralisation du critère de profil d'amélioration espérée sur les quantiles définie par :

$$PEQI_n(x) = E\left[\left(Q_{n+1}(x) - y^*_{max,n}(u)\right)_+ |\tilde{Y}(x_1), \cdots, \tilde{Y}(x_n)\right], \tag{4.3}$$

avec

$$y_{max,n} = \max_{1\leq i\leq n} f(x_i)$$

et

$$y^*_{max,n} = \min\left\{\max_{w\in V} q_n(x), y_{max,n}\right\}.$$

Il est montré dans Picheny *et al.* [106] que conditionnellement aux observations, Q_{n+1} est un processus gaussien. Nous pouvons donc en déduire une expression analytique de (4.3).

Désignons par Φ la fonction de répartition de la loi normale centrée réduite, alors (4.3) vaut:

$$\begin{aligned} PEQI_n(x) &= \left(\hat{\xi}_{n+1}(x) - y^*_{max,n}\right) \times \Phi\left(\frac{\hat{\xi}_{n+1}(x) - y^*_{max,n}}{s_{Q_{n+1}}(x)}\right) && (4.4) \\ &\quad + s_{Q_{n+1}}(x) \times \phi\left(\frac{\hat{\xi}_{n+1}(x) - y^*_{max,n}}{s_{Q_{n+1}}(x)}\right) && (4.5) \\ &=: \psi(\hat{\xi}_{n+1}(x) - y^*_{max,n}, s_{Q_{n+1}}(x)) && (4.6) \end{aligned}$$

où

$$\psi : \mathbb{R} \times [0, +\infty) \longrightarrow [0, +\infty)$$

est définie par

$$\psi(z,s) = \begin{cases} \sqrt{s}\phi\left(\frac{z}{\sqrt{s}}\right) + z\Phi\left(\frac{z}{\sqrt{s}}\right) & \text{si } s > 0 \\ \max(z,0) & \text{si } s = 0 \end{cases} \tag{4.7}$$

et satisfait aux exigences suivantes:

R_1 : ψ est continue;

R_2 : $\forall z \leq 0, \psi(z,0) = 0$;

R_3 : $\forall z \in R, \forall s > 0, \psi(z,s) > 0$.

Nous allons adapter la proposition 10 [140] à notre critère d'échantillonnage.

Proposition 1

Soient $(x_n)_{n\geq 1}$ et $(y_n)_{n\geq 1}$ deux suites dans $\mathbb{E}$. Supposons que la suite $(y_n)_{n\geq 1}$ converge vers y^* alors chacune des conditions ci-dessous implique la suivante :

(i) y^* est un point adhérent de l'ensemble $\{x_n, n \geq 1\}$

(ii) $s_n^2(y_n, \underline{x}_n) \to 0$ quand $n \to +\infty$

(iii) $\hat{\xi}_{n+1}(y_n, f, \underline{x}_n) \to f(y^*)$ quand $n \to +\infty$ pour tout $f \in H$

Preuve de la proposition 1

Montrons que $(i) \Rightarrow (ii)$
Supposons que $y^* \notin \{x_n, n \geq 1\}$ (autrement le résultat est trivial). Soit (x_{φ_n}) une sous suite de la suite $\{x_n, n \geq 1\}$ qui converge vers y^* et soit $\psi_n = \max_k \{\varphi_k; \varphi_k \leq n\}$ alors $s_n^2(y_n, \underline{x}_n) = Var\left[Y(y_n) - \hat{Y}_n(y_n, \underline{x}_n)\right] \leq Var\left[Y(y_n) - Y(x_{\psi_n}),\right]$.
Quand $\psi_n \longrightarrow \infty$, il résulte de la continuité de k que :

$$Var[Y(y_n) - Y(x_{\psi_n})] = k(y_n, y_n) + k(x_{\psi_n}, x_{\psi_n}) - 2k(x_{\psi_n}, y_n) \longrightarrow 0$$

Montrons que $(ii) \Rightarrow (iii)$
Nous avons:

$$|f(y^*) - \hat{\xi}_{n+1}(y_n, f, \underline{x}_n)| \leq |f(y^*) - f(y_n)| + |f(y_n) - \hat{\xi}_{n+1}(y_n, f, \underline{x}_n)|$$

Or

$$\hat{\xi}_{n+1}(y_n, f, \underline{x}_n) = m_n(y_n) + \Phi^{-1}(\beta) \times \sqrt{\frac{\tau^2 \times s_n^2(y_n, \underline{x}_n)}{\tau^2 + s_n^2(y_n, \underline{x}_n)}}$$

τ^2 étant la variance du bruit au point y_n.

$$\begin{aligned} |f(y_n) - \hat{\xi}_{n+1}(y_n, f, \underline{x}_n)| &= |< k(y_n, .), f > - < \sum_{i=0}^{n} \lambda_{n,i}(y_n) \times k(x_i, .), f > \\ &\quad -\Phi^{-1}(\beta)\sqrt{\frac{\tau^2 \times s_n^2(y_n, \underline{x}_n)}{\tau^2 + s_n^2(y_n, \underline{x}_n)}}| \\ &\leq |< k(y_n, .) - \sum_{i=0}^{n} \lambda_{n,i}(y_n) \times k(x_i, .), f > | \\ &\quad +|\Phi^{-1}(\beta) \sqrt{\frac{\tau^2 \times s_n^2(y_n, \underline{x}_n)}{\tau^2 + s_n^2(y_n, \underline{x}_n)}}| \end{aligned}$$

En utilisant l'inégalité de Cauchy-Schwarz dans H, nous avons :

$$\begin{aligned} |f(y_n) - \hat{\xi}_{n+1}(y_n, f, \underline{x}_n)| &\leq ||k(y_n, .) - \sum_{i=0}^{n} \lambda_{n,i}(y_n) \times k(x_i, .)\|_H \times \|f\|_H \\ &\quad +|\Phi^{-1}(\beta)\sqrt{\frac{\tau^2 \times s_n^2(y_n, \underline{x}_n)}{\tau^2 + s_n^2(y_n, \underline{x}_n)}}| \end{aligned}$$

D'autre part, on a:

$$\begin{aligned} &\|k(y_n, .) - \textstyle\sum_{i=0}^{n} \lambda_{n,i}(y_n) \times k(x_i, .)\|_H \\ &\quad = \sqrt{< k(y_n, .) - \textstyle\sum_{i=0}^{n} \lambda_{n,i}(y_n) \times k(x_i, .), k(y_n, .) - \textstyle\sum_{i=0}^{n} \lambda_{n,i}(y_n) \times k(x_i, .) >} \\ &\quad = \sqrt{(k(y_n, y_n) - \textstyle\sum_{i=0}^{n} \lambda_{n,i}(x) \times k(x_i, y_n))} \\ &\quad = \sqrt{s_n^2(y_n, \underline{x}_n))} \end{aligned}$$

Alors:

$$|f(y^*) - \hat{\xi}_{n+1}(y_n, f, \underline{x}_n)| \leq |f(y^*) - f(y_n)| + s_n(y_n, \underline{x}_n)\|f\|_H + |\Phi^{-1}(\beta)\sqrt{\frac{\tau^2 \times s_n^2(y_n, \underline{x}_n)}{\tau^2 + s_n^2(y_n, \underline{x}_n)}}| \to 0$$

quand $n \to +\infty$.

Soit $\nu_n = \sup_{x \in X} PEQI_n(x)$ où $PEQI_n$ est le critère défini en (4.3). Rappelons que, pour tout $n \geq 1, \nu_n = PEQI_n(X_{n+1}) = \psi\left(\hat{\xi}_{n+1}(X_{n+1}) - y^*_{max,n}, s^2_{Q_{n+1}}(X_{n+1})\right)$ avec

$$y^*_{max,n}(x) = y^*_{max,n}(u, v) \equiv y^*_{max,n}(u) = \min\left\{\max_{w \in V} q_n(u, w), \max_{1 \leq i \leq n} f(u_i, v_i)\right\}$$

La démonstration du théorème 1 est basée sur le résultat suivant :

Lemme 1

Pour tout $f \in H, \liminf_{n \longrightarrow +\infty} \nu_n(f) = 0$.

Preuve du lemme 1

Fixons $f \in H$. Pour tout $n \geq 1$, posons $x_n = X_n(f), \sigma_n = s^2_{Q_{n+1}}(x_{n+1}, f)$ et $z_n = \hat{\xi}_{n+1}(x_{n+1}, f) - y^*_{max,n}(x_{n+1})$; on a alors $\nu_n(f) = \psi(z_n, \sigma_n)$.

Montrons qu'il existe une sous-sous suite $\nu_{\varphi_{\phi_n}}$ de la suite ν_n telle que:
$\nu_{\varphi_{\phi_n}-1} = \psi\left(z_{\varphi_{\phi_n}-1}, v_{\varphi_{\phi_n}-1}\right) \to 0$ quand $n \to +\infty$.
Soit y^* un point adhérent pour la suite $\{x_n, n \geq 1\}$ ce qui implique qu'il existe une sous suite (x_{φ_n}) de la suite (x_n) qui converge vers y^*.
D'après la proposition 1, $(i) \Longrightarrow (iii)$ donc $\hat{\xi}_{\varphi_n}(x_{\varphi_n}, f) \to f(y^*)$.

Nous allons maintenant montrer que la suite $\left(y^*_{max,\varphi_n-1}(x_{\varphi_n})\right)$ est bornée. Il apparait clairement que $\left(y^*_{max,\varphi_n-1}(x_{\varphi_n})\right)$ admet y_{max,φ_n-1} comme borne supérieure car:

$$y^*_{max,\varphi_n-1}(x_{\varphi_n}) = \min\left\{\max_{w \in V} \hat{\xi}_{\varphi_n-1}(u_{\varphi_n}, w), y_{max,\varphi_n-1}\right\}$$

D'autre part y_{max,φ_n-1} est une suite croissante et bornée (et donc converge vers une limite inférieure bornée par $f(y^*)$), avec la propriété:

$$y_{max,\varphi_n-1} \geq y_{max,\varphi_{n-1}} \geq f\left(x_{\varphi_{n-1}}\right) \to f(y^*).$$

Il suffit alors de prouver que $\left(\max_{w \in V} \hat{\xi}_{\varphi_n-1}(u_{\varphi_n}, w)\right)$ admet une borne inférieure, et ce dernier résultat découle du fait que:

$$\max_{w \in V} \hat{\xi}_{\varphi_n-1}(u_{\varphi_n}, w) \geq \hat{\xi}_{\varphi_{n-1}}(x_{\varphi_n}) \to f(y^*)$$

Nous avons la suite $\left(y^*_{max,\varphi_n-1}(x_{\varphi_n})\right)$ qui est bornée, nous pouvons donc en extraire une sous suite $\left(y^*_{max,\varphi_{\phi_n}-1}(x_{\varphi_{\phi_n}})\right)$ qui converge et vérifie la propriété

$$y^*_{max,\varphi_{\phi_n}-1}(x_{\varphi_{\phi_n}}) = \min\left\{\max_{w \in V} \hat{\xi}_{\varphi_{\phi_n}-1}(u_{\varphi_{\phi_n}}, w), y_{max,\varphi_{\phi_n}-1}\right\} \geq$$

$$\min\left(f(x_{\varphi_{\phi_n}}), \hat{\xi}_{\varphi_{\phi_n}-1}(x_{\varphi_{\phi_n}})\right) \to f(y^*)$$

Donc

$$\lim_{n\to\infty} z_{\varphi_{\phi_n}-1} = \lim_{n\to\infty} \hat{\xi}_{\varphi_{\phi_n}}(x_{\varphi_{\phi_n}}) - \lim_{n\to\infty} y^*_{max,\varphi_{\phi_n}-1}(x_{\varphi_n}) \leq 0$$

Nous avons aussi d'après la proposition 1 $(i) \Rightarrow (ii)$ donc $s^2_{\varphi_{\phi_n}} \to 0$.
Picheny *et al.* [106] ont montré que: $s^2_{Q_{\varphi_{\phi_n}}}\left(x_{\varphi_{\phi_n}}, f\right) = \dfrac{[s^2_{\varphi_{\phi_n}}(x_{\varphi_{\phi_n}}, f)]^2}{\tau^2 + s^2_{\varphi_{\phi_n}}(x_{\varphi_{\phi_n}}, f)}$ où τ^2 désigne la variance du bruit au nouveau point $x_{\varphi_{\phi_n}}$, donc $\sigma_{\varphi_{\phi_n}-1} = s^2_{Q_{\varphi_{\phi_n}}}\left(x_{\varphi_{\phi_n}}, f\right) \to 0$

D'après R_1: ψ est continue et R_2: $\forall z \leq 0, \psi(z, 0) = 0$, on en déduit:

$$\nu_{\varphi_{\phi_n}-1}(f) = \psi\left(z_{\varphi_{\phi_n}-1}, \sigma_{\varphi_{\phi_n}-1}\right) \to \psi\left(\lim_{n\to+\infty} z_{\varphi_{\phi_n}-1}, 0\right) = 0$$

Preuve du théorème 1

Supposons que $\{x_n, n \geq 1\}$ n'est pas dense dans $\mathbb{E}$, alors il va exister un point $y^* \in X$ qui n'est pas adhérent à $\left\{(x_n)_{n\geq 1}\right\}$.
Ce qui implique d'après la propriété NEB que $\inf_{n\geq 1} s^2_n(y^*, f) > 0$, d'où $\inf_{n\geq 1} s^2_{Q_{n+1}}(y^*, f) > 0$.
D'autre part en utilisant l'inégalité de Cauchy-Schwarz dans H, on observe que la suite $\left(\hat{\xi}_{n+1}(y^*, f)\right)$ est bornée et nous avons :

$$\|\hat{\xi}_{n+1}(y^*, f) - f(y^*)\|^2 \leq s^2_n(y^*, f)\,\|f\|^2_H \leq k(y^*, y^*)\|f\|^2_H$$

La suite $(y_{max,n})$ est évidement bornée par $\|f\|_\infty$ et $y^*_{max,\varphi_n-1} \leq y_{max,n}$.
En utilisant le fait que pour toute valeur positive fixée de la deuxième variable, la fonction ψ définie en (4.7) est une fonction croissante de la première variable, nous avons:

$$\psi\left(\hat{\xi}_{n+1}(y^*, f) - y^*_{max,n}(y^*), s^2_{Q_{n+1}}(y^*, f)\right) \geq \psi\left(\hat{\xi}_{n+1}(y^*, f) - y_{max,n}, s^2_{Q_{n+1}}(y^*, f)\right)$$

$\hat{\xi}_{n+1}(y^*, f) - y_{max,n}$ étant un nombre réel et $\inf_{n\geq 1} s^2_{Q_{n+1}}(y^*, f) > 0$, on obtient en conséquence de R_1: ψ est continue et R_3 : $\forall z \in R$, $\forall s > 0$ $\psi(z, s) > 0$ que :

$$PEQI_n(y^*, f) \geq \inf_{n\geq 1} \psi\left(\hat{\xi}_{n+1}(y^*, f) - y^*_{max,n}(y^*), s^2_{Q_{n+1}}(y^*, f)\right) \geq$$

$$\inf_{n\geq 1} \psi\left(\hat{\xi}_{n+1}(y^*, f) - y_{max,n}, s^2_{Q_{n+1}}(y^*, f)\right) > 0$$

Ce qui est en contradiction avec le résultat du lemme 1 $\nu_n = \sup_{x\in X} PEQI_n(x)$.

4.4 Application sur des fonctions tests et comparaison de $PEQI$ avec PEI

Pour illustrer la bonne marche de notre algorithme d'optimisation, nous allons l'appliquer sur des fonctions tests classiquement utilisées en optimisation. Il s'agit des

fonctions Branin2 [14] et Rosenbrock [119] dont les expressions analytiques sont données dans le tableau 4.1

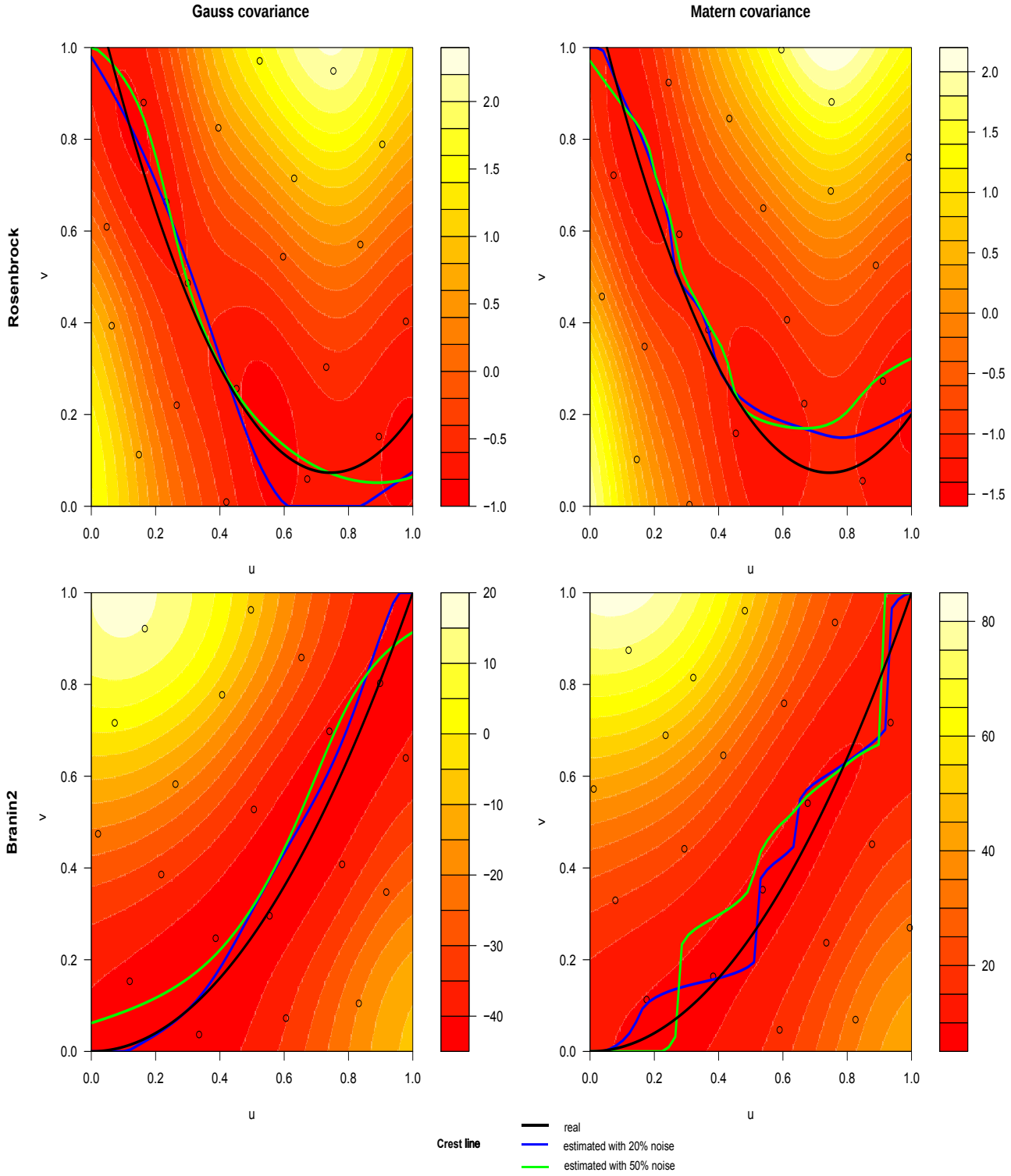


FIGURE 4.2 – *Lignes de crête réelles et prédites en utilisant le critère $PEQI$ avec $\beta = 0.1$. Des niveaux de bruits de 20 et 50% de la variance des réponses de la fonction test sont testés.*

Les fonctions présentent une ligne de crête dans l'espace $[0; 1]^2$. Sur cet espace, l'amplitude de variation de ces fonctions est mesurée à l'aide d'une variance empirique

Nom	Expression analytique
Branin2	$f(x) = \frac{1}{51.95}\left[\left(b - \frac{5.1a^2}{4\pi^2} + \frac{5a}{\pi} - 6\right)^2 + \left(10 - \frac{10}{8\pi}\right)\cos(a) - 44.81\right]$ où $a = 15 \times x_1 - 5$, $b = 15 \times x_2$
Rosenbrock2	$f(x) = 100\left(x_2 - x_1^2\right)^2 + \left(1 - x_1\right)^2$

TABLE 4.1 – *Fonctions tests*

obtenue après un tirage de 500 points (u, v) suivant une loi uniforme sur $[0; 1]^2$. Les fonctions étant déterministes, les variances des bruits aléatoires gaussiens iid ajoutés à ces fonctions sont exprimées en pourcentage de la variance empirique de la fonction à optimiser.
Les plans initiaux sont obtenus en générant des hypercubes latins à l'aide du package LHS [20] du logiciel $\mathbb{R}$. Leur nombre n de points satisfait la préconisation $n = 10 \times d$ [76], d étant la dimension de la fonction à optimiser. Des noyaux de covariance anisotropes Matérn $\nu = 3/2$ et gaussien sont utilisés et leurs paramètres sont estimés par maximum de vraisemblance en utilisant le package DiceKriging [120].

Dans la figure 4.2, nous avons au total 70 points d'évaluation, composés des 20 points du plan d'expérience initial et des 50 nouveaux points obtenus en maximisant séquentiellement le critère $PEQI$ avec $\beta = 0.1$. On peut également voir sur la figure 4.2 la prédiction finale de krigeage avec une variance du bruit égal à 20 % du signal ainsi que l'estimation de la ligne de crête obtenue avec un niveau de bruit égale à 20% et 50% de variance de la fonction.
Pour les deux fonctions test et les deux noyaux de covariance (Gauss et Matérn), on peut noter qu'après 50 itérations, nous avons une assez bonne estimation de la ligne de crête, même avec un niveau de bruit élevé. La meilleure estimation est obtenue avec le noyau gaussien, qui donne une ligne plus régulière.

4.5 Comparaison avec le profil d'amélioration espérée selon différents facteurs expérimentaux et algorithmiques

Avec les mêmes fonctions test, le nouveau critère d'échantillonnage proposé $PEQI$ est comparé avec le profil d'amélioration espérée PEI développé par Ginsbourger *et al.* [55].

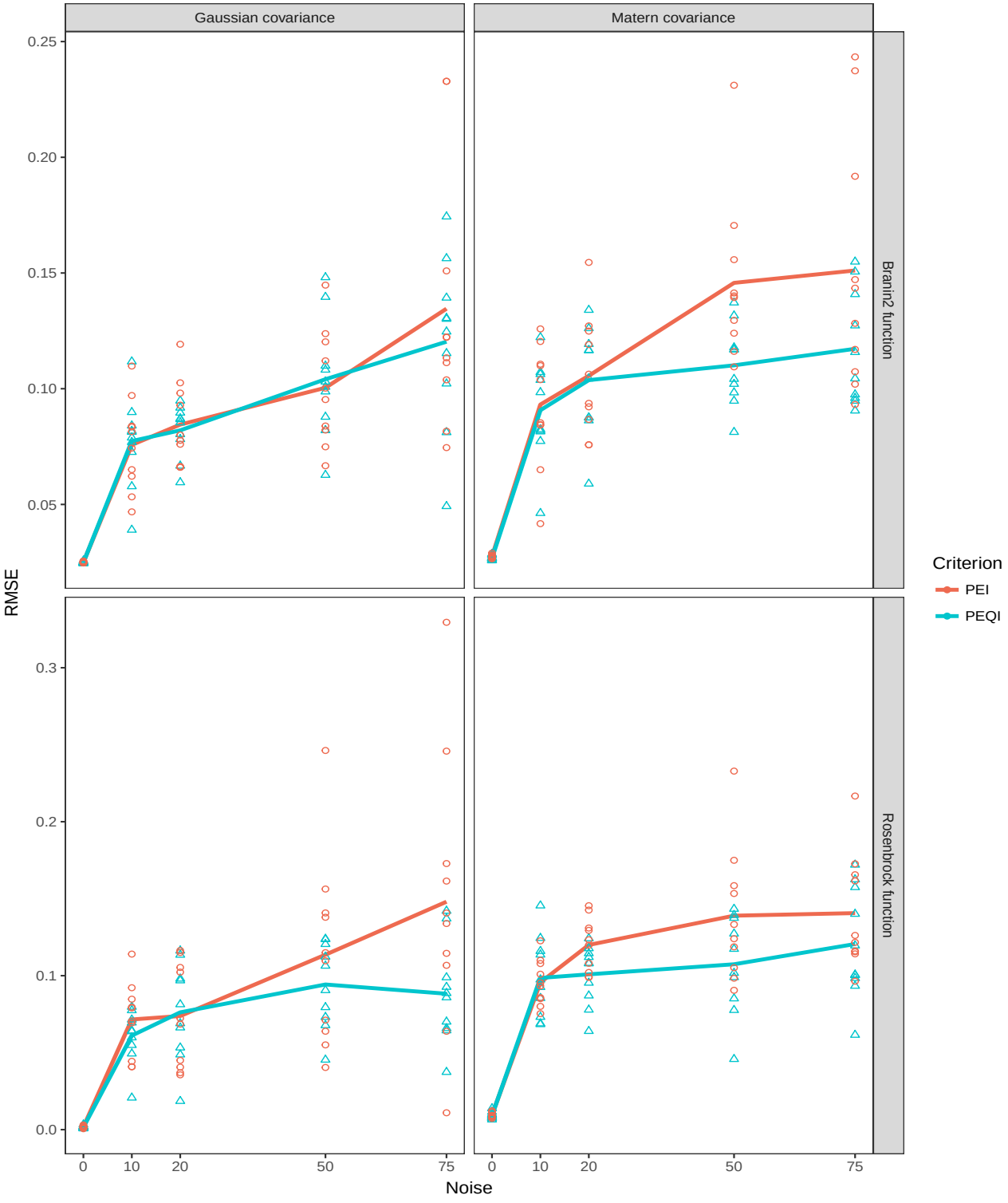


FIGURE 4.3 – *PEI vs PEQI. RMSE de la ligne de crête en fonction du niveau de bruit et du critère d'échantillonnage. Pour le critère $PEQI$, β est fixé à 0.1.*

Les deux critères sont comparés selon l'écart entre la ligne de crête de la fonction connue f (Rosenbrock ou Branin2) et la ligne de crête de la surface de krigeage g. Cette différence correspond à la $RMSE$ entre les échantillons de points des deux lignes de crête. Elle a été calculée en sélectionnant une grille régulière de 50 points $u1, \cdots u50$, et pour chaque point:

(i) recherche du maximum donné par la surface de krigeage g

(ii) évaluation du maximum de la fonction analytique f.

L'erreur quadratique moyenne $RMSE$ entre la prédiction et la ligne de crête réelle est la suivante:

$$RMSE = \sqrt{\frac{1}{50}\sum_{i=1}^{50}\big(\text{argmax}_{v\in V}g(u_i,v) - \text{argmax}_{v\in V}f(u_i,v)\big)}.$$

Les valeurs de $RMSE$ selon différents niveaux de bruit sont utilisées pour comparer PEI avec $PEQI$, en fixant $\beta = 0.1$ pour $PEQI$ (Figure 4.3).
La $RMSE$ est également utilisée pour comparer les effets des différents paramètres β pour $PEQI$ (figure 4.4) et pour comparer la sensibilité des performances de PEI et de $PEQI$ à la taille de la conception initiale (figures 4.5 et 4.6). Ces comparaisons ont été faites en utilisant deux fonctions test (Branin2 et Rosenbrock) et deux noyaux de covariance (Matérn et Gauss) pour effectuer le krigeage.
La plage de variation de la variance du bruit est de 0 à 75% de la variance empirique de la fonction à optimiser.
Dans un premier temps, des plans de $10 \times d$ points initiaux sur un budget total de 70 points sont retenus pour comparer les réponses obtenues avec PEI et $PEQI$ (figure 4.3), puis pour étudier l'influence du $\beta - quantile$ de $PEQI$ (figure 4.4). Dans un deuxième temps, les résultats obtenus avec ces mêmes plans de $10 \times d$ points initiaux sur un budget de 70 points sont comparés avec ceux obtenus par des plans de $20 \times d$ points initiaux sur un budget de 90 points (figures 4.5 et 4.6).
Les mêmes 10 hypercubes latins sont utilisés pour répéter 10 fois chacune des combinaisons de fonction à optimiser, critère d'échantillonnage, fonction de covariance et niveau de bruits.

A budget de simulation fixé, les performances respectives des deux critères PEI et $PEQI$ évoluent selon le niveau de bruit considéré. Comme le montre la figure 4.3, lorsque les deux critères ne donnent pas des résultats comparables, $PEQI$ surpasse toujours PEI. En l'absence de bruit, les critères sont similaires pour les deux fonctions test et les deux fonctions de covariance. Lorsque le bruit augmente, $PEQI$ devient meilleur que PEI à partir de niveaux de bruit qui diffèrent selon les fonctions test et les fonctions de covariance. Pour une fonction test et une covariance données, on peut remarquer que les performances de $PEQI$, contrairement à celles de PEI, restent relativement stables lorsque le bruit augmente. Ces différents résultats peuvent s'expliquer par le fait qu'avec PEI les nouveaux points à évaluer sont choisis sans tenir compte de leur niveau de bruit alors qu'avec $PEQI$ les nouveaux points sont sélectionnés en comparant les β-quantiles de prédiction du krigeage, donc intègrent leur niveau de bruit.
Tout comme EQI [104], les valeurs du critère $PEQI$ dépendent du choix fait pour β. Selon la valeur de β retenue, l'échantillonnage peut être plus ou moins guidé vers une exploration ou au contraire vers une exploitation. L'effet du choix de la valeur β est illustré figure 4.4. Trois valeurs distinctes sont considérées, allant d'une forte exploration à une forte exploitation $\beta \in \{0.1; 0.5; 0.9\}$.

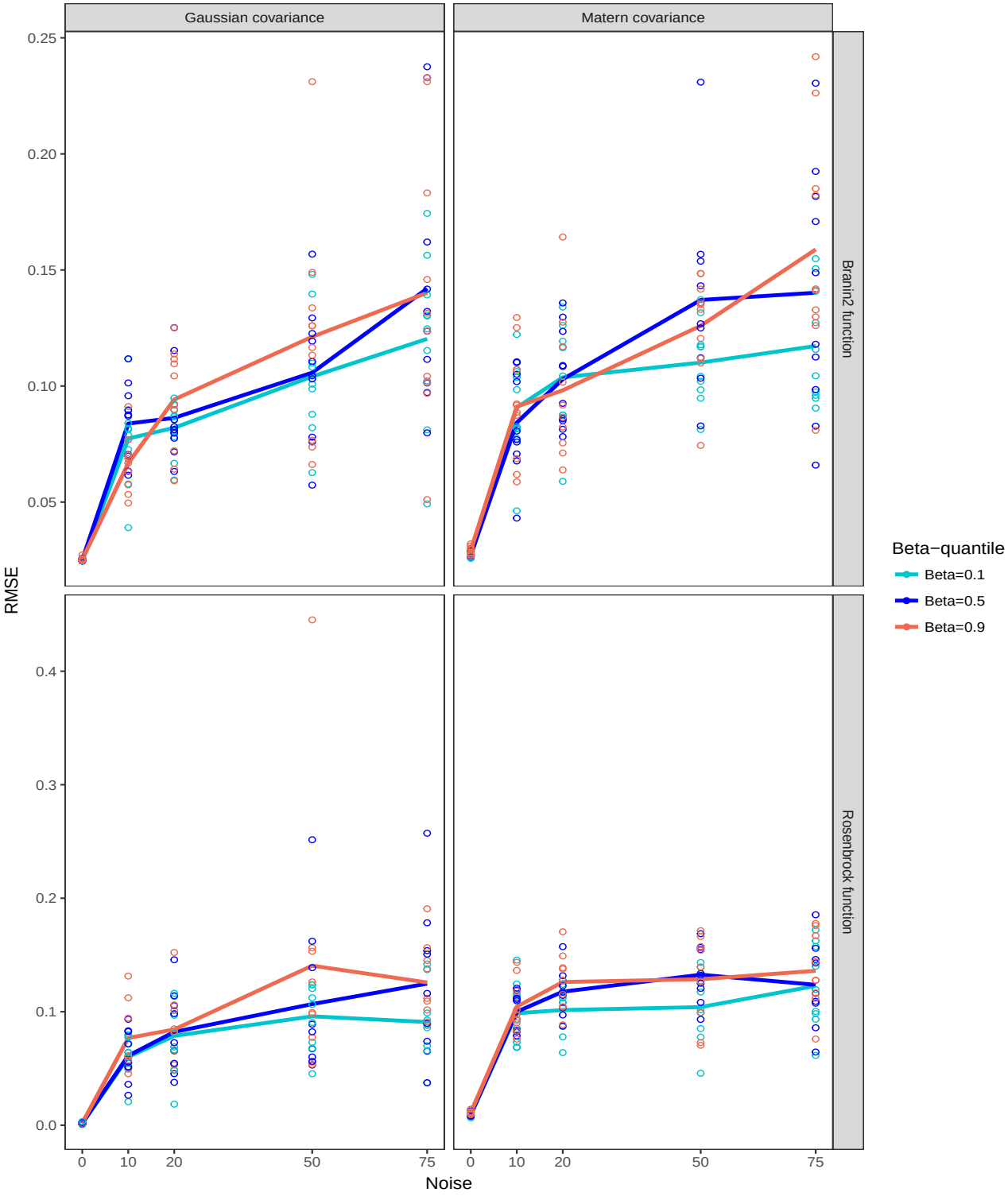


FIGURE 4.4 – *Influence de β sur $PEQI$. RMSE de la ligne de crête en fonction du niveau de bruit et de la taille du plan initial. Pour des plans initiaux de 20 et 40 points, 50 points sont ajoutés séquentiellement à l'aide du critère PEI*

Comme attendu, le critère $PEQI$ présente une sensibilité au choix de la valeur β. Le choix d'un β faible conduit à de meilleurs résultats pour des niveaux de bruits élevés mais avec un niveau faible de bruit, peu de différences apparaissent.
Les figures 4.5 et 4.6 illustrent l'influence de la taille du plan d'expérience initial sur

les performances respectives des critères PEI et $PEQI$. Deux tailles de plan initial sont comparées : 20 et 40 points, auxquels 50 points sont rajoutés séquentiellement. Le critère PEI apparaît beaucoup plus sensible que $PEQI$ à la taille initiale du plan d'expérience : $PEQI$ conduit à des résultats peu influencés par une division par deux de la taille du plan initial, et ceci quelque soit le niveau de bruit considéré.
Finalement, l'impact du noyau de covariance sur l'efficacité des deux méthodes d'optimisation est étudié. Les résultats présentés suggèrent que les noyaux influent essentiellement dans le cas de niveaux de bruit élevés.

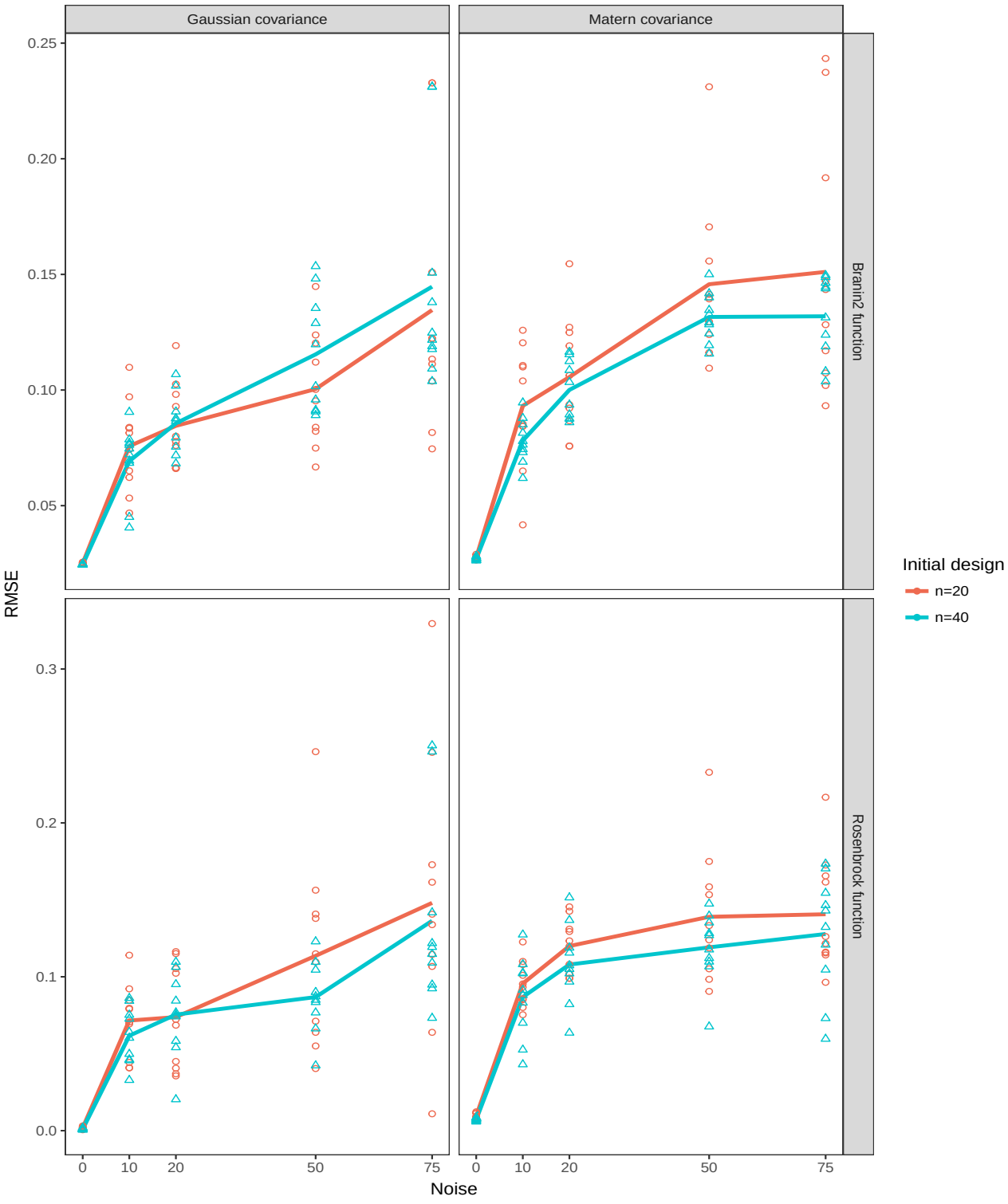


FIGURE 4.5 – *Influence de la taille du plan initial sur PEI. RMSE de la ligne de crête en fonction du niveau de bruit et de la taille du plan initial. À des plans initiaux de 20 et 40 points, 50 points sont ajoutés séquentiellement à l'aide du critère PEI*

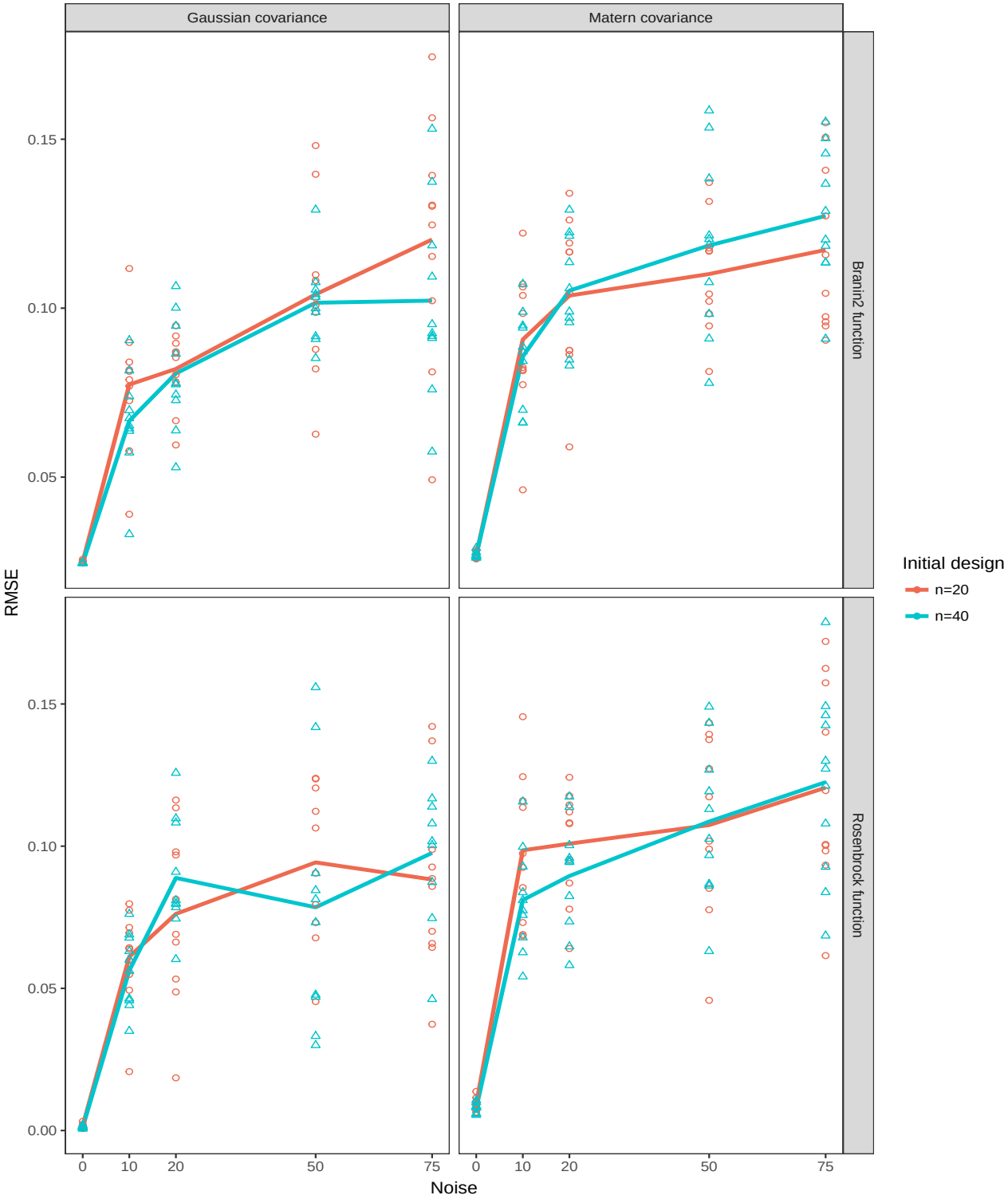


FIGURE 4.6 – *Influence de la taille du plan initial sur $PEQI$. RMSE de la ligne de crête en fonction du niveau de bruit et de la taille du plan initial. À des plans initiaux de 20 et 40 points, 50 points sont ajoutés séquentiellement à l'aide du critère $PEQI$ avec $\beta = 0.1$*

Pour les deux méthodes d'optimisation et sur l'ensemble des fonctions test, le noyau de covariance gaussien semble mieux convenir; cela peut venir du fait que les fonctions considérées sont assez régulières. En effet, la $RMSE$ est toujours la plus faible avec le noyau gaussien, jusqu'au niveau de bruit égal à 50% du signal. A niveau de bruit égal

à 75% du signal, avec le critère PEI les performances du noyau gaussien se dégradent notablement au point de devenir équivalentes à celles obtenues avec le noyau Matérn. Cette dégradation est beaucoup moins sensible, voire pas observée avec le critère $PEQI$. On note aussi que le choix du noyau de covariance influence moins les performances obtenues avec le critère $PEQI$ qu'avec le critère PEI.
En résumé, le critère $PEQI$ donne des résultats meilleurs que ceux de PEI, plus stables par rapport au choix du noyau de covariance et plus résistants au bruit.

V

Application sur un modèle de culture

L'objectif global de cette thèse est d'assister la sélection variétale dans le but d'améliorer la production agricole du sorgho au Sahel. Pour un génotype donné, la production est variable d'une condition climatique à une autre et les conditions climatiques sont très variables d'un endroit à un autre du Sahel. Le meilleur génotype est alors fonction de l'endroit cible.
Dans cette partie, nous avons défini une carte d'idéotypes de sorgho en utilisant un modèle de culture. Les modèles de culture sont différents dans les formalismes qu'ils utilisent pour représenter le système sol-plante en interaction avec le climat. Pour notre problème, nous allons utiliser le modèle SAMARA [33] qui est un modèle de simulation de culture fonctionnant au pas de temps journalier.

Cependant, l'utilisation du modèle SAMARA pour notre recherche de carte d'idéotypes demande un travail préalable. En effet, pour rendre compte de l'impact de la variabilité climatique sur la croissance de la plante, SAMARA a besoin de données météorologiques quotidiennes comme la vitesse moyenne du vent, la pluviométrie, l'heure d'ensoleillement, la température minimale et maximale et l'humidité relative maximale et minimale. Or dans les pays en développement, la disponibilité de données météorologiques continues constitue un réel problème. Les générateurs stochastiques de climat ont été adoptés comme un outil peu coûteux pouvant être utilisé pour produire des variables d'entrée pour des modèles de culture. Nous avons donc couplé le modèle SAMARA avec le générateur stochastique de climat Marksim [79].

5.1 Couplage modèle SAMARA et générateur stochastique de climat Marksim

5.1.1 Le modèle SAMARA

Les modèles de culture utilisent presque tous les mêmes concepts pour représenter les mécanismes de croissance de la plante. Cependant, ils différent dans la manière dont les hypothèses posées y sont formalisées. Le modèle Samara est fondé sur un bilan carboné. L'interception de la lumière et sa transformation par la photosynthèse constituent la source d'éléments carbonés. Ces éléments sont répartis entre différents puits : les réserves (sucre dans les tiges, amidon dans les grains), la constitution des structures (tige, racines, feuilles), et leur respiration. La répartition se fait à proportion des forces des différents puits. A chaque instant l'intensité de la photosynthèse dépend de la disponibilité en eau pour les racines. Les apports d'eau se limitent dans notre étude à la pluie; alors que les consommations correspondent à du ruissellement, de l'évaporation (du sol), à la transpiration des plantes et au drainage qui a lieu quand la réserve utile du sol est excédée. Le but est de maximiser l'espérance ou un quantile du rendement en grains. La floraison doit survenir idéalement un mois avant l'arrêt des pluies, pour éviter la moisissure des grains. Le remplissage des grains se fait alors principalement en exploitant la réserve utile du sol. Les grains mis en place vont accueillir les réserves en sucre accumulées dans les tiges et le produit de la photosynthèse faite sur la réserve en eau du sol. Suivant les réserves disponibles et le nombre de grains prêts à les accueillir, le remplissage des grains sera plus ou moins complet.

Pour notre problème de recherche d'idéotypes, nous voulons un modèle dont certains paramètres changent en fonction des caractéristiques des génotypes. Ceci permettra de rendre compte de la différence de fonctionnement des génotypes.
Le modèle SAMARA est adapté à notre problème car il est développé pour étudier les concepts de plantes in-silico (idéotypes) soumises à différents environnements climatiques, différents types de sol et différents itinéraires techniques [33]. En tant que tel, il permet de faire des compromis entre les caractères d'une plante dans le but d'optimiser un résultat agronomique comme le rendement.
Avec ce modèle sont simulés à un pas de temps journalier le développement de la plante, sa production de biomasse et son allocation entre organes végétatifs et reproductifs.

5.1.2 Générateur stochastique de climat Marksim

Le générateur stochastique de climat MarkSim a été développé dans les années 90 pour simuler des séries climatiques à partir de sources connues de données climatiques provenant du monde entier [79, 80, 81, 82, 83]. Le monde y est divisé en 720 groupes climatiques distincts les uns des autres. L'algorithme de base de MarkSim est un simulateur de précipitations quotidiennes. La pluviométrie est modélisée à l'aide d'une chaîne de Markov de troisième ordre en deux étapes. On détermine d'abord si un jour donné est pluvieux ou pas, et ensuite, la quantité de pluie. La probabilité qu'un jour i soit pluvieux dépend de la pluviosité des trois jours précédents. Elle est définie par un modèle probit:

$$P(W/D_1D_2D_3) = \Phi^{-1}(b_i + a_{i-1}d_1 + a_{i-2}d_2 + a_{i-3}d_3)$$

où Φ^{-1} est l'inverse de la fonction de répartition de la loi normale centrée réduite, b_i est le probit de base mensuel d'un jour humide après 3 jours secs consécutifs, a_i une indicatrice égale à 1 s'il a plu le jour i et 0 sinon et d_i un paramètre de persistence. La probabilité d'une journée humide est donc spécifiée par 15 paramètres: les probabilités de référence, b_i, pour chaque mois, et trois paramètres de persistence, d_1, d_2 et d_3, qui ne varient pas d'un mois à l'autre.
Les précipitations en cas de pluie sont ensuite modélisées à l'aide de la distribution gamma tronquée, restreinte aux valeurs supérieures ou égales à 0.1 mm, pour déterminer la quantité quotidienne de pluie les jours où des précipitations sont observées. La méthode du maximum de vraisemblance est utilisée pour estimer la moyenne et les paramètres de cette distribution pour chaque mois, ce qui fait au total 24 paramètres supplémentaires.
Pour générer des relevés pluviométriques, les probabilités mensuelles de base (probabilité de pluie après 3 jours de pluie pendant 3 jours consécutifs) sont interpolées en probabilités quotidiennes à l'aide de la transformée de Fourier en 12 points [77].
De même, les paramètres mensuels de moyenne et de forme de la distribution gamma des quantités de précipitations sont interpolés aux valeurs quotidiennes à l'aide de la transformée de Fourier en 12 points.
Marksim permet une estimation des températures maximales et minimales quotidiennes et des valeurs quotidiennes du rayonnement solaire en se basant sur les méthodes utilisées dans Richardson [113]. La température maximale, la température minimale et le

rayonnement solaire sont considérés comme des processus stochastiques multivariés continus avec des moyennes quotidiennes et les écarts-types conditionnés à l'état humide ou sec du jour. La série chronologique de chaque variable est ensuite réduite à une série chronologique d'éléments résiduels en centrant et réduisant les observations, selon les équations suivantes:

$$\chi_{p,i}(j) = \begin{cases} \dfrac{X_{p,i}(j) - \bar{X}_i^0(j)}{\sigma_i^0(j)} & \text{si } Y_{p,i} = 0 \\ \dfrac{X_{p,i}(j) - \bar{X}_i^1(j)}{\sigma_i^1(j)} & \text{si } Y_{p,i} > 0 \end{cases} \tag{5.1}$$

où $Y_{p,i}$ représente la quantité de précipitation le jour i de l'année p, $\bar{X}_i^0(j)$ et $\sigma_i^0(j)$ sont respectivement la moyenne et l'écart-type pour une journée sèche ($Y_{p,i} = 0$), $\bar{X}_i^1(j)$ et $\sigma_i^1(j)$ la moyenne et l'écart-type pour une journée pluvieuse ($Y_{p,i} > 0$), $\chi_{p,i}(j)$ représente donc un élément résiduel pour la variable j, le jour i de l'année p.
Un procédé de génération proposé par Matalas [97] est utilisé pour ajuster les résidus des températures maximales, des températures minimales et des rayonnements solaires. Le modèle est sous la forme

$$\chi_{p,i}(j) = A\chi_{p,i-1}(j) + B\epsilon_{p,i}(j) \tag{5.2}$$

où $\chi_{p,i}(j)$ et $\chi_{p,i-1}(j)$ représentent respectivement les résidus pour les jours i et $i-1$ de l'année p de la variable j (de la température maximale ($j = 1$), de la température minimale ($j = 2$) et du rayonnement solaire ($j = 3$)); $\epsilon_{p,i}(j)$ variable aléatoire normalement distribuée, centrée réduite, et A et B les paramètres du modèle. L'utilisation du modèle 5.2 suppose que les résidus de la température maximale, de la température minimale et du rayonnement solaire sont normalement distribués et que la corrélation temporelle de chacune de ces variables peut être décrite par un modèle autorégressif linéaire d'ordre 1.
Comme pour la pluviométrie, une série de Fourier est utilisée pour lisser les moyennes saisonnières et les écarts-types.
Les paramètres de la chaîne de Markov ont été estimés sur 9200 stations météorologiques où des données journalières étaient disponibles, et les résultats ont été classifiés en groupes climatiques [78]. Pour chaque groupe, a été estimée une régression de ces paramètres sur les moyennes mensuelles de la pluviométrie totale et des températures journalières minimales et maximales. Ces moyennes mensuelles sont disponibles sur quasiment toutes les terres émergées. En effet, elle ont été interpolées sur une grille de pas 30 secondes d'arc, ce qui représente une distance au sol de 1,1 km à l'équateur, et qui diminue comme le cosinus de la latitude; et elles ont été mises en ligne sur le base de données WordClim, que l'on interroge facilement à partir de R à l'aide des fonctions getData et extract du package raster. Par exemple, pour les moyennes mensuelles de la température minimale :

$$coordonnee = cbind(lat, lon)$$

$$grid.tmin = getData('worldclim', var =' tmin', res = 0.5, lat = lat, lon = lon)$$

$$tmin = extract(grid.tmin, coordonnees)$$

Les auteurs du générateur ont trouvé que ce modèle de dépendance à 3 jours à coefficients saisonniers ne suffisait pas à retrouver la variabilité de la pluviométrie annuelle totale. Pour retrouver cette variance, ils modulent les probabilités de transition de la chaîne de Markov d'une année simulée à l'autre, suivant une loi qui a pour variance la variance d'estimation de ces paramètres.

5.2 Détermination d'une carte d'idéotypes de sorgho

5.2.1 Choix des paramètres pour l'optimisation

Les modèles de culture possèdent généralement de nombreux paramètres, et Samara n'échappe pas à cette règle puisqu'il en a 76. Il n'est donc pas question de les optimiser tous en comptant sur l'algorithme pour décider de leur intérêt.

Parmi ces paramètres, la plupart ne sont pas de bons candidats à une optimisation à l'aide d'un algorithme. Soit parce qu'ils ne sont pas variables au sein de l'espèce, soit parce que la sortie d'intérêt (ici le rendement en grains) n'y est pas sensible. Toutefois cette sensibilité dépend du milieu, il n'est pas donc pas judicieux de rejeter des paramètres sur la base d'une analyse de sensibilité menée sur une saison des pluies particulière. A rejeter aussi est le paramètre dont on sait que la sortie d'intérêt en est une fonction monotone. C'est le cas par exemple de l'efficience de conversion de l'énergie lumineuse en glucides : l'augmenter fait croître invariablement le rendement, inutile d'utiliser un algorithme d'optimisation pour s'en apercevoir. C'est par avis d'expert que l'on connaît cette monotonie plutôt que par analyse de sensibilité, car là encore la dérivée de la sortie d'intérêt par rapport aux paramètres dépend du milieu. C'est aussi par avis d'expert que l'on peut choisir quelques paramètres sur lesquels la décision ne va pas de soi, car ils ont des effets antagonistes sur le rendement, et doivent sans doute faire l'objet d'un compromis. On peut penser par exemple que pour augmenter le rendement, il faudra agir sur les paramètres suivants :

- la longueur maximale des racines (RootFrontMax) : investir dans de longues racines consomme de l'énergie qui ne se retrouvera pas dans les grains. Mais cela permet aussi de résister à des séquences de jours sans pluie (des épisodes secs) et de sauver alors la production. C'est donc un paramètre tout indiqué pour la recherche d'un compromis.
- la longueur du cycle végétatif (ndjbvp): l'augmenter permet de produire de nombreuses feuilles et d'accumuler de la biomasse qui va permettre de produire du grain. Mais cela fait prendre aussi le risque que la saison des pluies soit trop courte pour qu'à son terme le grain soit rempli et prêt à mûrir
- le coefficient de sensibilité de la mortalité des feuilles (leaf.death) : règle la mortalité des feuilles en fin de cycle, pendant la phase de remplissage des grains. Réglé à 0 on a un sorgho "stay green", qui garde ses feuilles vertes; il sera donc capable de faire de la photosynthèse et de valoriser ainsi la réserve en eau du

sol, ce qui est favorable; mais d'un autre côté les feuilles vertes respirent, et donc consomment des réserves; il y a donc un compromis à faire entre la consommation de réserves de carbone et l'espérance de transformation de la réserve en eau en carbohydrates.

- la capacité de réserve dans les tiges (stem.reserve) : à la fin de la saison des pluies le sorgho peut mobiliser des réserves accumulées dans sa tige (plus précisément dans les entre-noeuds). Mais ces réserves peuvent s'avérer inutiles si les conditions sont favorables.

Dans la section suivante, nous allons faire une étude de la sensibilité de la distribution du rendement aux 4 paramètres retenus dans quelques endroits du Sahel.

5.2.2 Étude de la sensibilité de la distribution du rendement

Dans cette partie, nous étudions la distribution des rendements aux quatre paramètres retenus pour l'optimisation : la longueur de la phase végétative, de 200 à 500 degrés-jours, la profondeur maximale d'enracinement, de 200 à 1500 mm, le coefficient de sensibilité de la mortalité des feuilles à la restriction des réserves carbonées, de 0 à 0.1, et la capacité maximale de réserve dans les tiges, exprimée en proportion de la masse structurale des tiges, de 0 à 2. Pour chacun de ces paramètres 3 valeurs ont été retenues : les deux bornes de l'intervalle et son centre. Les 3 valeurs de chacun des 4 paramètres ont été combinées suivant un plan complet à $3^4 = 81$ éléments.
Pour chaque lieu retenu dans cette analyse, 99 années de données météo ont été générées pour servir d'entrées à SAMARA avec les 81 combinaisons de valeurs des paramètres. Les 81 distributions obtenues ont été représentées par un diagramme quantile-quantile permettant de repérer les situations à risque, où la distribution des rendements n'est pas gaussienne. La répartition des 81 valeurs des différents paramètres est illustrée par le graphique 5.1. En colonnes, nous avons les valeurs prises par le coefficient de sensibilité de la mortalité des feuilles (pour un sorgho "stay green" leaf.death=0, pour un sorgho avec des feuilles à sensibilité intermédiaire leaf.death=0.05 et pour un sorgho asséchant ses feuilles leaf.death=0.1). Les trois valeurs de la capacité maximale de réserve dans les tiges sont imbriqués dans chaque valeur du coefficient de sensibilité de la mortalité des feuilles. En ligne, nous avons les différentes valeurs de la longueur de la phase végétative (sorgho cycle court ndjbvp=200, sorgho cycle moyen ndjbvp=350 et sorgho cycle long ndjbvp=500). Les trois valeurs de la longueur maximale des racines sont imbriquées dans chaque valeur de la longueur du cycle. Pour une longueur de cycle et un coefficient de sensibilité de la mortalité des feuilles fixées, une lecture en colonne correspond à une augmentation de la capacité de réserve dans les tiges tandis qu'une lecture en ligne correspond à une augmentation de la longueur maximale des racines. Les figures (5.3, 5.4, 5.5, 5.6, 5.7) montrent la réponse de la distribution des rendements aux quatre paramètres retenus dans cinq lieux.

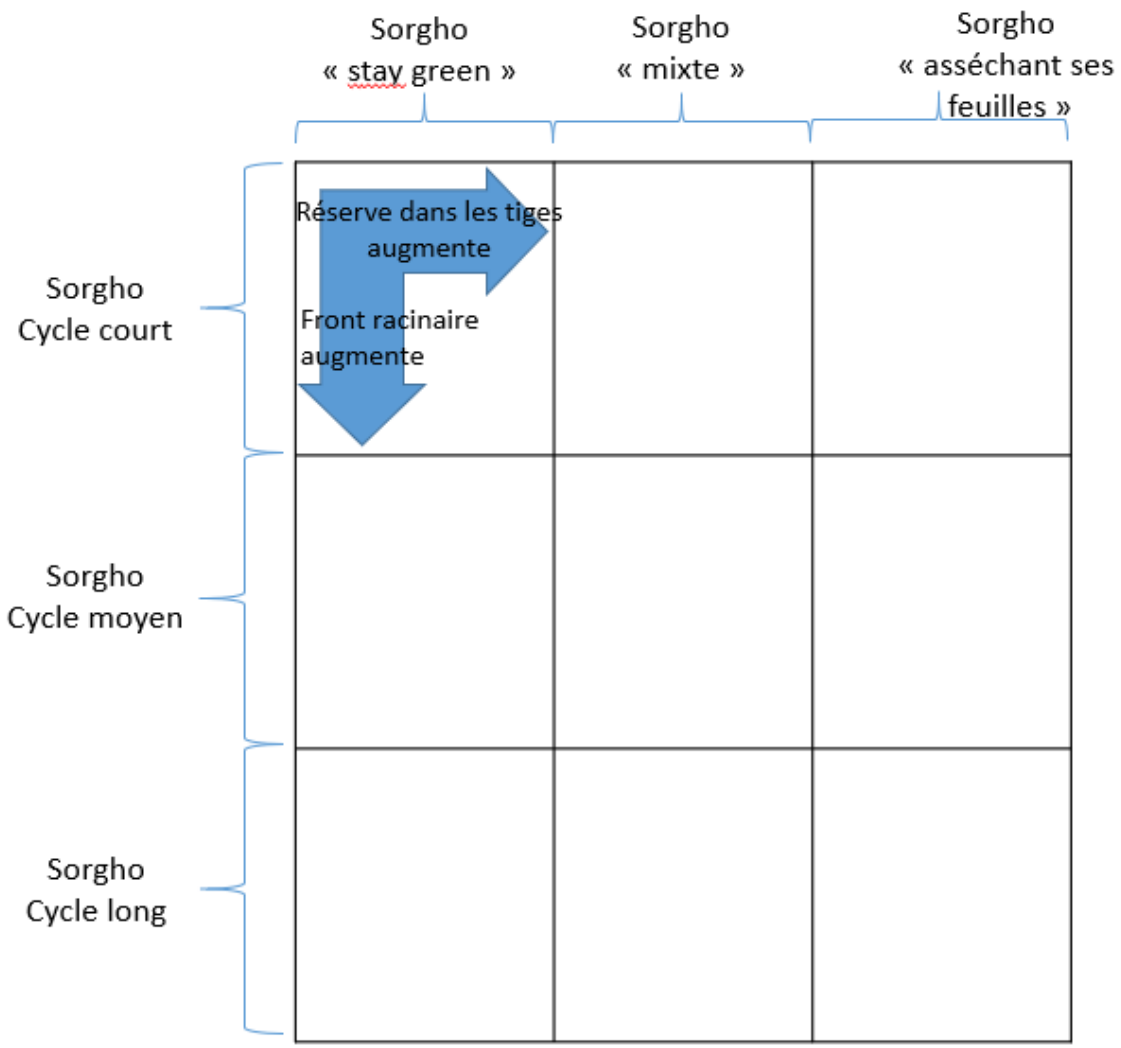


FIGURE 5.1 – *Disposition des différentes valeurs des différents paramètres*

Les lieux retenus sont régulièrement répartis sur une ligne Nord-Sud à -15.09 degrés de longitude Ouest, et une latitude variant de 11.30 à 15.85 degrés Nord. Ces points sont situés respectivement non loin de Catio, Ziguinchor, Nioro du Rip, Diourbel, et entre Louga et Saint Louis suivant un gradient Nord Sud (figure 5.2).

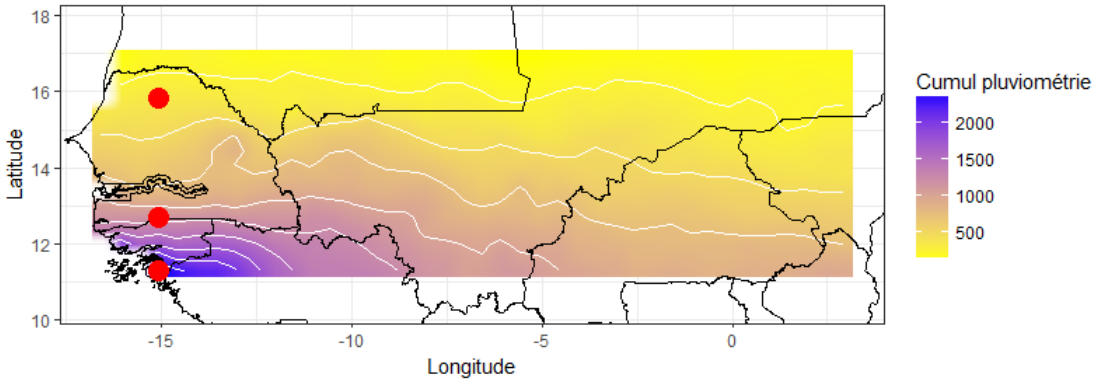


FIGURE 5.2 – *Pluviométrie moyenne annelle en (mm)*

Près de Catio (figure 5.3)
Les meilleurs rendements sont fournis par des variétés avec une longueur de cycle maximale, de courtes racines, une forte sensibilité à la mortalité des feuilles et une capacité moyenne de réserve dans les tiges.

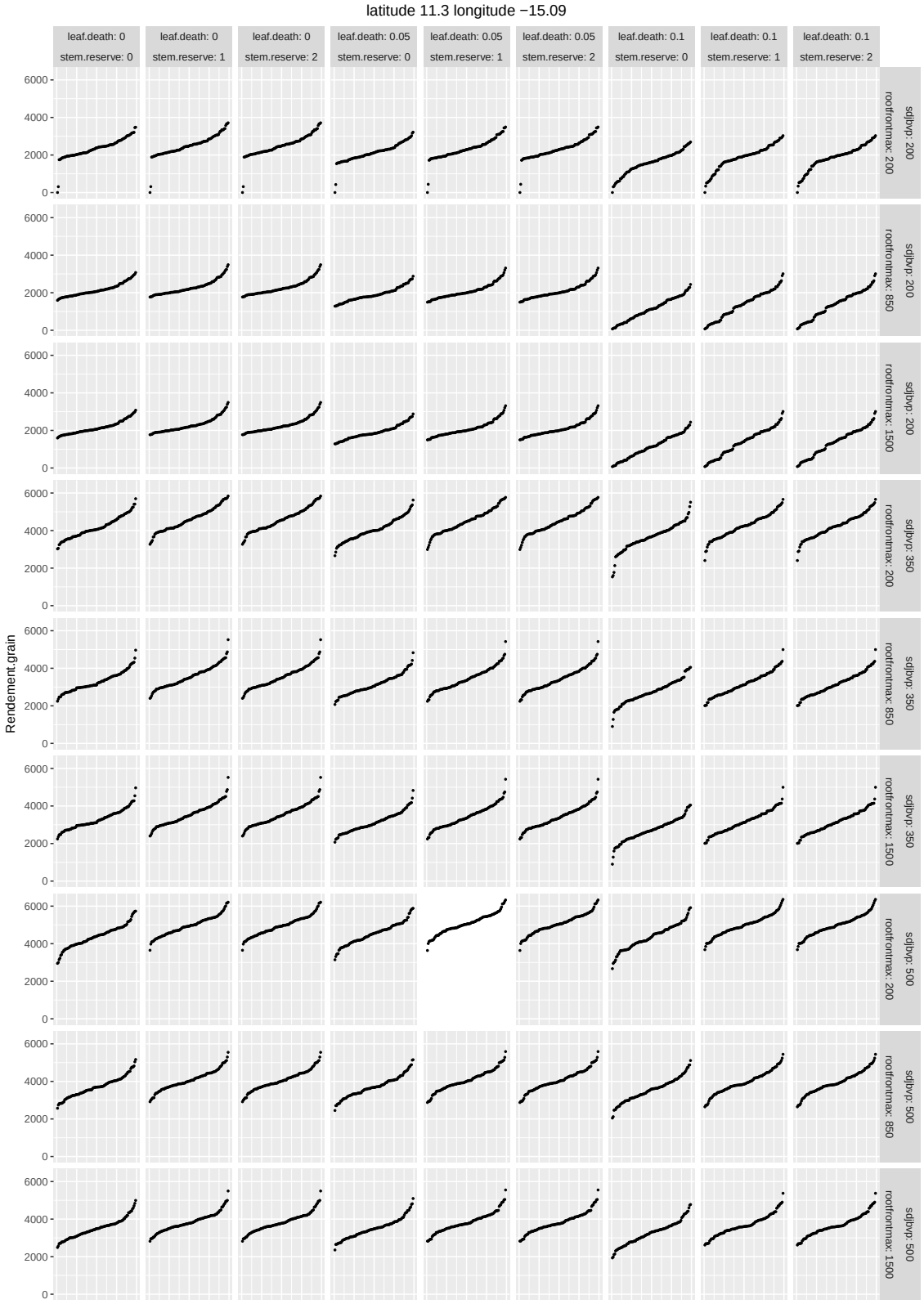


FIGURE 5.3 – *Sensibilité de la distribution des rendements à 4 paramètres près de Catio (Guinée Bissao)*

Près de Ziguinchor (figure 5.4)
Non loin de Ziguinchor, la pluie est largement suffisante et régulière pour la culture du sorgho.
Examinons le bloc de gauche sans mortalité des feuilles, sorgho dit "stay green". Avec un sorgho de cycle court (sdjbvp=200°j), passer des racines les plus courtes (200 mm) à des racines moyennes (850 mm) sécurise la production en augmentant les rendements les plus faibles. Cela correspond bien à l'idée que l'investissement dans les longues racines consomme des ressources, mais permet d'aller chercher l'eau loin dans le sol quand les pluies font défaut, et permet ainsi de sauver la production. Même à Ziguinchor, le compromis à trouver entre l'espérance de production et la limitation des risques ne se situe pas du coté des plus courtes racines.
Toujours avec un sorgho "stay green" à Ziguinchor, augmenter la longueur du cycle végétatif jusqu'à 350°j augmente le rendement moyen de 2 tonnes à 4 tonnes avec de courtes racines. Mais on voit aussi qu'il y a toujours des accidents et l'intérêt est plus grand de passer à des racines moyennes voire longues.
Augmenter encore la longueur du cycle végétatif jusqu'à 500°j augmente à peine le rendement médian, et bien plus le rendement maximum. L'intérêt de racines longues est encore de sécuriser la production en affectant légèrement le rendement médian. Dans toutes ces situations, augmenter la mortalité des feuilles ou la capacité de mise en réserve dans les tiges présente peu d'intérêt pour le rendement médian.

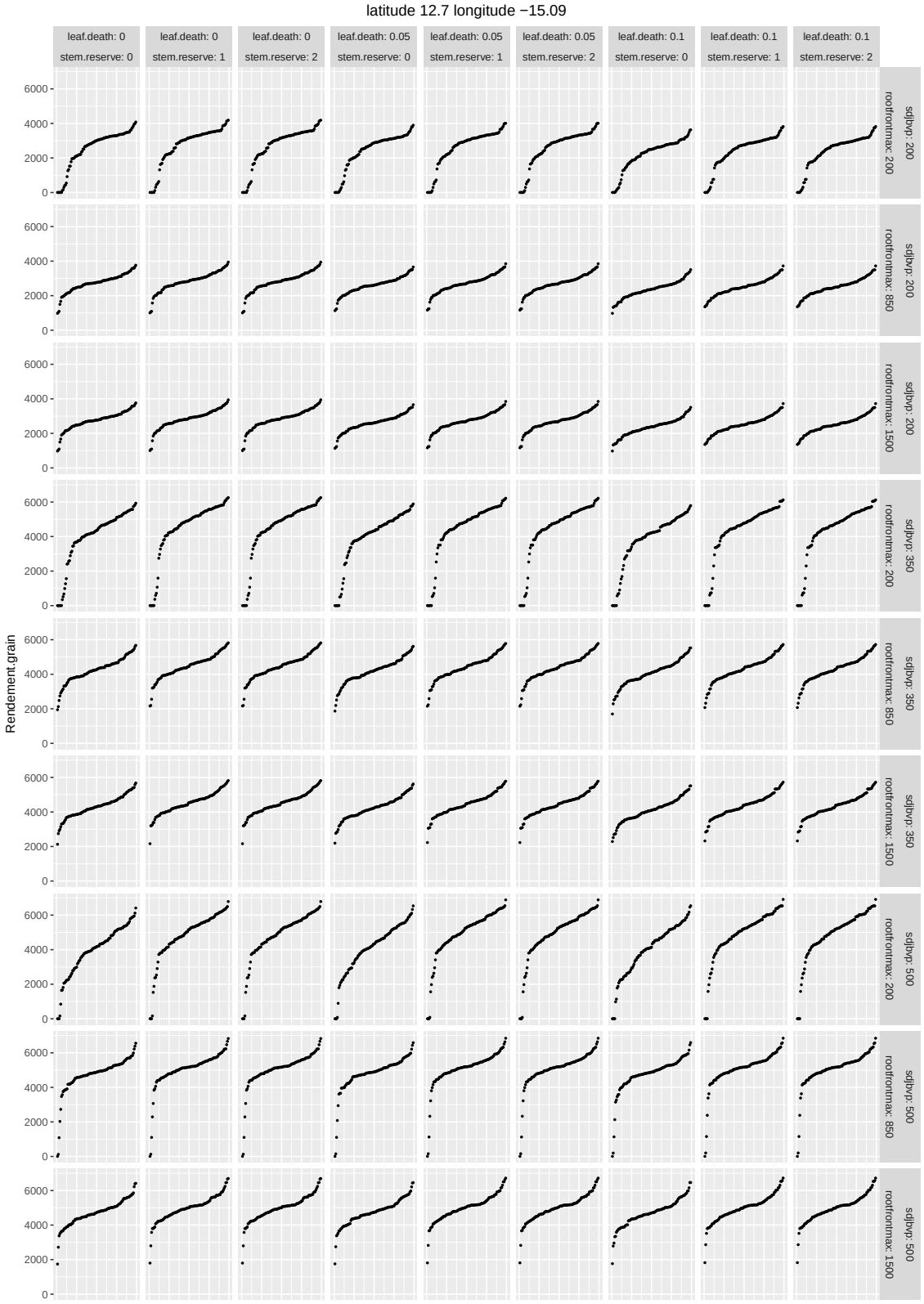


FIGURE 5.4 – *Sensibilité de la distribution des rendements à 4 paramètres près de Ziguinchor (Sénégal)*

Près de Nioro du Rip (figure 5.5)
A la latitude 13.75, près de Nioro du Rip, les pluies sont moins régulières et les rendements sont beaucoup plus variables qu'à Ziguinchor. Quelle que soit la longueur de cycle choisie, éviter les rendements nuls passe par l'abandon des racines courtes. Avec des racines moyennes (850 mm), une longueur de cycle végétatif de 350°j donne des rendements plus élevés et pas plus variables qu'avec la longueur inférieure. Augmenter cette longueur de cycle jusqu'à 500°j n'augmente pas systématiquement la médiane de la distribution: cela dépend des valeurs des autres paramètres.

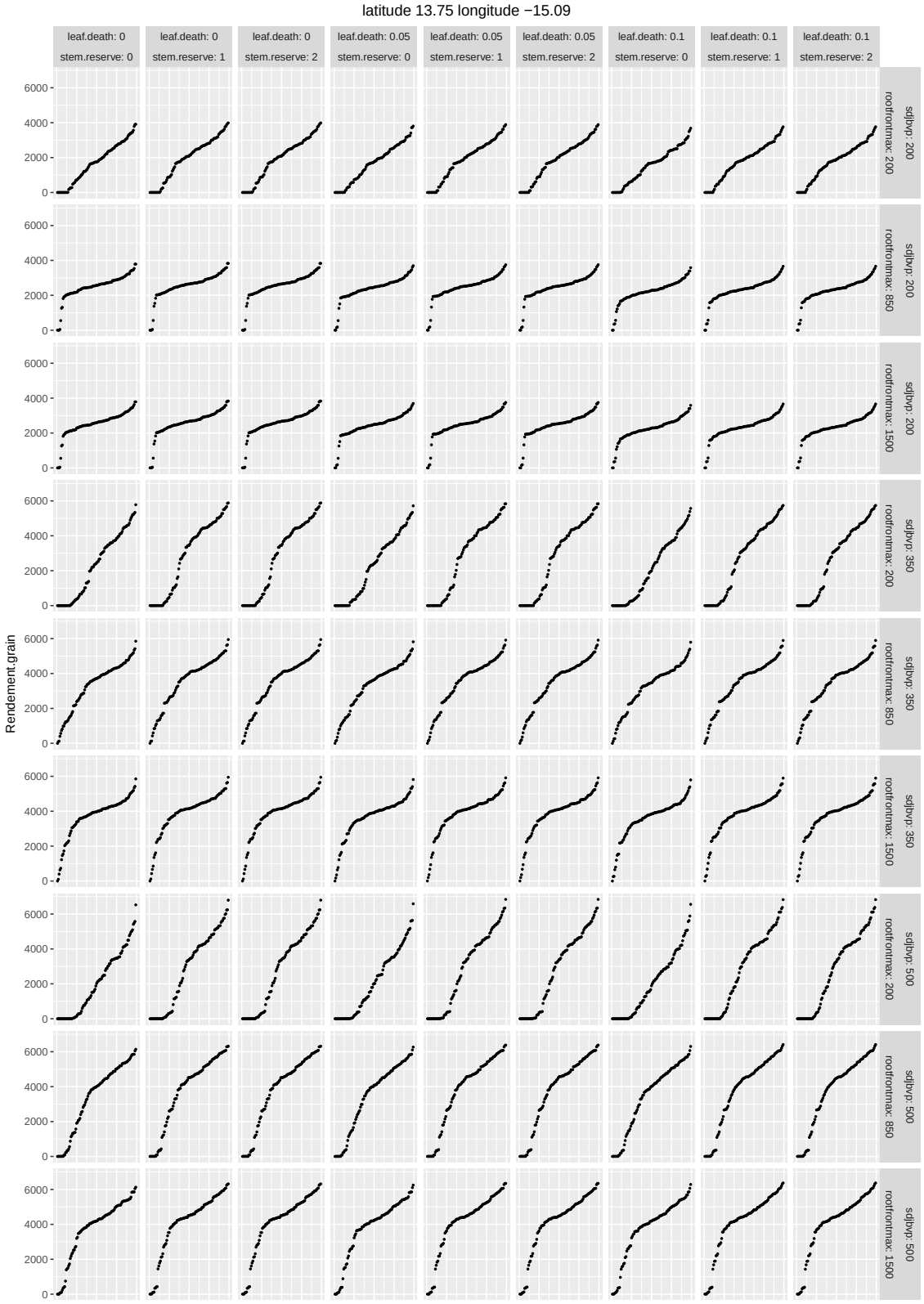


FIGURE 5.5 – *Sensibilité de la distribution des rendements à 4 paramètres près de Nioro du Rip (Sénégal)*

Près de Diourbel (figure 5.6)
A la latitude 14.8 près de Diourbel, la situation est encore plus difficile avec une saison des pluies d'arrivée plus tardive et des pluies plus rares. Avec le cycle le plus court, des racines moyennes évitent la plupart des rendements nuls, et les racines longues ne font pas mieux.
L'intérêt de racines longues est plus manifeste pour un cycle de durée moyenne, où elles déterminent une nette remontée de l'ensemble de la distribution. La durée du cycle la plus longue expose le sorgho à ne pas finir de remplir ses grains, ce qui détermine un rendement bien plus variable et une diminution de l'ensemble de la distribution. Là encore l'abandon du caractère "stay green" et la mise en réserve dans la tige ne présentent pas d'intérêt.

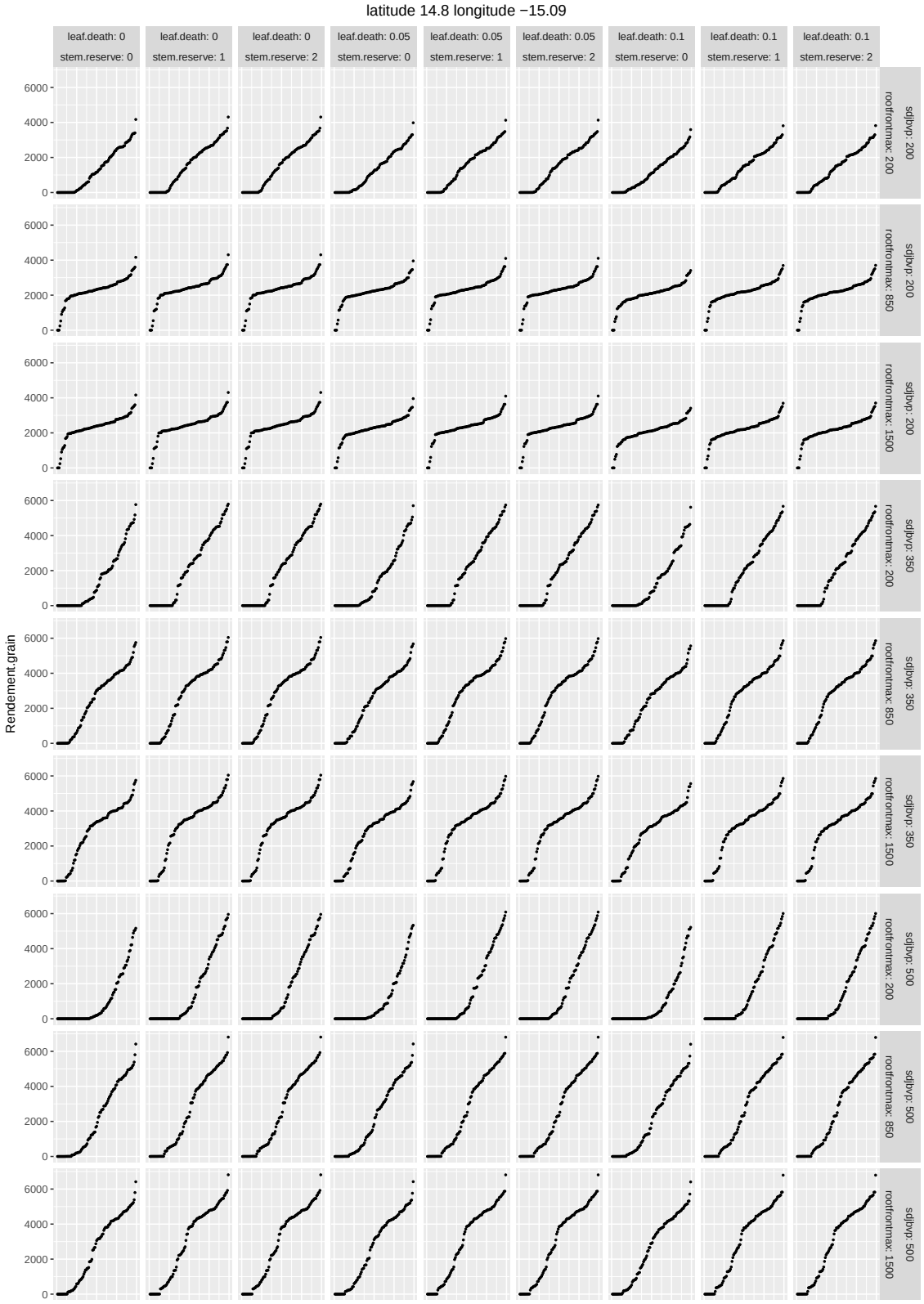


FIGURE 5.6 – *Sensibilité de la distribution des rendements à 4 paramètres près de Diourbel (Sénégal)*

Près de Louga (figure 5.7)
A la latitude 15.85 au Nord de Louga les rendements nuls ne peuvent être évités même avec le cycle le plus court et les racines les plus longues. La culture du sorgho est

risquée dans cette partie du Sénégal.

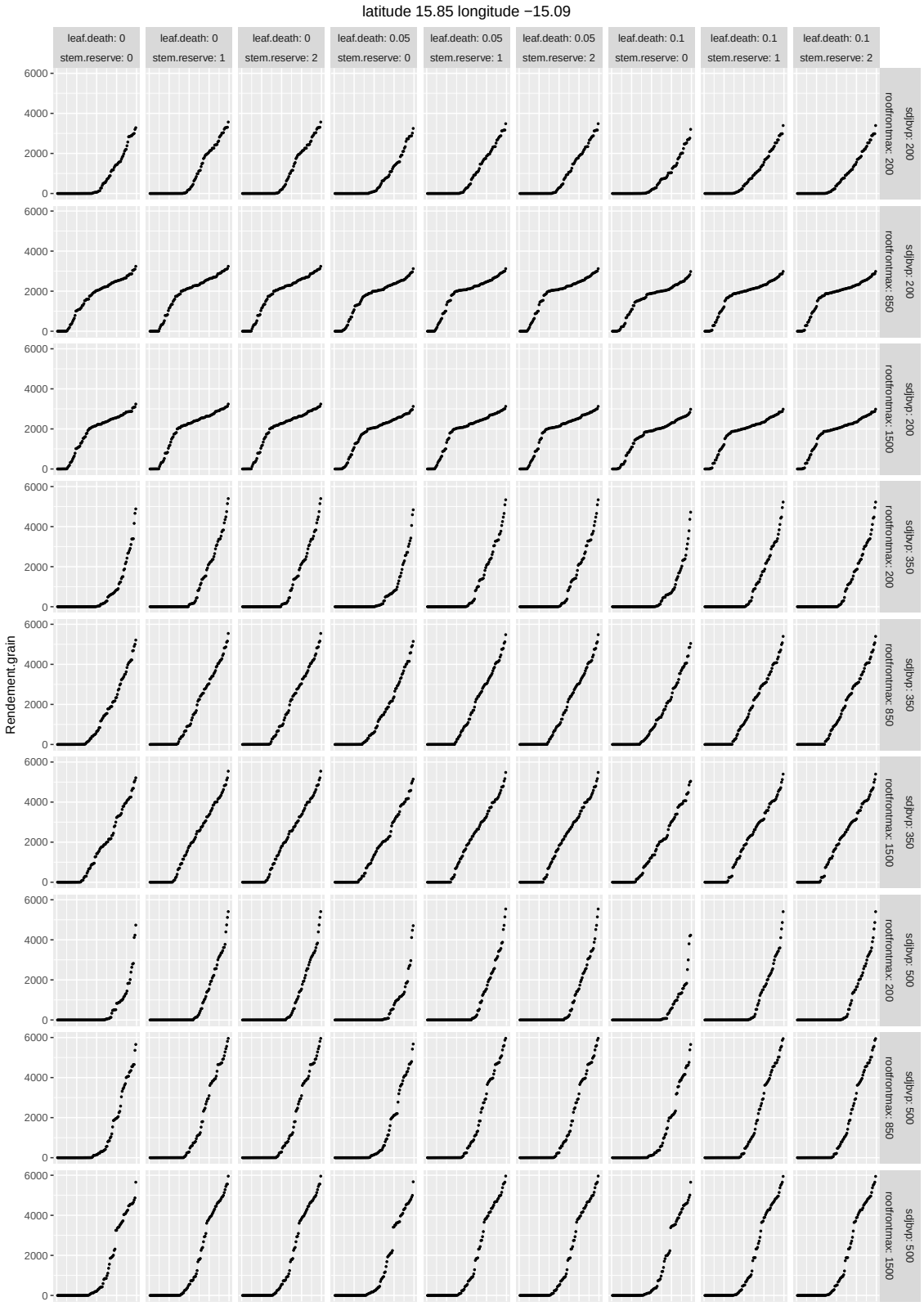


FIGURE 5.7 – *Sensibilité de la distribution des rendements à 4 paramètres près de Louga (Sénégal)*

5.3 Mise en œuvre de l'algorithme d'optimisation à base de $PEQI$ pour la détermination de la carte d'idéotypes

Nous voulons définir une carte d'idéotypes de sorgho au Sahel en utilisant le modèle de culture SAMARA. Plus spécifiquement, nous voulons déterminer un gradient des meilleures variétés sur un gradient d'environnements cibles. Une fois que le modèle de culture et le modèle de climat sont calibrés et validés avec des données observées, leur couplage nous permet de simuler la performance de différents génotypes selon différents types de climat, différents types de sol et différents itinéraires techniques. Dans le cadre de cette thèse, nous nous plaçons dans un cas simplifié où les caractéristiques du sol et les itinéraires techniques sont supposés identiques quelque soit le lieu à l'exception de la date de semis qui a été fixée pour chaque lieu et chaque année considérés. Selon les recommandations de Balme *et al* [2], il est plus prudent d'attendre d'avoir une pluie cumulée de 20 mm sur 3 jours pour semer mais nous avons porté ce seuil à 30 mm . Par conséquent, les sorties du modèle ne dépendent plus que de la variété et de l'environnement. Notre objectif revient donc à déterminer pour chaque environnement cible u, le jeu de paramètres variétaux v^* qui maximise le rendement. Cependant comme illustré sur la figure 5.8, les rendements calculés pour un même jeu de paramètres variétaux avec des saisons des pluies différentes, seront différents. Pour chaque environnement u fixé et variété v donnée, on n'a accès qu'à des rendements de la forme $\tilde{y_j}(u,v) = y(u,v) + \epsilon_j$ où ϵ_j représente un bruit de mesure que l'on supposera être la réalisation d'une variable aléatoire et $y(u,v)$ représente l'espérance du rendement pour la variété v dans l'environnement u. Dans la suite, nous faisons l'hypothèse que les bruits d'observation sont normalement distribués, centrés, et indépendantes d'une exécution à l'autre, $\epsilon_j \sim N(0,\tau^2)$.

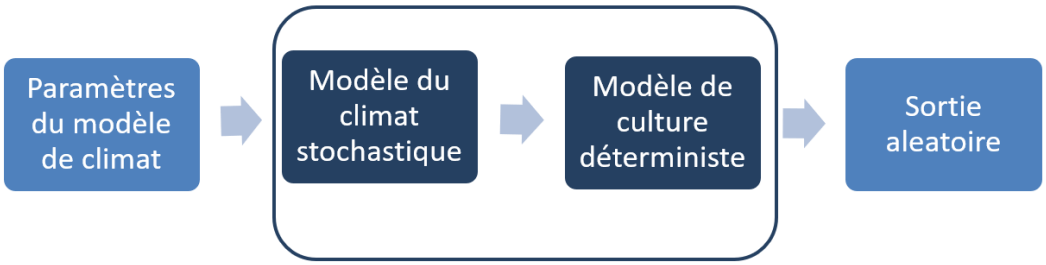


FIGURE 5.8 – *Couplage du modèle de culture avec le modèle de climat*

Pour la mise en œuvre de notre algorithme d'optimisation plus particulièrement du critère $PEQI$, un métamodèle de krigeage avec prise en compte du bruit et une fonction de covariance Matérn ont été utilisés. Les paramètres de la fonction de covariance

ont été estimés par maximum de vraisemblance en utilisant le package DiceKriging [120].
Soit $D = \{(u_1, v_1), \cdots, (u_{100}, v_{100}))\}$ le plan d'expérience initial obtenu avec la fonction optimumLHS du package lhs. La première étape de construction de la carte d'idéotypes consiste à évaluer le modèle aux points du plan d'expérience et à estimer la variance du bruit ϵ. Nous avons généré 99 réalisations du climat pour chaque environnement u_i du plan d'expérience, ce qui nous a permis d'estimer la variance de l'espérance du rendement pour une variété v_i par: $\tau_i^2 = \frac{1}{99 \times 98} \sum_{j=1}^{99} \left(\tilde{y}_j(u_i, v_i) - y(u_i, v_i)\right)^2$ où $y(u_i, v_i) = \frac{1}{99} \sum_{j=1}^{99} \tilde{y}_j(u_i, v_i)$ est la moyenne de 99 rendements.
Les résultats des évaluations et l'estimation de la variance du bruit ont été utilisés pour construire le modèle de krigeage et déterminer le nouveau point à échantillonner, c'est à dire celui qui maximise le critère $PEQI$. Nous avons ainsi évalué de façon séquentielle 200 points supplémentaires. Le plan d'expérience de même que le modèle de krigeage ont été mis à jour au cours des itérations.
A partir du modèle de krigeage obtenu avec 300 points d'évaluations, la carte d'idéotypes prédite est présentée par les cartes figures (5.9, 5.10, 5.11 et 5.12) .

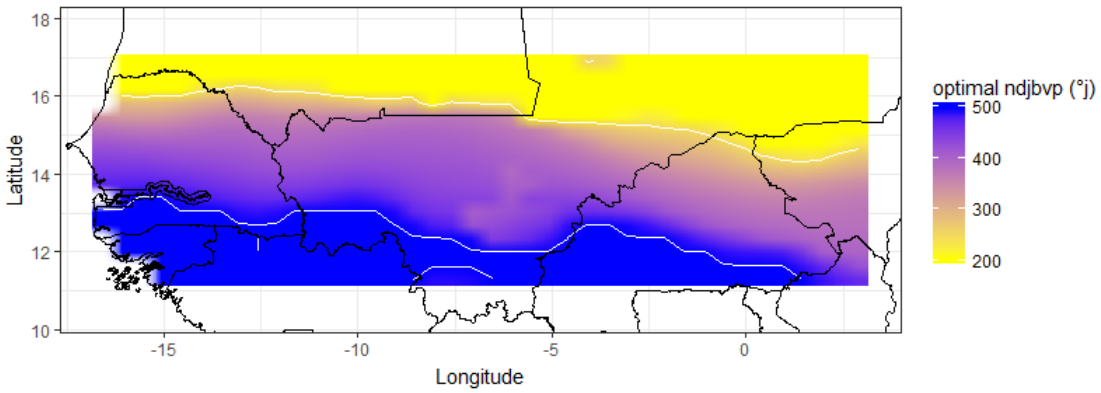


FIGURE 5.9 – *Carte des valeurs optimales de la longueur du cycle (ndjbvp) pour le rendement moyen*

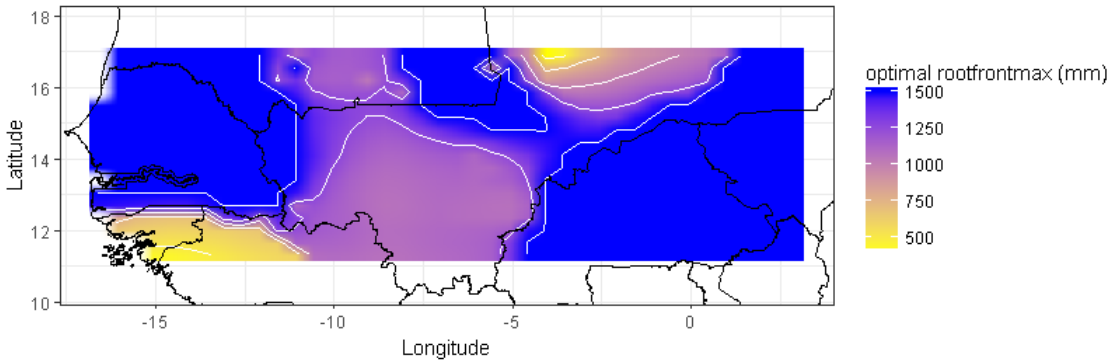


FIGURE 5.10 – *Carte des valeurs optimales de la longueur des racines (rootfrontmax) pour le rendement moyen*

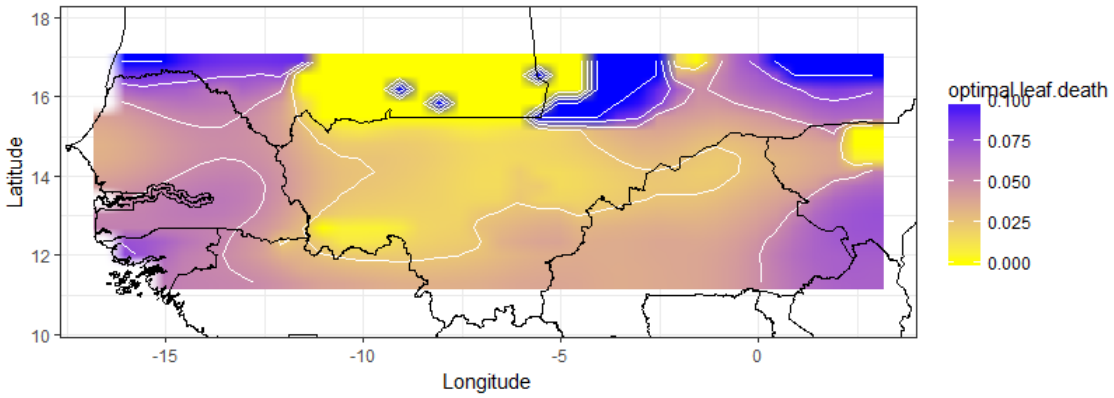


FIGURE 5.11 – *Carte des valeurs optimales de la mortalité des feuilles (leaf.death) pour le rendement moyen*

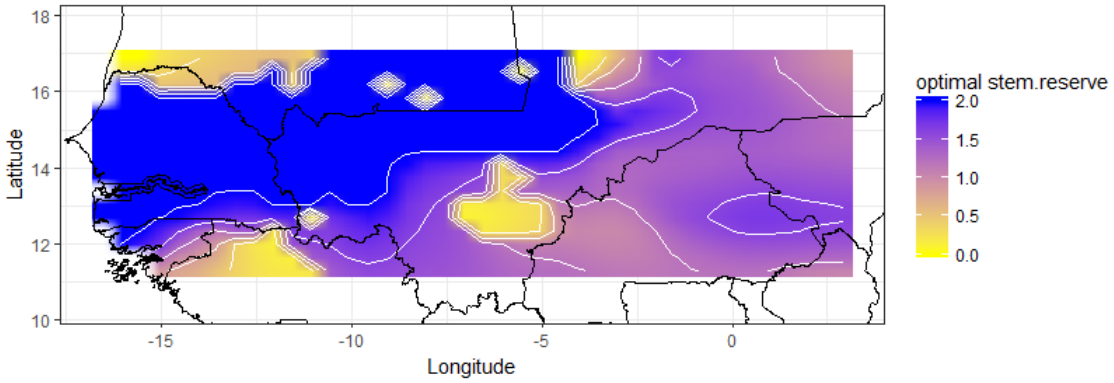


FIGURE 5.12 – *Carte des valeurs optimales de la capacité de réserves dans les tiges (stem.reserve) pour le rendement moyen*

Pour la longueur du cycle, les résultats de l'optimisation sont cohérents avec les résultats de l'analyse de sensibilité, on voit qu'à basse latitude où la pluviométrie est régulière, il vaut mieux un sorgho de cycle long. Par contre à une latitude élevée où les pluies sont rares, un sorgho de cycle court donne de meilleurs rendements.
En ce qui concerne la longueur des racines, à 15 degrés de longitude Ouest, ce n'est qu'à Cation en Guinée Bissao que le sorgho pourra se contenter de racines courtes. La transition est très rapide en se déplaçant vers le Nord où l'optimum est obtenu pour des racines longues sur tout le Sénégal. Vers 10 degrés de longitude Ouest (Mali), l'optimum est obtenu pour des racines de longueur moyenne (environ 1000 mm).
Cette apparente contradiction recoupe les analyses de sensibilité que l'on peut faire près de Ziguinchor au Sénégal (lat=12.7, lon=-15.09), près de Kassaro au Mali (lat=12.7, lon=-8.59) et à Catio en Guinée Bissao (lat=11.3, lon=-15.09).

Près de Ziguinchor, l'optimum est obtenu pour le cycle le plus long, des racines longues et un stockage maximum dans la tige, alors que le coefficient de mortalité des feuilles est sans influence (cf ligne 9 de la figure 5.4).

A la même latitude mais au Mali, c'est le cycle le plus long qui est optimal, mais cette fois ci avec des racines moyennes. Le coefficient de sensibilité de la mortalité des feuilles et la capacité de réserve dans les tiges ont peu d'influence sur le rendement (cf ligne 8 de la figure 5.13).

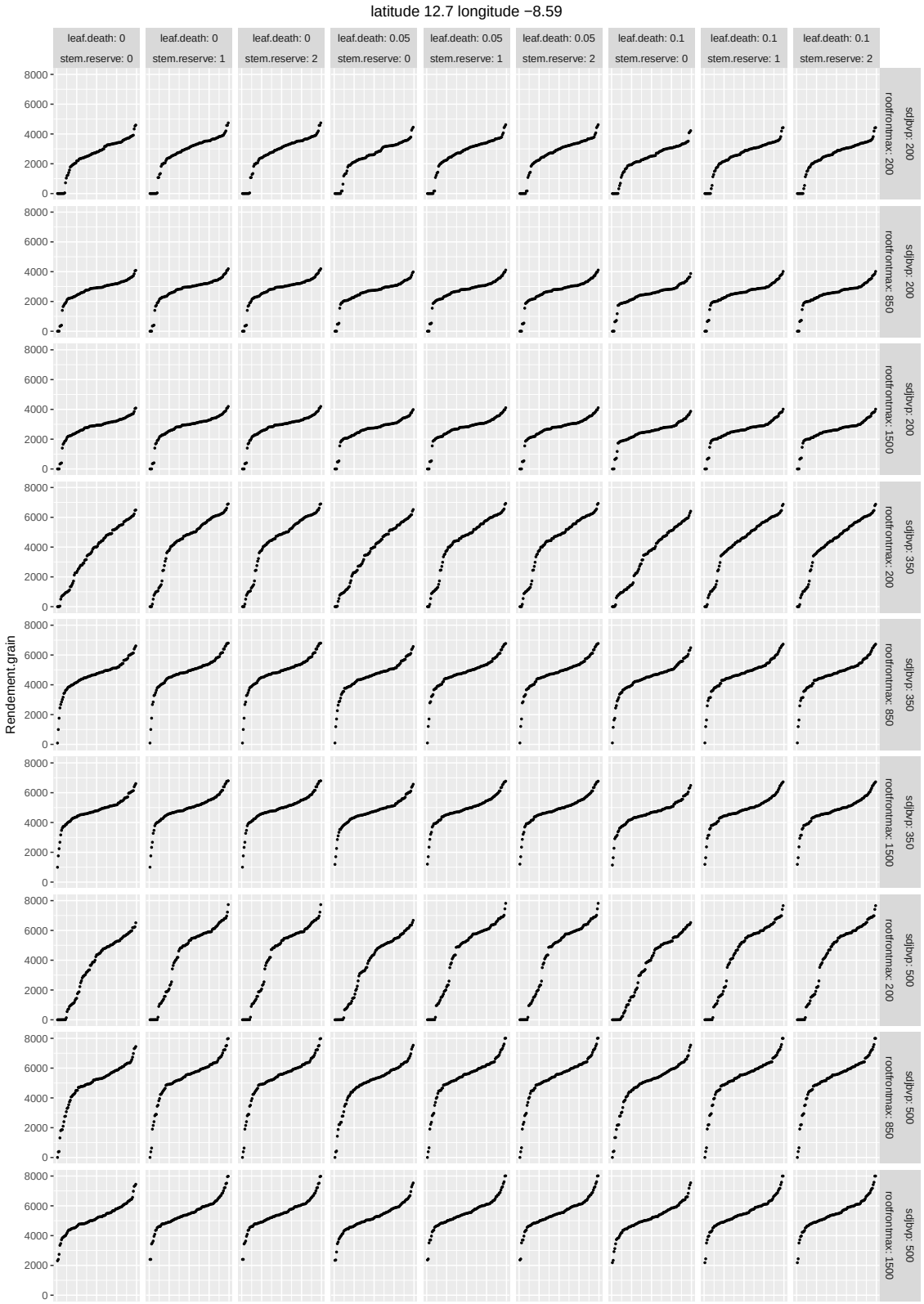


FIGURE 5.13 – *Sensibilité de la distribution des rendements à 4 paramètres près de Kassaro*

A 15.09 de longitude Ouest et 11.3° de latitude près de Catio en Guineé-Bissao, l'optimum est obtenu pour le cycle le plus long, de courtes racines, une forte sensibilité à la mortalité des feuilles et une capacité de réserve dans les tiges moyenne (cf ligne 7 de la figure 5.3).

Dans certains lieux, les paramètres autres que la longueur du cycle et la longueur des racines n'ont que peu d'influence sur le rendement. Cela détermine des cartes apparemment anormales
Ainsi les cartes des valeurs optimales pour la sensibilité de la mortalité des feuilles et la capacité de réserve dans les tiges montrent des anomalies: en Mauritanie et au Nord du Mali, vers 16 degrés de latitude Nord et 5, 8, et 9 degrés de longitude Ouest, les valeurs optimales du coefficient de mortalité des feuilles, normalement faibles s'inverse pour atteindre la valeur la plus élevée. Aux mêmes endroits, les valeurs du coefficient de réserve dans les tiges sont égales à la valeur minimale alors qu'elles atteignent la valeur maximale aux alentours. L'examen des graphiques de sensibilité (cf ligne 2 figures 5.14, ligne 2 figure 5.15 et ligne 2 figure 5.16) montrent que ces paramètres influencent peu la distribution du rendement, leurs valeurs sont donc indifférentes.

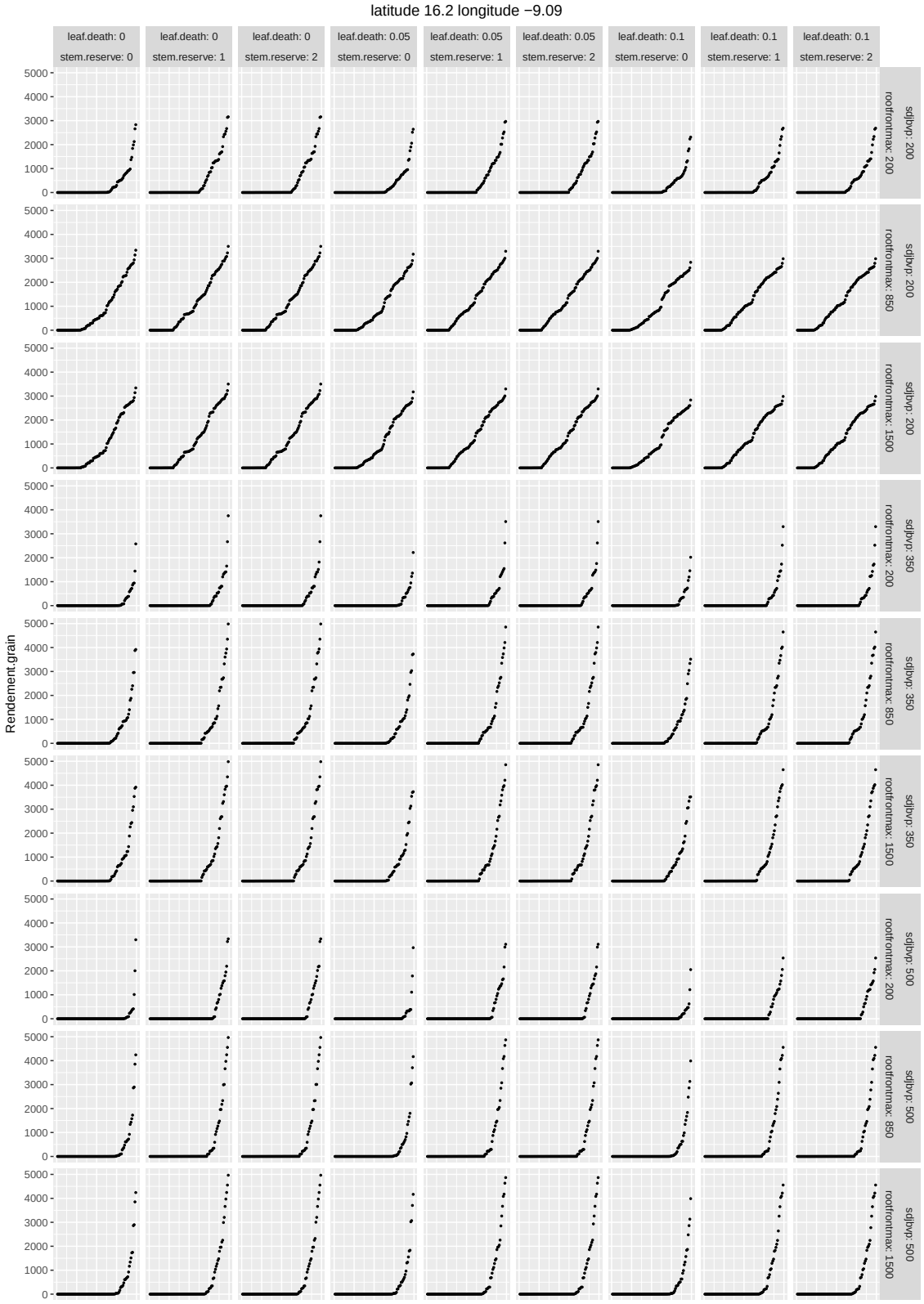


FIGURE 5.14 – *Sensibilité de la distribution des rendements à 4 paramètres près de Kobenri*

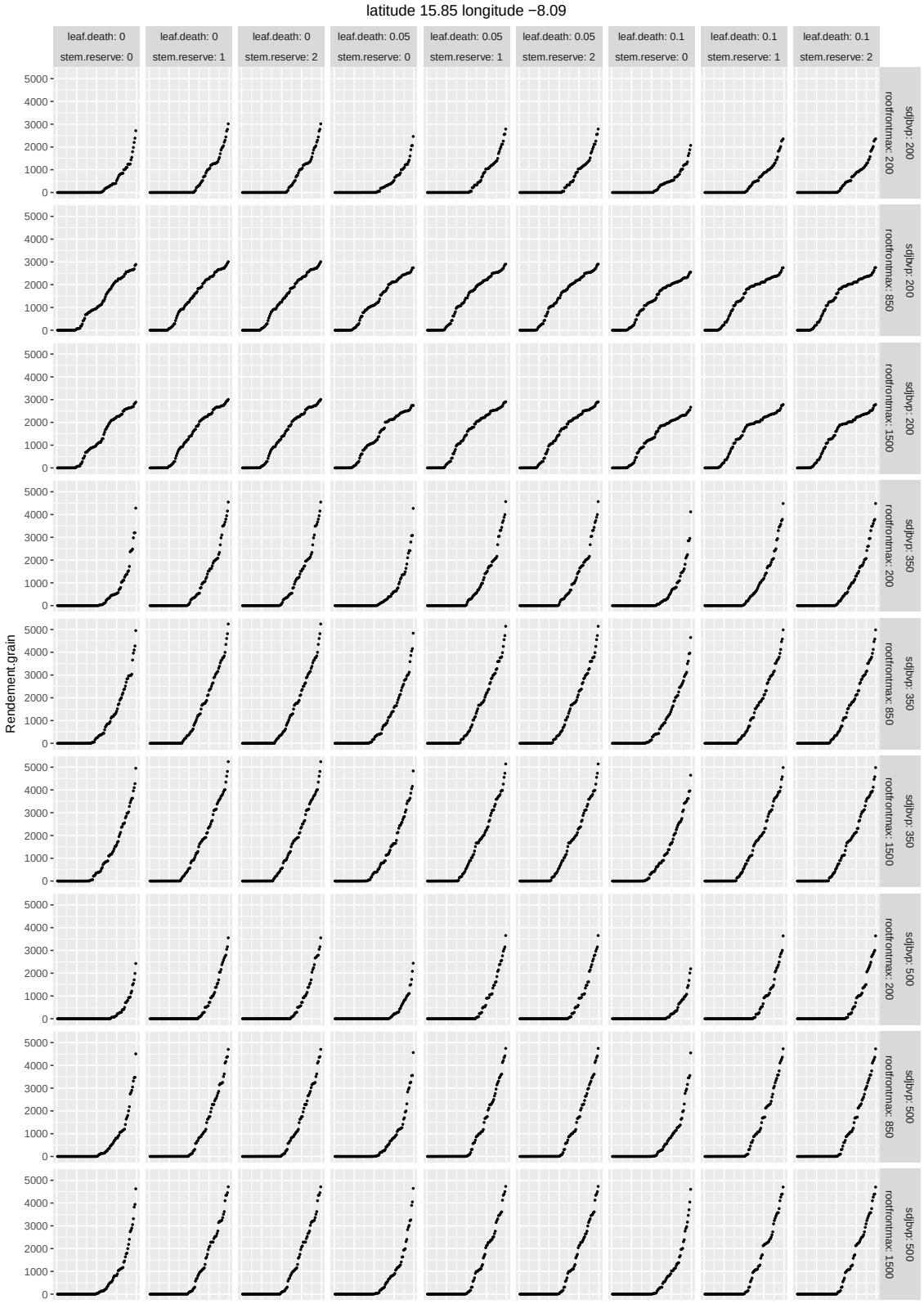


FIGURE 5.15 – *Sensibilité de la distribution des rendements à 4 paramètres près de Timbedra*

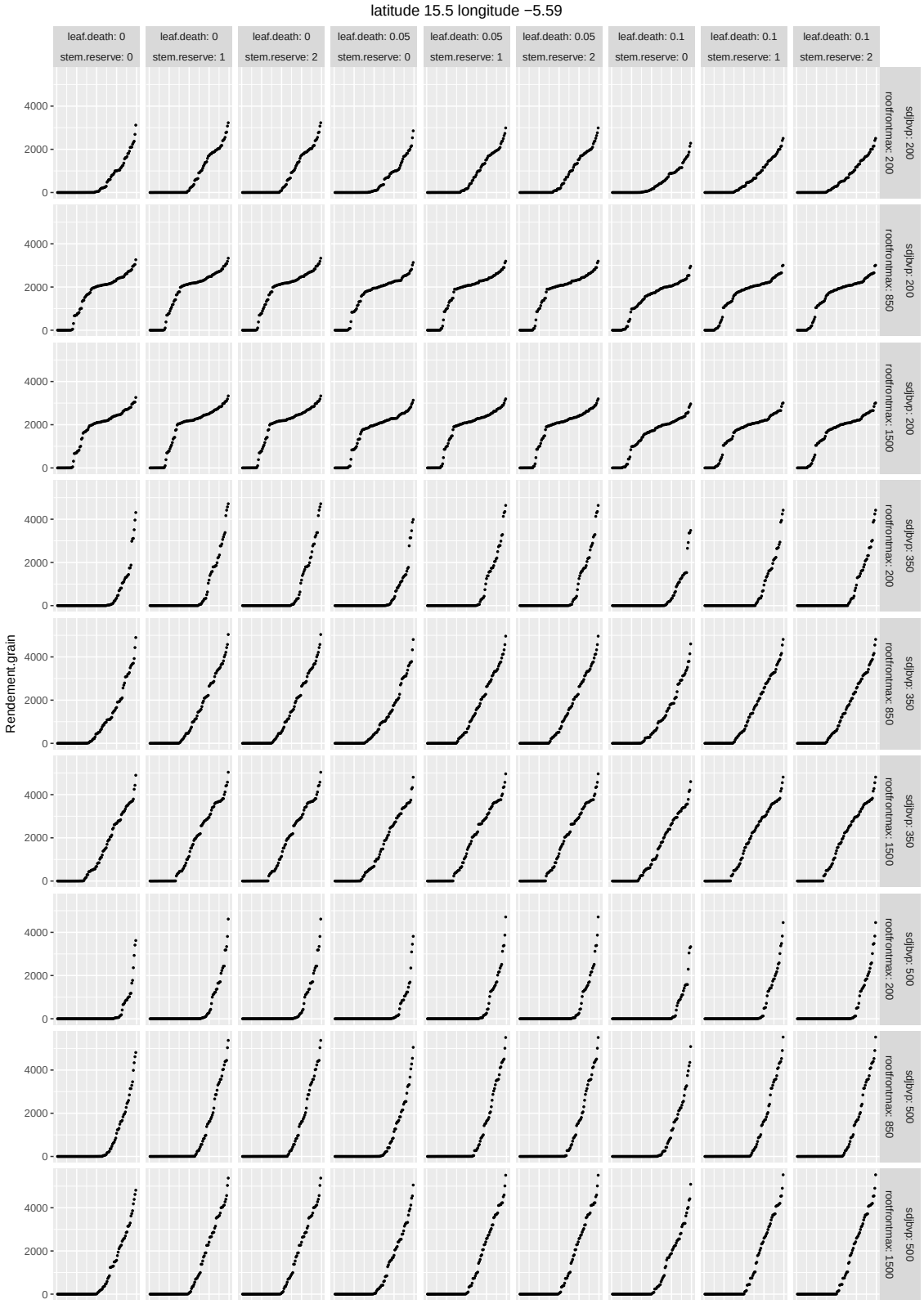


FIGURE 5.16 – *Sensibilité de la distribution des rendements à 4 paramètres près de près de Niono*

5.4 Conclusion

Il ressort de cette application que l'algorithme d'optimisation utilisant le critère $PEQI$ est capable de fournir en un temps raisonnable des résultats satisfaisants pour la recherche de carte optimale. En comparaison, l'algorithme génétique, utilisé directement sur ce même modèle SAMARA, prend bien plus de temps. Nous avons obtenu une carte d'idéotype en seulement 8 h de temps de calcul au lieu des 40 jours de calcul qu'il aurait fallu pour un algorithme génétique. Pourtant, le temps de calcul assez faible du modèle SAMARA ne le prédispose pas à l'optimisation à l'aide de métamodèle. L'avantage de ce dernier vient sans doute de la prise en compte de la corrélation des résultats entre lieux géographiquement proches, alors qu'un algorithme génétique est réinitialisé à chaque nouvelle position géographique. Un autre avantage du métamodèle sur une fonction bruitée peut être aussi le lissage de la surface de réponse des sorties aux paramètres. En effet, comme les modèles de culture sont généralement programmés avec des équations aux différences à pas de temps journalier, avec des changements de stade physiologique intervenant à heures fixes, les réponses des sorties aux paramètres peuvent être en escalier, ce qui ne facilite pas l'optimisation. Le lissage opéré par le métamodèle sur fonction bruitée gomme ces aspérités.
La carte produite gagne en précision avec le nombre d'itérations. Sur l'exemple que nous avons traité, une carte satisfaisante était obtenue après 200 itérations.
Au delà d'un plan initial minimum fonction du nombre de paramètres à estimer, chacun peut arriver au niveau de précision qu'il désire en jouant sur le nombre d'itérations. Dans ce but, l'examen de la carte et son recoupement avec des analyses de sensibilité comme nous l'avons montré permettent de juger si le niveau de précision est suffisant.
Optimiser un rendement moyen n'est pas optimiser toute une distribution. La prise en compte de la variabilité devrait peser aussi dans le choix d'une variété que l'on veut à la fois productive et stable. Un compromis entre espérance et variance peut être peut envisagé. On peut aussi maximiser non pas l'espérance mais un quantile du rendement. Dans ce cas, une adaptation de notre algorithme doit être envisagée pour remplacer le krigeage ordinaire par un krigeage d'indicatrices [84] ou un krigeage disjonctif [116, 133].

VI

Conclusion et perspectives

L'objectif global de la thèse était de trouver des idéotypes de sorgho adaptés au climat sahélien présent et futur. Le climat étant très variable quand on se déplace du Nord au Sud, la meilleure variété dépend de l'environnement cible. De ce fait, notre objectif d'amélioration variétale doit nécessairement intégrer une étude des interactions $G \times E$. Une revue des outils statistiques d'analyse de l'interaction $G \times E$ est faite dans le chapitre 1. Cette revue met en avant l'incapacité de prédiction de l'interaction $G \times E$ par les modèles descriptifs. Pour les modèles explicatifs comme la régression factorielle, les performances des génotypes peuvent être prédites dans des environnements caractérisés par un ensemble de covariables environnementales. Cependant, les covariables candidates à l'explication sont souvent trop nombreuses.
Une alternative est l'utilisation d'un modèle de culture qui permet de prédire la performances des génotypes dans différents environnements. En un lieu donné qui constitue l'environnement cible, la variabilité de cet environnement est caractérisé par un modèle de génération de données météorologiques où le climat en chaque point de l'espace est résumé par des paramètres de probabilité et de hauteur de pluie, complété par des informations sur le vent et le rayonnement. Notre problème revient donc à trouver pour chaque environnement fixé, le jeu de paramètres variétaux qui maximise une moyenne ou une médiane ou un quantile de la production. Nous avons un problème d'optimisation conditionnelle d'une fonction bruitée à résoudre. Dans le chapitre 2 de cette thèse, différentes méthodes classiques (méthodes avec calcul de dérivées, directes, stochastiques et utilisant un métamodèle) sont comparées. Pour notre travail, la fonction à optimiser est issue d'un modèle de culture. Les modèles de culture ne sont pas toujours dérivables en fonction des paramètres variétaux et la variabilité de l'environnement détermine un bruit sur la sortie moyenne du modèle. Un calcul numérique de la dérivée va donc tendre vers quelque chose de complètement aléatoire et mener l'algorithme d'optimisation à l'échec. Les méthodes directes et les méthodes stochastiques sont peu sensibles aux imprécisions du calcul de la fonction f mais sont très gourmandes en évaluations de f. Les méthodes d'optimisation basées sur un métamodèle de la fonction objectif sont moins coûteuses, en effet un critère d'échantillonnage définie à l'aide du métamodèle permet d'évaluer la fonction uniquement vers des

zones d'intérêt. Le métamodèle est mis à jour pour être suffisamment précis autour des zones d'intérêt. Un maximum d'informations concernant la fonction est alors exploitable au cours de chaque itération, permettant ainsi de limiter le nombre d'évaluations de la fonction. Mais pour cela, le critère d'échantillonnage doit être défini en prenant en compte le type de problème posé.
Pour notre problème de recherche de cartes d'idéotypes, on a intérêt à définir un critère qui permet l'évaluation des meilleures variétés conditionnellement à l'environnement, plutôt qu'un critère permettant l'évaluation des meilleures variétés en un seul environnement. Cela revient non seulement à chercher un optimum sur une surface de réponse aux paramètres pour un environnement donné, mais aussi le lieu des optima, leur ensemble, quand l'environnement change. Cela conduit à proposer un plan d'expérience adaptatif de façon à obtenir la meilleure précision possible dans ce lieu des optima. Jusqu'ici aucun critère n'existe à notre connaissance pour l'optimisation conditionnelle d'une fonction bruitée. Les travaux présentés dans ce manuscrit de thèse mènent à la proposition d'un nouveau critère d'échantillonnage $PEQI$, donc à la présentation d'un nouvel algorithme d'optimisation à base de krigeage. Une preuve de convergence de cet algorithme est également donnée. Ce nouvel algorithme permet la résolution d'un problème d'optimisation conditionnelle d'une fonction bruitée. Une étude des performances de cet algorithme est faite sur des fonctions test. Ce critère aboutit à l'estimation d'une ligne de crête avec un budget de simulation limité. Une comparaison du nouveau critère $PEQI$ avec un critère existant PEI, menée sur deux fonctions tests pour différents niveaux de bruits, différents noyaux de covariance et différentes tailles de plan d'expérience initial, montre que le critère $PEQI$ est toujours au moins aussi bien que le critère PEI, pour des niveaux de bruits faibles et le surpasse pour des niveaux de bruits plus élevés. Le critère $PEQI$ est aussi moins sensible que PEI au choix de la fonction de covariance et de la taille du plan d'expérience initial.
Dans cette thèse, nous avons supposé que la variance des bruits de simulation était homogène, ce qui n'est pas toujours le cas en pratique. Une application du critère pour l'optimisation conditionnelle de fonction bruitée avec une variance des bruits hétérogène est envisagée.
Une carte d'idéotypes par modèle a été produite en optimisant le rendement moyen en fonction de 4 paramètres du modèle Samara. D'autres options peuvent être envisagées pour le choix de ces paramètres, ainsi que pour la variable à optimiser : suivant les objectifs des paysans et leur aversion au risque, un quantile du rendement donnera un meilleur compromis entre son espérance et sa variabilité. Parmi les caractéristiques des plantes reflétées par les paramètres des modèles, certaines sont plus faciles que d'autres à améliorer génétiquement, qu'elles soient moins chères à mesurer ou plus héritables. Une évolution d'une algorithme d'optimisation serait la prise en compte du coût de l'amélioration, soit sous forme de contrainte, soit en faisant intervenir le coût dans une optimisation multicritère.

VII

Annexes

7.1 Quelques notions de topologie

7.1.1 Espace de Hilbert à noyau reproduisant

Rappelons qu'un produit scalaire sur un espace vectoriel H sur $\mathbb{R}$ est une application $(f, g) \rightarrow < f, g >_H$ de H^2 dans $\mathbb{R}$ qui est bilinéaire, symétrique et définie positive (c'est à dire pour tout $f \in H \backslash \{0\}$).
Un espace vectoriel muni d'un produit scalaire est appelé pré-hilbertien. Il est muni d'une norme associée au produit scalaire définie par $\|f\|_H =< f, f >_H^{1/2}$.
Un espace de Hilbert est un espace pré-hilbertien complet pour la norme associée.
Soit H un espace de Hilbert de fonctions $f : \mathbb{R}^d \rightarrow \mathbb{R}$. H est un espace de Hilbert à noyau reproduisant s'il existe une fonction $k_H : \mathbb{R}^d \rightarrow \mathbb{R}$ telle que $\forall f \in H$, $\forall x \in \mathbb{R}^d$, $f(x) =< k_H(., x), f >_H$.
k_H est alors défini de façon unique et est appelé noyau reproduisant de H. On peut remarquer que l'on a la propriété suivante (propriété reproduisante): $\forall x, y \in \mathbb{R}^d$, $< k_H(., x), k_H(., y) >_H = k_H(x, y)$.

7.1.2 Point adhérent

Soient E espace métrique, A une partie non vide de E, et y un élément de E. On dit que le point y est adhérent à A dans E si et seulement si il existe une suite $(u_n)_{n \in \mathbb{N}}$ de point de A qui converge vers y.

7.1.3 Partie dense

Soit E un espace vectoriel normé et A une partie de E. On dit que A est dense dans E si l'une des conditions équivalentes suivante est vérifiée :

- pour tout point $x \in E$, il existe une suite (y_n) d'éléments de A qui converge vers x
- pour tout x de E, pour tout $\epsilon > 0$, il existe $y \in A$ tel que $\|y - x\| \leq \epsilon$
- l'adhérence $\bar{A}$ de A est égale à E

Ces définitions peuvent être généralisées au cas des espaces métriques (en remplaçant la norme par la distance), et la dernière peut même l'être aux espaces topologiques généraux.

7.1.4 Suite extraite ou sous suite

Une suite extraite (ou sous suite) d'une suite donnée est une suite obtenue en sélectionnant dans l'ordre, un sous ensemble infini de termes. On peut également dire que l'on obtient une sous suite en omettant une partie des termes de la suite principale.

7.1.5 Inégalité de Cauchy-Schwarz

Soit H un espace vectoriel (réel ou complexe) muni du produit scalaire $< ., . >$. Alors pour tous $x, y \in H$

$$| < x, y > | \leq \|x| \, \|y\|$$

7.2 Quantile de krigeage

7.2.1 Propriétés du quantile de krigeage

Soit x_{n+1} le point à évaluer à l'étape $(n+1)$, τ^2_{n+1} et $\tilde{Y}_{n+1} = Y(x_{n+1}) + \epsilon_{n+1}$ respectivement la variance du bruit et la réponse bruitée. Nous allons étudier les propriétés de la moyenne et de la variance de krigeage à l'étape $n+1$ vue de l'étape n.
Soit $M_{n+1} = E(Y(x)|\tilde{A}_n, \tilde{Y}_{n+1})$ la fonction moyenne de krigeage à l'étape $n+1$ et $S^2_{n+1} = Var(Y(x)|\tilde{A}_n, \tilde{Y}_{n+1})$ la variance conditionnelle correspondante. Vu de l'étape n, la moyenne, de même que la variance, sont des processus aléatoires car elles dépendent de la mesure $\tilde{Y}_{n+1}$ non encore observée. Nous allons maintenant montrer qu'elles sont en fait des processus gaussiens, de même que le quantile associé $Q_{n+1} = M_{n+1} + \Phi^{-1}(\beta) S_{n+1}(x)$. Ces résultats découlent du fait que le prédicteur de krigeage est

linéaire par rapport aux observations et que la variance de krigeage est indépendante aux observations [106].

$$M_{n+1} = \sum_{j=1}^{n} \tilde{Y}_j + \lambda_{n+1,n+1}(x)\left(Y(x_{n+1}) + \epsilon_{n+1}\right) \tag{7.1}$$

où

$$\lambda_{n+1,n+1}(x) = \left(K_{n+1}(x)^T + \frac{1 - k_{n+1}(x)^T(K_{n+1} + \Gamma_{n+1})^{-1}\mathbf{1}_{n+1}}{\mathbf{1}_{n+1}^T(K_{n+1} + \Gamma_{n+1})^{-1}\mathbf{1}_{n+1}}\mathbf{1}_{n+1}^{-1}\right) \times (K_{n+1} + \Gamma)^{-1}$$

Il apparait clairement que conditionnellement à $\tilde{A}_n = \{\tilde{Y}_1, \cdots, \tilde{Y}_n\}$, M_{n+1} est un processus gaussien avec une espérance et une covariance données par:

$$E\left[M_{n+1}(x)|\tilde{A}_n\right] = \sum_{j=1}^{n} \lambda_{n+1,j}(x)\tilde{y}_j + \lambda_{n+1,n+1}(x)m_n(x) \tag{7.2}$$

et

$$cov\left[M_{n+1}(x), M_{n+1}(x')|\tilde{A}_n\right] = \lambda_{n+1,n+1}(x)\lambda_{n+1,n+1}(x')\left(S_n^2(x_{n+1}) + \tau_{n+1}^2\right). \tag{7.3}$$

En utilisant le fait que $Q_{n+1} = M_{n+1} + \Phi^{-1}(\beta)S_{n+1}(x)$, nous pouvons aussi conclure que conditionnellement à $\tilde{A}_n$, $Q_{n+1}(.)$ est un processus gaussien, étant la somme d'un processus gaussien et d'un processus déterministe. Nous avons:

$$E\left[Q_{n+1}(x)|\tilde{A}_n\right] = \sum_{j=1}^{n} \lambda_{n+1,j}(x)\tilde{y}_j + \lambda_{n+1,n+1}(x)m_n(x) + \Phi^{-1}(\beta)s_{n+1}(x) \tag{7.4}$$

et

$$cov\left[Q_{n+1}(x), Q_{n+1}(x')|\tilde{A}_n\right] = \lambda_{n+1,n+1}(x)\lambda_{n+1,n+1}(x')\left(s_n^2(x_{n+1}) + \tau_{n+1}^2\right) \tag{7.5}$$

$$m_{Q_{n+1}} = E\left[Q_{n+1}(x)|\tilde{A}_n\right] \tag{7.6}$$

$$\begin{aligned} s_{Q_{n+1}}^2 &= var\left[Q_{n+1}(x)|\tilde{A}_n\right] && (7.7)\\ &= \lambda_{n+1,n+1}(x)^2\left(s_n^2(x_{n+1}) + \tau_{n+1}^2\right) && (7.8)\end{aligned}$$

7.2.2 Mise à jour du quantile de krigeage par l'ajout d'un point

L'écriture de $m_{Q_{n+1}}$ et $s_{Q_{n+1}}^2$ en fonction des valeurs du krigeage à l'étape n est donnée ci-dessous. D'après Emery *et al.* [40], la moyenne, la variance et les poids du krigeage sont mis à jour avec une mesure additionnelle $\tilde{y}_{n+1}$ réponse de x_{n+1} avec une variance τ_{n+1}^2.

$$m_{n+1}(x) = m_n(x) + \frac{c_n(x, x_{n+1})}{s_n^2(x) + \tau_{n+1}^2}\left(\tilde{y}_{n+1} - m_n(x)\right) \tag{7.9}$$

$$s_{n+1}^2 = s_n^2(x) - \frac{c_n(x, x_{n+1})^2}{s_n^2(x) + \tau_{n+1}^2} \tag{7.10}$$

où c_n est la covariance du krigeage qui est égale (pour le krigeage ordinaire) à:

$$c_n(x, x^{'}) = k(x, x^{'}) - k(x)^T K^{-1} k(x^{'}) + \frac{(1 - \mathbf{1}_n K^{-1} k(x^{'}))(1 - k(x) K^{-1} \mathbf{1}_n)}{\mathbf{1}_n K^{-1} \mathbf{1}_n} \tag{7.11}$$

Notons que $c_n(x, x) = s_n^2(x)$ et que $\tilde{y}_{n+1} = m_n(x_{n+1})$. Cela implique que $m_{n+1}(x) = m_n(x)$ et

$$m_{Q_{n+1}} = m_n(x_{n+1}) + \Phi^{-1}(\beta) \sqrt{\frac{\tau_{n+1}^2 s_n^2(x_{n+1})}{\tau_{n+1}^2 + s_n^2(x_{n+1})}} \tag{7.12}$$

$$s_{Q_{n+1}}^2 = \frac{[s_n^2(x_{n+1})]^2}{\tau_{n+1}^2 + s_n^2(x_{n+1})} \tag{7.13}$$

VIII

Annexes

8.1 Propriétés du quantile de krigeage

Soit x_{n+1} le point à évaluer à l'étape $(n+1)$, τ^2_{n+1} et $\tilde{Y}_{n+1} = Y(x_{n+1}) + \epsilon_{n+1}$ respectivement la variance du bruit et la réponse bruitée. Nous allons étudier les propriétés de la moyenne et de la variance de krigeage à l'étape $n+1$ vue de l'étape n.
Soit $M_{n+1} = E(Y(x)|\tilde{A}_n, \tilde{Y}_{n+1})$ la fonction moyenne de krigeage à l'étape $n+1$ et $S^2_{n+1} = Var(Y(x)|\tilde{A}_n, \tilde{Y}_{n+1})$ la variance conditionnelle correspondante. Vu de l'étape n, la moyenne, de même que la variance, sont des processus aléatoires car elles dépendent de la mesure $\tilde{Y}_{n+1}$ non encore observée. Nous allons maintenant montrer qu'elles sont en fait des processus gaussiens, de même que le quantile associé $Q_{n+1} = M_{n+1} + \Phi^{-1}(\beta)S_{n+1}(x)$. Ces résultats découlent du fait que le prédicteur de krigeage est linéaire par rapport aux observations et que la variance de krigeage est indépendante aux observations [106].

$$M_{n+1} = \sum_{j=1}^{n} \tilde{Y}_j + \lambda_{n+1,n+1}(x)\left(Y(x_{n+1}) + \epsilon_{n+1}\right) \tag{8.1}$$

où

$$\lambda_{n+1,n+1}(x) = \left(K_{n+1}(x)^T + \frac{1 - k_{n+1}(x)^T(K_{n+1} + \Gamma_{n+1})^{-1}\mathbf{1}_{n+1}}{\mathbf{1}^T_{n+1}(K_{n+1} + \Gamma_{n+1})^{-1}\mathbf{1}_{n+1}}\mathbf{1}^{-1}_{n+1}\right) \times (K_{n+1} + \Gamma)^{-1}$$

Il apparait clairement que conditionnellement à $\tilde{A}_n = \{\tilde{Y}_1, \cdots, \tilde{Y}_n\}$, M_{n+1} est un processus gaussien avec une espérance et une covariance données par:

$$E\left[M_{n+1}(x)|\tilde{A}_n\right] = \sum_{j=1}^{n} \lambda_{n+1,j}(x)\tilde{y}_j + \lambda_{n+1,n+1}(x)m_n(x) \tag{8.2}$$

et

$$cov\left[M_{n+1}(x), M_{n+1}(x')|\tilde{A}_n\right] = \lambda_{n+1,n+1}(x)\lambda_{n+1,n+1}(x')\left(S^2_n(x_{n+1}) + \tau^2_{n+1}\right). \tag{8.3}$$

En utilisant le fait que $Q_{n+1} = M_{n+1} + \Phi^{-1}(\beta)S_{n+1}(x)$, nous pouvons aussi conclure que conditionnellement à $\tilde{A}_n$, $Q_{n+1}(.)$ est un processus gaussien, étant la somme d'un processus gaussien et d'un processus déterministe. Nous avons:

$$E\left[Q_{n+1}(x)|\tilde{A}_n\right] = \sum_{j=1}^{n} \lambda_{n+1,j}(x)\tilde{y}_j + \lambda_{n+1,n+1}(x)m_n(x) + \Phi^{-1}(\beta)s_{n+1}(x) \tag{8.4}$$

et

$$cov\left[Q_{n+1}(x), Q_{n+1}(x^{'})|\tilde{A}_n\right] = \lambda_{n+1,n+1}(x)\lambda_{n+1,n+1}(x^{'})\left(s_n^2(x_{n+1}) + \tau_{n+1}^2\right) \tag{8.5}$$

$$m_{Q_{n+1}} = E\left[Q_{n+1}(x)|\tilde{A}_n\right] \tag{8.6}$$

$$\begin{aligned} s_{Q_{n+1}}^2 &= var\left[Q_{n+1}(x)|\tilde{A}_n\right] && (8.7)\\ &= \lambda_{n+1,n+1}(x)^2\left(s_n^2(x_{n+1}) + \tau_{n+1}^2\right) && (8.8)\end{aligned}$$

8.2 Mise à jour du quantile de krigeage par l'ajout d'un point

L'écriture de $m_{Q_{n+1}}$ et $s_{Q_{n+1}}^2$ en fonction des valeurs du krigeage à l'étape n est donnée ci-dessous. D'après Emery *et al.* [40], la moyenne, la variance et les poids du krigeage sont mis à jour avec une mesure additionnelle $\tilde{y}_{n+1}$ réponse de x_{n+1} avec une variance τ_{n+1}^2.

$$m_{n+1}(x) = m_n(x) + \frac{c_n(x, x_{n+1})}{s_n^2(x) + \tau_{n+1}^2}\left(\tilde{y}_{n+1} - m_n(x)\right) \tag{8.9}$$

$$s_{n+1}^2 = s_n^2(x) - \frac{c_n(x, x_{n+1})^2}{s_n^2(x) + \tau_{n+1}^2} \tag{8.10}$$

où c_n est la covariance du krigeage qui est égale (pour le krigeage ordinaire) à:

$$c_n(x, x^{'}) = k(x, x^{'}) - k(x)^T K^{-1} k(x^{'}) + \frac{(1 - \mathbf{1}_n K^{-1} k(x^{'}))(1 - k(x)K^{-1}\mathbf{1}_n)}{\mathbf{1}_n K^{-1}\mathbf{1}_n} \tag{8.11}$$

Notons que $c_n(x, x) = s_n^2(x)$ et que $\tilde{y}_{n+1} = m_n(x_{n+1})$. Cela implique que $m_{n+1}(x) = m_n(x)$ et

$$m_{Q_{n+1}} = m_n(x_{n+1}) + \Phi^{-1}(\beta)\sqrt{\frac{\tau_{n+1}^2 s_n^2(x_{n+1})}{\tau_{n+1}^2 + s_n^2(x_{n+1})}} \tag{8.12}$$

$$s_{Q_{n+1}}^2 = \frac{[s_n^2(x_{n+1})]^2}{\tau_{n+1}^2 + s_n^2(x_{n+1})} \tag{8.13}$$

Bibliographie

[1] S Asseng, H van Keulen, and W Stol. Performance and application of the APSIM Nwheat model in the Netherlands. *European Journal of Agronomy*, 12(1):37–54, jan 2000. 3

[2] Maud Balme, Sylvie Galle, and Thierry Lebel. Démarrage de la saison des pluies au Sahel : variabilité aux échelles hydrologique et agronomique, analysée à partir des données EPSAT-Niger. *Science et changements planétaires/Sécheresse*, 16(1):15–22, 2005. 85

[3] Jean baptiste Denis. Two way analysis using covarites1. *Statistics*, 19(1):123–132, jan 1988. 12, 14

[4] Pierre Barbillon, Gilles Celeux, Agnès Grimaud, Yannick Lefebvre, and Étienne De Rocquigny. Nonlinear methods for inverse statistical problems. *Computational Statistics & Data Analysis*, 55(1):132–142, jan 2011. 4

[5] Aude Barbottin. Utilisation d'un modèle de culture pour évaluer le comportement des génotypes : Pertinence de l'utilisation d'Azodyn pour analyser la variabilité du rendement et de la teneur en protéines du blé tendre Jury. *PhD Thesis*, page 178, 2004. 15

[6] C. P. Baril. Factor regression for interpreting genotype-environment interaction in bread-wheat trials. *Theoretical and Applied Genetics*, 83(8):1022–1026, 1992. 15

[7] Russell R. Barton and Martin Meckesheimer. Chapter 18 Metamodel-Based Simulation Optimization. *Handbooks in Operations Research and Management Science*, 13:535–574, jan 2006. 33

[8] Hans-Georg Beyer and Bernhard Sendhoff. Robust optimization – A comprehensive survey. *Computer Methods in Applied Mechanics and Engineering*, 196(33-34):3190–3218, 2007. 34

[9] R Bonhomme. Modélisation du fonctionnement d'une culture: caractérisation de la contrainte hydrique et prise en compte de ses effets. *L'eau dans l'espace rural. Riou, C., Bonhomme, R., Chassin, P., Neveu, A., Papy, F., Paris*, 1997. 16

[10] M H A Bonte, A H van den Boogaard, and J Huétink. An optimisation strategy for industrial metal forming processes. *Structural and Multidisciplinary Optimization*, 35(6):571–586, mar 2008. 4

[11] M. H.A. Bonte, A. H. Van Den Boogaard, and J. Huétink. A metamodel based optimisation algorithm for metal forming processes. In *Advanced Methods in Material Forming: With 264 Figures and 37 Tables*, pages 55–72. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. 33

[12] Kenneth J. Boote, James W. Jones, and Nigel B. Pickering. Potential Uses and Limitations of Crop Models. *Agronomy Journal*, 88(5):704, 1996. 16

[13] M Brancourt-Hulmel, C Lecomte, M Leleu, P Bérard, N Galic, B Trouvé, and C Sausseau. Sélection et stabilité du rendement chez le blé tendre d'hiver. *Agronomie*, 14(9):611–625, 1994. 15

[14] F H Branin and S K Hoo. A method for finding multiple extrema of a function of n variables. *Numerical Methods*, pages 231–237, 1972. 59

[15] N Brisson, C Gary, E Justes, R Roche, B Mary, D Ripoche, D Zimmer, J Sierra, P Bertuzzi, P Burger, F Bussière, Y.M Cabidoche, P Cellier, P Debaeke, J.P Gaudillère, C Hénault, F Maraux, B Seguin, and H Sinoquet. An overview of the crop model stics. *European Journal of Agronomy*, 18(3-4):309–332, jan 2003. 21

[16] N Brisson, F Ruget, P Gate, J Lorgeou, B Nicoullaud, X Tayot, D Plenet, M H Jeuffroy, A Bouthier, D Ripoche, B Mary, and E Justes. STICS: a generic model for the simulation of crops and their water and nitrogen balances. II. Model validation for wheat and maize. *Agronomie*, 22(5-6):69–92, 2002. 16

[17] Nadine Brisson, Bruno Mary, Dominique Ripoche, Marie Hélène Jeuffroy, Françoise Ruget, Bernard Nicoullaud, Philippe Gate, Florence Devienne-Barret, Rodrigo Antonioletti, Carolyne Durr, Guy Richard, Nicolas Beaudoin, Sylvie Recous, Xavier Tayot, Daniel Plenet, Pierre Cellier, Jean-Marie Machet, Jean Marc Meynard, and Richard Delécolle. STICS: a generic model for the simulation of crops and their water and nitrogen balances. I. Theory and parameterization applied to wheat and corn. *Agronomie*, 18(5-6):311–346, 1998. 16

[18] C. G. BROYDEN. The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations. *IMA Journal of Applied Mathematics*, 6(1):76–90, mar 1970. 28

[19] Francois Brun, Daniel Wallach, David Makowski, and James W Jones. *Working with dynamic crop models: evaluation, analysis, parameterization, and applications*. Elsevier, 2006. 15, 17, 19

[20] R Carnell. lhs: Latin Hypercube Samples, available at: h ttp. *CRAN. R-project. org/package= lhs, r package version 0.5, last access*, 15, 2011. 60

[21] Pierre Casadebaig. *Analyse et modélisation de l'interaction génotype - environnement - conduite de culture : application au tournesol (Helianthus annuus L.)*. PhD thesis, Université Toulouse, apr. 15

[22] Pierre Casadebaig, Philippe Debaeke, and Jérémie Lecoeur. Thresholds for leaf expansion and transpiration response to soil water deficit in a range of sunflower genotypes. *European Journal of Agronomy*, 28(4):646–654, may 2008. 16

[23] Augustine-Louis Cauchy. Méthode générale pour la résolution des systèmes d'équations simultanées [Translated: (2010)]. *Compte Rendu des S'eances de L'Acad'emie des Sciences*, 25(2):536–538, 1847. 27

[24] W. G. Cochran and F. Yates. The analysis of groups of experiments. *The Journal of Agricultural Science*, 28(04):556–580, oct 1938. 13

[25] Dennis D Cox and Susan John. SDO: A statistical method for global optimization. *Multidisciplinary design optimization: state of the art*, pages 315–329, 1997. 40

[26] Noel Cressie. *Statistics for Spatial Data*. 2015. 35, 36

[27] Noel A. C. Cressie and Christopher K. Wikle. *Statistics for spatio-temporal data*. Wiley, 2011. 37

[28] J. Crossa, H. G. Gauch, and R. W. Zobel. Additive Main Effects and Multiplicative Interaction Analysis of Two International Maize Cultivar Trials. *Crop Science*, 30(3):493, 1990. 12, 13

[29] C Darwin. On the origin of species by means of natural selection. 1859. *Murray*, 1968. 32

[30] William C. Davidon. Variable Metric Method for Minimization. *SIAM Journal on Optimization*, 1(1):1–17, feb 1991. 29

[31] Jean-Baptiste Denis and Patrick Vincourt. Panorama des méthodes statistiques d'analyse des interactions génotype x milieu. *Agronomie*, 2:219–230, 1982. 12, 14

[32] Ibnou Dieng. *Prédiction de l'interaction génotype x environnement par linéarisation et régression PLS-mixte*. PhD thesis, Montpellier 2, jan 2007. 15

[33] M Dingkuhn, J C Soulié, and T Lafarge. Samara V2: A cereal crop model to study G x E x M interaction and phenotypic plasticity, and explore ideotypes. In *AgMIP Rice International Workshop*, pages 28–30, 2011. 4, 70, 71

[34] C M Donald. The design of a wheat ideotype. *Finlay, KW and Shepherd*, 1968. 3, 20, 39

[35] M. Dorigo, V. Maniezzo, and A. Colorni. Ant system: optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, 26(1):29–41, 1996. 33

[36] Olivier Dubrule. Cross validation of kriging in a unique neighborhood. *Journal of the International Association for Mathematical Geology*, 15(6):687–699, dec 1983. 38

[37] David E. Goldberg. Genetic Algorithms In Search, Optimization, and Machine Learning. *Addison-Wesley Professional Reading*, 1989. 32

[38] S. a. Eberhart and W. a. Russell. Stability Parameters for Comparing Varieties1. *Crop Science*, 6(1):36, 1966. 13

[39] Bryan Eisenhower, Zheng O'Neill, Satish Narayanan, Vladimir A. Fonoberov, and Igor Mezić. A methodology for meta-model based optimization in building energy models. *Energy and Buildings*, 47:292–301, apr 2012. 33

[40] Xavier Emery. The kriging update equations and their application to the selection of neighboring data. *Computational Geosciences*, 13(3):269–280, sep 2009. 99, 102

[41] K. W. Finlay and G. N. Wilkinson. The analysis of adaptation in a plant-breeding programme. *Australian Journal of Agricultural Research*, 14(6):742–754, 1963. 12, 13

[42] R Fletcher. Practical methods of optimization john wiley & sons. *New York*, 80, 1987. 28

[43] R. Fletcher and M. J. D. Powell. A Rapidly Convergent Descent Method for Minimization. *The Computer Journal*, 6(2):163–168, aug 1963. 29

[44] R. Fletcher and C. M. Reeves. Function minimization by conjugate gradients. *The Computer Journal*, 7(2):149–154, feb 1964. 28

[45] Lawrence Jerome Fogel. Autonomous Automata. *Industrial Research*, 4(2):14–19, 1962. 32

[46] P. N. Fox, A. A. Rosielle, and W. J.R. Boyd. The nature of genotype × environment interactions for wheat yield in Western Australia. *Field Crops Research*, 11(C):387–398, jan 1985. 13

[47] Lennart Frimannslund and Trond Steihaug. A generating set search method using curvature information. *Computational Optimization and Applications*, 38(1):105–121, jul 2007. 30

[48] K R Gabriel. Least Squares Approximation of Matrices by Additive and Multiplicative Models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 40(2):186–196, 1978. 14

[49] Mathias Neumann Gabrielle, B and Denoroy, P and Gosse, G and Justes, E and Andersen. Development and evaluation of a CERES-type model for winter oilseed rape. *Field Crops Research*, 57(1):95–111, may 1998. 16

[50] Hugh G. Gauch and Jr. Model Selection and Validation for Yield Trials with Interaction. *Biometrics*, 44(3):705, sep 1988. 12, 13

[51] Jean Charles Gilbert and Jorge Nocedal. Global Convergence Properties of Conjugate Gradient Methods for Optimization. *SIAM Journal on Optimization*, 2(1):21–42, feb 1992. 28

[52] PE Gill, W Murray, and MH Wright. Practical optimization. *Academic press*, 1981. 28

[53] D. Ginsbourger, V. Picheny, Olivier Roustant, and Y. Richet. A new look at Kriging for the Approximation of Noisy Simulators with Tunable Fidelity. *8th annual conference of ENBIS*, pages –, sep 2008. 38

[54] David Ginsbourger. *Multiples métamodèles pour l'approximation et l'optimisation de fonctions numériques multivariables*. PhD thesis, Ecole Nationale Supérieure des Mines de Saint-Etienne, jan 2009. 4

[55] David Ginsbourger, Jean Baccou, Clément Chevalier, Frédéric Perales, Nicolas Garland, and Yann Monerie. Bayesian Adaptive Reconstruction of Profile Optima and Optimizers. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):490–510, 2014. 4, 40, 45, 46, 52, 60

[56] Fred Glover. Future paths for integer programming and links to artificial intelligence. *Computers and Operations Research*, 13(5):533–549, jan 1986. 33

[57] Donald Goldfarb. A family of variable-metric methods derived by variational means. *Mathematics of Computation*, 24(109):23–23, jan 1970. 28

[58] Harry F. Gollob. A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika*, 33(1):73–115, mar 1968. 12, 13

[59] Eric Gozé. *Modèle Stochastique de la Pluviométrie au Sahel, Application à l'Agronomie*. PhD thesis, Montpellier 2, 1990. 22

[60] J. Gu, G. Y. Li, and Z. Dong. Hybrid and Adaptive Metamodel Based Global Optimization. In *Volume 5: 35th Design Automation Conference, Parts A and B*, pages 751–765. ASME, jan 2009. 33

[61] G L Hammer, M J Kropff, T R Sinclair, and J R Porter. Future contributions of crop modelling - from heuristics and supporting decision making to understanding genetic regulation and aiding crop improvement. *European Journal of Agronomy*, 18(1):15–31, 2002. 16

[62] Yannick Hébert, Christophe Plomion, and Nathalie Harzic. Genotype x environment interaction for root traits in maize, as analysed with factorial regression models. *Euphytica*, 81(1):85–92, 1995. 15

[63] Nicolas Heslot, Deniz Akdemir, Mark E. Sorrells, and Jean-Luc Jannink. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theoretical and Applied Genetics*, 127(2):463–480, feb 2014. 15

[64] M.R. Hestenes and Eduard Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49(6):409, 1952. 28

[65] John H. Holland. Outline for a Logical Theory of Adaptive Systems. *Journal of the ACM*, 9(3):297–314, jul 1962. 32

[66] Robert Hooke and T. A. Jeeves. " Direct Search" Solution of Numerical and Statistical Problems. *Journal of the ACM*, 8(2):212–229, apr 1961. 4, 30

[67] Deng Huang, Theodore T Allen, William I Notz, and Ning Zeng. Global optimization of stochastic black-box systems via sequential kriging meta-models. *Journal of global optimization*, 34(3):441–466, 2006. 40, 46

[68] Egea J. A. Larrosa. New heuristics for global optimization of complex bioprocesses. *PhD Thesis, Universidade de Vigo*, 2008. 34

[69] Alexander I. J. Forrester, Andy J. Keane, and Neil W. Bressloff. Design and Analysis of "Noisy" Computer Experiments. *AIAA Journal*, 44(10):2331–2339, oct 2006. 40

[70] T. Jansson, A. Andersson, and L. Nilsson. Optimization of draw-in for an automotive sheet metal part: An evaluation using surrogate models and response surfaces. *Journal of Materials Processing Technology*, 159(3):426–434, feb 2005. 33

[71] Janis Janusevskis and Rodolphe Le Riche. Simultaneous kriging-based estimation and optimization of mean response. *Journal of Global Optimization*, 55(2):313–336, jan 2012. 4

[72] S Jeuffroy, M-H and Recous. Azodyn: a simple model simulating the date of nitrogen deficiency for decision support in wheat fertilization. *European Journal of Agronomy*, 10(2):129–144, mar 1999. 16

[73] Gregory L. Johnson, Clayton L. Hanson, Stuart P. Hardegree, and Edward B. Ballard. Stochastic Weather Simulation: Overview and Analysis of Two Commonly Used Models. *Journal of Applied Meteorology*, 35(10):1878–1896, oct 1996. 22

[74] Mark E Johnson, Leslie M Moore, and Donald Ylvisaker. Minimax and maximin distance designs. *Journal of statistical planning and inference*, 26(2):131–148, 1990. 39

[75] Donald R. Jones. A Taxonomy of Global Optimization Methods Based on Response Surfaces. *Journal of Global Optimization*, 21(4):345–383, 2001. 4, 33, 41

[76] Donald R. Jones, Matthias Schonlau, and William J. Welch. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*, 13(4):455–492, 1998. 40, 42, 45, 60

[77] P. G. Jones. Current availability and deficiencies data relevant to agro-ecological studies in the geographical area covered in IARCS. In *Workshop on Agro-Ecological Characterization, Classification and Mapping, Rome (Italy), 14-18 Apr 1986*, pages 69–82. CAB International, 1987. 71

[78] P G Jones. MarkSim_standalone dor DSSAT users. (March):8, 2012. 72

[79] P. G. Jones and P. K. Thornton. A rainfall generator for agricultural applications in the tropics. *Agricultural and Forest Meteorology*, 63(1-2):1–19, feb 1993. 70, 71

[80] P. G. Jones and P. K. Thornton. Spatial and temporal variability of rainfall related to a third-order Markov model. *Agricultural and Forest Meteorology*, 86(1-2):127–138, aug 1997. 71

[81] P. G. Jones and P. K. Thornton. Fitting a third-order Markov rainfall model to interpolated climate surfaces. *Agricultural and Forest Meteorology*, 97(3):213–231, nov 1999. 71

[82] Peter G. Jones and Philip K. Thornton. MarkSim: Software to generate daily weather data for Latin America and Africa. *Agronomy Journal*, 92(3):445–453, 2000. 71

[83] Peter G. Jones and Philip K. Thornton. Generating downscaled weather data from a suite of climate models for agricultural modelling applications. *Agricultural Systems*, 114:1–5, jan 2013. 71

[84] A. G. Journel. Nonparametric estimation of spatial distributions. *Journal of the International Association for Mathematical Geology*, 15(3):445–468, jun 1983. 94

[85] R Kennedy. J. and Eberhart, Particle swarm optimization. In *Proceedings of IEEE International Conference on Neural Networks IV, pages*, volume 1000, 1995. 33

[86] J. Kiefer and J. Wolfowitz. Stochastic Estimation of the Maximum of a Regression Function. 29

[87] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by Simulated Annealing. *Science*, 220(4598):671–680, 1983. 32

[88] Scott Kirkpatrick. Optimization by simulated annealing: Quantitative studies. *Journal of Statistical Physics*, 34(5-6):975–986, mar 1984. 32

[89] D G Krige. *A statistical approach to some mine valuation and allied problems on the Witwatersrand: By DG Krige*. PhD thesis, University of the Witwatersrand, 1951. 35

[90] H. J. Kushner. A New Method of Locating the Maximum Point of an Arbitrary Multipeak Curve in the Presence of Noise. *Journal of Basic Engineering*, 86(1):97, mar 1964. 41

[91] Vanesse Labeyrie, Mathieu Thomas, Zachary K. Muthamia, and Christian Leclerc. Seed exchange networks, ethnicity, and sorghum diversity. *Proceedings of the National Academy of Sciences*, 113(1):98–103, jan 2016. 2

[92] J. J. Landsberg and C. T. de Wit. Simulation of Assimilation, Respiration and Transpiration of Crops. *The Journal of Ecology*, 68(1):333, 1980. 3

[93] Rodolphe Le Riche, Victor Picheny, Andre Meyer, Nam-Ho Kim, and David Ginsbourger. Gears Design with Shape Uncertainties Using Controlled Monte Carlo Simulations and Kriging. In *50th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Reston, Virigina, may 2009. American Institute of Aeronautics and Astronautics. 33

[94] Ch Loyce, J P Rellier, and J M Meynard. Management planning for winter wheat with multiple objectives (1): the BETHA system. *Agricultural Systems*, 72(1):9–31, 2002. 16

[95] J M Machet, P Dubrulle, and P Louis. AZOBIL: a computer program for fertilizer N recommendations based on a predictive balance sheet method. In *Proceedings of the 1st Congress of the European Society of Agronomy, Paris, France*, pages 2–21, 1990. 16

[96] K. V. Mardia and R. J. Marshall. Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71(1):135–146, apr 1984. 38

[97] N. C. Matalas. Mathematical assessment of synthetic hydrology. *Water Resources Research*, 3(4):937–945, dec 1967. 72

[98] Georges Matheron. Le krigeage universel. *Ecole nationale supérieure des mines de Paris*, 1969. 35

[99] M. D. McKay, R. J. Beckman, and W. J. Conover. Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, 21(2):239–245, may 1979. 37

[100] J Mockus, V Tiesis, and A Zilinskas. Toward Global Optimization, volume 2, chapter Bayesian Methods for Seeking the Extremum. pages 117–128, 1978. 42

[101] J. A. Nelder and R. Mead. A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308–313, jan 1965. 4, 30, 31

[102] Harald Niederreiter. *Random number generation and quasi-Monte Carlo methods*. SIAM, 1992. 37

[103] F A O OCDE. Perspectives agricoles de l'OCDE et de la FAO 2014-2023. *Organisation de Coopération et de Développement Economiques, Paris*, 2014. 1

[104] V Picheny, D Ginsbourger, and Y Richet. Noisy Expected Improvement and On-line Computation Time Allocation for the Optimization of Simulators with Tunable Fidelity. *2nd International Conference on Engineering Optimization*, pages 1–10, jun 2010. 62

[105] Victor Picheny and David Ginsbourger. Noisy kriging-based optimization methods: A unified implementation within the DiceOptim package. *Computational Statistics and Data Analysis*, 71:1035–1053, 2014. 46

[106] Victor Picheny, David Ginsbourger, Yann Richet, and Gregory Caplin. Quantile-Based Optimization of Noisy Computer Experiments With Tunable Precision. *Technometrics*, 55(1):2–13, feb 2013. 38, 40, 46, 47, 51, 54, 58, 99, 101

[107] Victor Picheny, Tobias Wagner, and David Ginsbourger. A benchmark of kriging-based infill criteria for noisy optimization. *Structural and Multidisciplinary Optimization*, 48(3):607–626, apr 2013. 46

[108] Moshe J. Pinthus. Estimate of genotypic value: A proposed method. *Euphytica*, 22(1):121–123, mar 1973. 13

[109] M. J. D. Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The Computer Journal*, 7(2):155–162, feb 1964. 30

[110] M. J. D. Powell. Direct search algorithms for optimization calculations. *Acta Numerica*, 7:287, jan 1998. 30

[111] Nestor V. Queipo, Raphael T. Haftka, Wei Shyy, Tushar Goel, Rajkumar Vaidyanathan, and P. Kevin Tucker. Surrogate-based analysis and optimization. *Progress in Aerospace Sciences*, 41(1):1–28, jan 2005. 4

[112] P. Racsko, L. Szeidl, and M. Semenov. A serial approach to local stochastic weather models. *Ecological Modelling*, 57(1-2):27–41, oct 1991. 22

[113] C. W. Richardson. Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, 17(1):182–190, feb 1981. 22, 71

[114] C. W. Richardson and D. A. Wright. WGEN : A Model for Generating Daily Weather Variables. *U. S. Department of Agriculture, Agricultural Research Service*, ARS-8:83, 1984. 22

[115] R. Rincent, E. Kuhn, H. Monod, F.-X. Oury, M. Rousset, V. Allard, and J. Le Gouis. Optimization of multi-environment trials for genomic selection based on crop models. *Theoretical and Applied Genetics*, 130(8):1735–1752, aug 2017. 15

[116] J Rivoirard. Introduction au krigeage disjonctif et à la géostatistique non-linéaire. *Document de cours. Centre de géostatistique, Ecole des Mines de Paris*, 1991. 94

[117] Herbert Robbins and Sutton Monro. A Stochastic Approximation Method. *Institute of Mathematical Statistics*, 22(3):400–407, 1951. 29

[118] B Rolland, F X Oury, C Bouchard, and C Loyce. Vers une évolution de la création variétale pour répondre aux besoins de l'agriculture durable? L'exemple du blé tendre. *Dossier de l'Environnement de l'INRA*, (30):78–89, 2006. 10

[119] H. H. Rosenbrock. An Automatic Method for Finding the Greatest or Least Value of a Function. *The Computer Journal*, 3(3):175–184, mar 1960. 30, 59

[120] Olivier Roustant, David Ginsbourger, and Yves Deville. DiceKriging, DiceOptim: two R packages for the analysis of computer experiments by kriging-based metamodelling and optimization. *Journal of Statistical Software*, 51(1):54p, oct 2012. 60, 86

[121] Françoise Ruget, Nadine Brisson, Richard Delécolle, and Robert Faivre. Sensitivity analysis of a crop simulation model , STICS , in order to choose the main parameters to be estimated. 2002. 19

[122] J Sacks, WJ Welch, TJ Mitchell, and HP Wynn. Design and analysis of computer experiments. *Statistical science*, 1989. 39

[123] Jerome Sacks, William J Welch, Toby J Mitchell, and Henry P Wynn. Design and analysis of computer experiments. *Statistical science*, pages 409–423, 1989. 33, 37

[124] Larry L. Schumaker. *Spline Models for Observational Data (Grace Wahba)*, volume 33. Society for Industrial and Applied Mathematics, jan 1991. 33

[125] Warren Scott, Peter Frazier, and Warren Powell. The Correlated Knowledge Gradient for Simulation Optimization of Continuous Parameters using Gaussian Process Regression. *SIAM Journal on Optimization*, 21(3):996–1026, jul 2011. 40

[126] Mikhail A. Semenov and Elaine M. Barrow. Use of a stochastic weather generator in the development of climate change scenarios. *Climatic Change*, 35(4):397–414, 1997. 23

[127] D. F. Shanno. Conditioning of quasi-Newton methods for function minimization. *Mathematics of Computation*, 24(111):647–647, sep 1970. 28

[128] Timothy W Simpson, J D Poplinski, Patrick N Koch, and Janet K Allen. Metamodels for computer-based engineering design: survey and recommendations. *Engineering with computers*, 17(2):129–150, 2001. 34

[129] Afshin. Soltani and Thomas R. Sinclair. *Modeling Physiology of Crop Development, Growth and Yield*. CABI, 2012. 17

[130] Dong Song, Rosa H M Chan, Vasilis Z Marmarelis, Robert E Hampson, Sam A Deadwyler, and Theodore W Berger. Neural Networks. In *Neural Networks*, volume 22, pages 1340–1351. 2009. 33

[131] G.H. Stringfield and R.M. Salter. Differential response of corn varieties to fertility levels and to seasons. *Journal of Agricultural Research*, 49(11):991–1000, 1934. 13

[132] Jean-Denis Taupin and E. Bonef. Erreur d'estimation de la pluie à l'échelle décadaire sur des surfaces de 25 à 10 000 km2 en fonction de la densité du réseau sol en milieu sahélien : implication dans l'intercomparaison données sols - données satellitaires. *Coll. Colloques et Séminaires, Orstom*, pages 45–62, 1996. 22

[133] Clemens Tilke. *Introduction to Disjunctive Kriging and Non-Linear Geostatistics*, volume 25. Oxford, GB: Clarendon Press, 1997. 94

[134] Virginia Torczon. On the Convergence of the Multidirectional Search Algorithm. *SIAM Journal on Optimization*, 1(1):123–145, feb 1991. 30

[135] Virginia Torczon. On the Convergence of Pattern Search Algorithms. *SIAM Journal on Optimization*, 7(1):1–25, feb 1997. 30

[136] V Tourczon. Multi-directional search: a direct search algorithm for parallel machines. *PhD Thesis, Rice University*, 1989. 30

[137] F A Van Eeuwijk, J B Denis, and M S Kang. Incorporating additional information on genotypes and environments in models for two-way genotype by environment tables. *Genotype-by-environment interaction*, pages 15–50, 1996. 12, 14

[138] Fred A. van Eeuwijk. Multiplicative Interaction in Generalized Linear Models. *Biometrics*, 51(3):1017, sep 1995. 13

[139] Emmanuel Vazquez. Modélisation comportementale de systèmes non-linéaires multivariables par méthodes à noyaux et applications. *PhD Thesis, Université Paris Sud-Paris XI*, may 2005. 38

[140] Emmanuel Vazquez and Julien Bect. Convergence properties of the expected improvement algorithm with fixed mean and covariance functions. *Journal of Statistical Planning and Inference*, 140(11):3088–3095, nov 2010. 53, 54, 55

[141] W. Vent. Rechenberg, Ingo, Evolutionsstrategie - Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. 170 S. mit 36 Abb. Frommann-Holzboog-Verlag. Stuttgart 1973. Broschiert. *Feddes Repertorium*, 86(5):337–337, apr 2008. 32

[142] J. Voltas, F. A. Van Eeuwijk, J. L. Araus, and I. Romagosa. Integrating statistical and ecophysiological analyses of genotype by environment interaction for grain filling of barley II. Grain growth. *Field Crops Research*, 62(1):75–84, jun 1999. 15

[143] Daniel Wallach, Bruno Goffinet, Jacques-Eric Bergez, Philippe Debaeke, Delphine Leenhardt, and Jean-Noël Aubertot. Parameter Estimation for Crop Models. *Agronomy Journal*, 93(4):757, 2001. 19

[144] Frederick H Walters, Stephen L Morgan, Lloyd R Parker, and Jr Stanley N. *Sequential simplex optimization: a technique for improving quality and productivity in research, development, and manufacturing*. CRC Press, aug 2001. 30

[145] X. C. X. C. Zhang and J. D. J. D. Garbrecht. Evaluation of Cligen Precipitation Parameters and Their Implication on Wepp Runoff and Erosion Prediction. *Transactions of the ASAE*, 46(2):311, 2003. 22

[146] G. Udny Yule. Mendel'S Laws and Their Probable Relations To Intra-Racial Heredity. *New Phytologist*, 1(10):222–238, dec 1902. 2

Résumé : Au Sahel, la répartition des pluies irrégulière dans le temps et dans l'espace engendre une forte interaction variété $\times$ année et variété $\times$ lieu. Pour déterminer les variétés les plus productives en espérance en fonction des lieux, il faudrait de nombreuses années d'expérimentation en chaque lieu, ce qui prendrait beaucoup de temps. Une alternative est de maximiser la production prédite à l'aide d'un modèle de culture décrivant la croissance et le développement de cultures en interaction avec leurs conditions agro-environnementales.
La production à maximiser est moyenne sur une distribution de probabilité d'entrées environnementales, spécifique du lieu, alors que les paramètres variétaux qui maximisent cette production définissent un but de sélection que l'on appelle idéotype.
Dans ce travail, nous voulons déterminer une carte d'idéotypes de sorgho. Nous sommes donc confrontés à un problème d'optimisation d'un modèle complexe. Une méthode classiquement utilisée dans ce contexte est la méthode Efficient Global Optimization (EGO), fondée sur un métamodèle de krigeage. Ici, une telle approche n'est pas adaptée. En effet, la distribution des entrées météorologiques suit un modèle stochastique dont les paramètres varient continûment dans l'espace en suivant un gradient Nord-Sud. L'optimisation des paramètres variétaux est alors conditionnelle à ces paramètres de climat. D'autre part, la fonction à maximiser n'est connue que par un nombre limité de simulations, donc à une erreur près.
Notre cadre de travail concerne donc l'optimisation conditionnelle d'une fonction bruitée. Les extensions existantes de l'algorithe EGO ne prennent pas en charge ce cadre. Dans cette thèse, un nouveau critère pour l'optimisation conditionnelle d'une fonction bruitée est proposé et étudié. Une preuve de convergence de l'algorithme associé est également fournie. Une métaphore de l'optimisation conditionnelle est la recherche d'une ligne de crête. A partir de simulations sur des fonctions test, une étude des performances de ce nouveau critère est proposée, de même qu'une comparaison avec le critère habituellement utilisé pour la recherche de ligne de crête. Les résultats de cette étude montrent l'intérêt de notre critère.
L'application à la cartographie d'idéotypes de sorgho a été testée sur l'espace couvert par le Sénégal, le sud du Mali et le Burkina Faso. Elle a consisté à maximiser le rendement espéré en fonction de 4 paramètres du modèle Samara : la longueur de la phase végétative, la longueur maximale des racines, le potentiel de réserve des tiges, et la mortalité des feuilles. Les résultats de cette optimisation recoupent en partie l'analyse de sensibilité menée sur ces mêmes paramètres.

Mots clés : Modèle de culture, Paramètres variétaux, Distribution des entrées, Modèle stochastique, Méthode EGO, Ligne de crête, Processus gaussien, critère d'échantillonnage, plan séquentiel, Fonction bruitée.

Abstract : In the Sahel region, the irregular rainfall distribution in time and space generates variety $\times$ year and variety $\times$ location interactions. Therefore, determining variety with the best expected yield would take many years of experimentation in each location.
Alternatively, the best variety could be identified by maximizing the predicted yield using a crop simulation model that describes growth and development of a crop in interaction with agro-environmental conditions.
The average yield depends on the probability distribution of environmental inputs, which is location specific, while the cultivar parameters that maximize this yield define the ideotype, i.e. the selection target.
In this work, we want to draw the map of sorghum ideotypes in Sub Saharan Africa. To face the problem of optimizing a complex model, an algorithm conventionally used in this context is the Efficient Global Optimization method (EGO), based on kriging as a surrogate model. Here, the distribution of meteorological inputs follows a stochastic model whose parameters vary continuously in space along a North-South gradient. Consequently, the optimization of varietal parameters is conditional on these climate parameters. Moreover, the function to maximize is noisy, because expectation and quantiles are merely estimated with a limited number of simulations. We aimed at adapting the EGO algorithm to the conditional optimization of a noisy function. Extensions exist either for the optimization of noisy functions or for the conditional optimization of deterministic functions, ie the search for the values of a subset of parameters that optimize the function conditionally to the values taken by another subset, which are fixed. Proof of convergence of the associated algorithm is also provided. A metaphor for conditional optimization is the search for a crest line. No method has yet been developed for the conditional optimization of noisy functions: this is what we propose in this thesis. Testing this new method on test functions shows that, in case of a high level of noise on the function, the PEQI criterion that we propose is better than the PEI criterion usually implemented in such a situation.
The application of this new optimization method sorghum ideotypes parameters mapping has been tested in the area covered by Senegal, southern Mali and Burkina Faso. It consisted in maximizing the expected yield with respect to 4 parameters of Samara model: vegetative phase length, maximum root length, stem reserve potential, and leaf mortality. The results of this optimization partly coincide with the sensitivity analysis conducted on these same parameters.

Keywords : crop model, cultuvar parameters, inputs distribution, stochastic model, EGO method, crest line, Gaussian process, sampling criterion, sequential design, noisy function.