



Prédictions génomiques des interactions Génotype x Environnement à l'aide d'indicateurs agro-climatiques chez le blé tendre (*Triticum aestivum* L.)

Delphine Ly

► To cite this version:

Delphine Ly. Prédictions génomiques des interactions Génotype x Environnement à l'aide d'indicateurs agro-climatiques chez le blé tendre (*Triticum aestivum* L.). Sciences agricoles. Université Blaise Pascal - Clermont-Ferrand II, 2016. Français. NNT : 2016CLF22669 . tel-01477172

HAL Id: tel-01477172

<https://theses.hal.science/tel-01477172v1>

Submitted on 27 Feb 2017

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

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

Ecole Doctorale Sciences de la Vie, Santé, Agronomie, Environnement

N° d'ordre : 683

THESE DE DOCTORAT

Présenté à l'Université Blaise Pascal pour l'obtention du grade de

DOCTEUR D'UNIVERSITE

Spécialité : Physiologie et génétique moléculaires

Soutenue le 25 janvier 2016 par :

Delphine LY

**Prédictions génomiques des interactions Génotype x Environnement à l'aide
d'indicateurs agro-climatiques chez le blé tendre (*Triticum aestivum* L.)**

Directeur de thèse : Gilles CHARMET

Co-encadrant de thèse : Arnaud GAUFFRETEAU

Membres du Jury

MOUZEYAR Saïd	Professeur, Université Blaise Pascal, Clermont-Ferrand	Président
BASTIEN Catherine	Directrice de Recherche INRA, Orléans	Rapporteur
MANGIN Brigitte	Directrice de Recherche INRA, Toulouse	Rapporteur
MANZANARES-DAULEUX Maria	Professeur, Agro-Campus Ouest, Rennes	Examineur
PRAUD Sébastien	Coordinateur Recherche Génétique, Biogemma, Chappes	Examineur
ROUMET Pierre	Chargé de Recherche INRA, Montpellier	Examineur
CHARMET Gilles	Directeur de Recherche INRA, Clermont-Ferrand	Directeur de thèse

UMR 1095 GDEC

**Genomic Predictions of Genotype x Environment interactions
using weather data in wheat (*Triticum aestivum* L.)**

LABORATOIRE D'ACCUEIL

GDEC

Génétique Diversité Eco-physiologie des Céréales

UMR 1095 INRA/Université Blaise Pascal

5 chemin de Beaulieu

63039 Clermont-Ferrand Cedex 2

FRANCE

Résumé

Un des principaux enjeux de l'amélioration des plantes consiste aujourd'hui à faire face au changement climatique, en assurant un rendement élevé et plus stable dans des systèmes agricoles économes en intrants (eau, fertilisants) et respectueux de l'environnement. Les nouvelles variétés de blé devront non seulement être tolérantes aux stress hydriques et aux fortes températures, mais aussi continuer à être productives avec des apports limités en fertilisation, tout en maintenant une qualité du grain adaptés aux différents usages. De nouvelles méthodes de prédiction des réponses des blés à ces stress sont indispensables pour avancer dans cette direction.

Dans ce travail, nous avons tout d'abord identifié les stress qui régissaient les interactions entre génotypes et les environnements (GxE) dans les essais considérés, puis développé un modèle génomique de l'adaptation à un stress environnemental (Factorial Regression genomic Best Linear Unbiased Prediction ou FR-gBLUP), en particulier pour le stress hydrique. En émettant l'hypothèse que plus des variétés de blés sont génétiquement proches, plus elles répondront de façon similaire à un stress environnemental donné, nous avons mesuré par validation croisée des gains de précision de prédiction par rapport à un modèle additif variant entre 3.5% et 15.4%. Des simulations complètent l'étude en démontrant que plus la part de variance expliquée par les réponses au stress considéré est importante, plus le modèle FR-gBLUP apporte un gain de précision.

Pour prédire les réponses variétales à un stress particulier, les environnements doivent être finement caractérisés pour les stress limitant le développement des plantes. En nous intéressant plus particulièrement au stress azoté en France, nous avons établi des indicateurs de stress à partir d'un modèle de culture, et les avons comparés à des indicateurs classiques, tels que le type de conduite azotée ou l'azote disponible. Nous avons ainsi mis en évidence l'intérêt des modèles de culture pour caractériser les interactions GxE et pour prédire la réponse génomique au stress azoté, à condition que le signal d'interaction soit assez fort.

Au-delà de l'application potentielle de ces méthodes pour la sélection ou la recommandation de variétés de blés plus adaptées ou plus résistantes au changement climatique, les résultats de ce travail démontrent aussi l'intérêt de la complémentarité des approches éco-physiologiques et génétiques.

Mots Clés : Interaction génotype x environnement, Prédiction génomique, Stress abiotique, Régression factorielle, Régression aléatoire, Modèle de culture, Caractérisation environnementale

Abstract

In a climate change context, assuring high and stable yield in more sustainable agricultural systems is a major challenge for plant breeding. We are aiming for future wheat varieties which will be heat and drought tolerant, and also productive in limited fertilization input environments. New prediction methods of the response to these stresses are needed to move forward.

In this study, we first identified stresses that generated interactions between genotypes and environments (GxE) in our experimental trials and then developed a genomic model for adaptation to a particular environmental stress (Factorial Regression genomic Best Linear Unbiased Prediction ou FR-gBLUP), in our case drought. This model hypothesizes that the more individuals are genetically close, the more their response to a stress will resemble. We used cross-validations to measure prediction accuracy gains compared to an additive model and observed gains between 3.5% and 15.4%. Besides, simulation studies showed that the more the variance explained by the responses to the stress is important, the more the FR-gBLUP model will improve the additive model.

Furthermore, fine characterization of the stresses limiting the plants' growth is required to predict varietal responses to a particular stress. We focused on the particular case of nitrogen stress in France. By establishing crop model based stress indicators and comparing them to classical indicators, such as the management system or the available nitrogen, we pointed out the interest of crop model to characterize GxE interactions and to predict the genomic response to nitrogen stress, as long as the GxE interaction signal is strong enough.

Beyond the potential applications of these methods for breeding or recommendation for varieties more adapted or tolerant to environmental stresses, this study also raises the interest of coupling eco-physiological and genetics approaches.

Key words : Genotype x environment interaction, Genomic prediction, Abiotic stress, Factorial regression, Random regression, Crop model, Environmental characterization

Avant-propos

Cette thèse a été financée par l'Institut National de la Recherche Agronomique (INRA) dans le cadre de son méta-programme SelGen (Sélection Génomique) et la région d'Auvergne. Un financement complémentaire SelGen a permis les dépenses liées aux déplacements. L'Ambassade de France en Australie a financé le voyage à Toowoomba (Australie).

Ce manuscrit est présenté sous la forme :

- d'une première partie introductive qui vise à mettre en place la question des prédictions génomiques des interactions GxE et à comprendre l'intérêt de travailler chez le blé,
- d'un recueil de deux chapitres, qui correspondent à des publications, ainsi que des travaux complémentaires,
- suivis d'une discussion générale, qui resitue les résultats par rapport aux travaux faits dans la littérature et évalue les possibilités d'application de nos études, à court et à plus long terme.

Publications dans des revues scientifiques avec comité de lecture

Ly, D., Chenu, K., Gauffreteau, A., Rincet, R., Huet, S., Gouache, D., and Charmet, G. How can crop modeling improve genomic predictions? Illustration for N deficiency in a wheat Core Collection. [En préparation]

Ly, D., Huet, S., Gauffreteau, A., Cormier, F., Rincet, R., Jannink, J-L., and Charmet, G. Genomic random regression for adaptation breeding: Whole-genome prediction of the reaction norms to environmental stress in bread wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* [Soumise]

Ly, D., Hamblin, M., Rabbi, I., Melaku, G., Moshood, B., Gauch, H.G., Okechukwu, R., Dixon, A.G.O., Kulakow, P., Jannink, J-L. (2013) Relatedness and Genotype x Environment Interaction Affect Prediction Accuracies in Genomic Selection: A Study in Cassava. *Crop Sci.*53:1. [Publié]

Cormier, F., Ly, D., Praud, S., Le Gouis, J., Crossa, J. Improving genomic prediction using a GWAS-based method to pre-select marker in multi-environment data. [En préparation]

Communications dans des séminaires ou congrès

Ly, D., Huet, S., Gauffreteau, A., Cormier, F., Rincet, R., Jannink, J-L., and Charmet, G. Genomic random regression for adaptation breeding: Whole-genome prediction of the reaction norms to environmental stress in bread wheat (*Triticum aestivum* L.). 16th Eucarpia Biometrics in Plant Breeding Meeting, du 9 au 11 septembre 2015, Wageningen (Pays-Bas). [Poster]

Ly, D., Huet, S., Gauffreteau, A., Martre, P., and Charmet, G. Usefulness of Crop Models to characterize environmental conditions in genomic predictions of Genotype-by-Environment interactions in wheat, Plant & Animal Genome Conference XXIV, du 9 au 13 janvier 2014, San Diego (USA). [Poster]

Remerciements

Je remercie Catherine Bastien, Brigitte Mangin, d'avoir accepté d'être les rapporteurs de ce manuscrit, ainsi que Maria Manzanares-Dauleux, Saïd Mouzeyar, Sébastien Praud et Pierre Roumet de compléter la composition du jury.

Ces 3 dernières années ont été une belle et grande aventure, sur le plan scientifique mais aussi humain, et je voudrais remercier tous ceux qui y ont participé. C'est une histoire qui remonte à l'agro de Montpellier. Tout d'abord, merci à toute l'équipe enseignante APIMET pour m'avoir fait découvrir les joies de l'amélioration des plantes ! C'était une grande chance d'avoir des enseignants passionnés, qui nous ont poussés et encouragés. Merci également aux encadrants de stage, Anne-Marie Bochard et Jean-Luc Jannink et Martha Hamblin, qui m'ont donné envie de commencer l'aventure de la thèse.

Dans cette grande aventure, mes premiers remerciements vont à Gilles Charmet, mon directeur de thèse, tout d'abord pour ne pas avoir été trop regardant pour le choix du candidat de la thèse : je crois bien que c'était l'entretien le plus court de ma vie jusqu'à aujourd'hui. En tout cas, cela a été très formateur pour moi de pouvoir réaliser cette thèse avec vous. Votre « culture » du blé dans tous ses aspects a été un atout majeur dans la réalisation de ce travail. Merci pour tout cela, et surtout, merci de m'avoir toujours soutenue.

J'adresse également mes remerciements à Arnaud Gauffreteau. Tu as été un co-encadrant de choc. Merci pour ta bonne humeur et ton grand esprit critique. Ca a toujours été très enrichissant, presque épuisant, de ressortir d'une de nos petites réunions, ou de nos longues conversations téléphoniques.

De la même manière, je remercie Sylvie Huet, auprès de qui j'ai redécouvert le gBLUP. Merci de m'avoir reçue à Jouy avec le sourire et avec bienveillance. Merci aussi à l'équipe MAIAGE (pour leur accueil sympathique, leur barbecue et leur trousse à pharmacie) !

Je remercie aussi Karine Chenu pour ton accueil dans le monde d'APSIM et aussi celui des kangourous. Je pense que ce sont tes qualités humaines ont rendu mon séjour à Toowoomba très stimulant sur le plan scientifique. Merci pour ce regain de motivation (et de vitamine D) que tu m'as apporté en milieu de thèse.

Un grand merci à Pierre Martre qui m'a initié aux joies de la calibration des modèles de culture. Merci pour tes conseils et pour les discussions qui ont définis la thèse.

Dans le même esprit, je remercie également tous les membres de comité de thèse David Gouache, Pierre Roumet et Jean-Luc Jannink, et aussi Nicolas Heslot, Jacques Bordes, Jacques Legouis, l'équipe DGS, qui m'ont énormément aidé à construire cette aventure, pas par pas : Merci pour tous vos conseils inestimables pour progresser dans les pistes à suivre.

Je tiens aussi à remercier Thierry Langin, directeur d'unité, pour son écoute et son soutien, ainsi que les personnes de l'UMR, pour avoir fait du GDEC un espace de travail agréable à vivre.

En particulier, je remercie les copains, liés aux interactions GxE de près ou de loin : Renaud, un grand merci pour toute l'aide que tu as pu apporter, tu as été un interlocuteur privilégié pour ton expertise, ton analyse et tes idées ; Fabien, merci pour les données et les discussions enrichissantes, pour les buibui et aussi la non-organisation de voyage; Jonathan et Rémi pour les pauses-café ou pauses-eau golf, volley, duel quizz ; Titouan, merci d'avoir été un compagnon de thèse, à l'écoute et motivant, merci de m'avoir accueillie dans ton bureau pendant les nausées, et de m'avoir laissée gagner au ping-pong ; Pierre, Alice, Caroline, Anne, Julio, Elodie, Anthony, Robin, Pauline et Magali, merci pour votre simplicité et bonne humeur, qui ont égayé chacune de mes journées au GDEC, merci d'avoir également supporté mes impatiences à 11h58, parfois même en m'offrant des choses à grignoter ... un énorme merci !

Je tiens aussi à remercier sportivement les équipes de volley (ADAS, CUC, Toowoomba) ainsi que l' « équipe » des fléchettes, (Pascal, Olive, Magali, ...) pour ces grands moments, peut-être pas toujours de sport, mais de bonne humeur et détente.

J'ai une pensée particulière pour l'équipe Sogeres de la cantine de Crouël ! Votre gentillesse et vos petites attentions surpassent toutes les difficultés que l'on aurait à choisir entre deux desserts. Merci d'avoir fait de nos pauses du midi un véritable moment gastronomique.

Merci aux copains, qui sont au-delà de la chaîne des puy, ou encore plus loin, mais toujours présents quand il le faut : Alberto, Aude, Irène et Louise. Merci d'avoir toujours été présents.

Merci à tous ceux que je n'ai cités et qui ne sont pas moins importants.

Mes derniers remerciements s'adressent à la famille : à mes parents et à Titi. Merci d'y avoir cru et de m'avoir soutenue et encouragée.

Surtout, un grand merci à Charles. Merci d'être toi, d'avoir été là pour cette aventure, et d'être là pour la prochaine.

TABLE DES MATIERES

INTRODUCTION	16
1- Les interactions entre génotype et environnement en sélection face au changement climatique	19
1.1- De nouvelles conditions agricoles dues aux effets du changement climatique	19
1.2- Les interactions entre génotypes et environnements en sélection	20
1.3- Les modèles classiques d'analyse des interactions entre génotypes et environnements	21
1.4- Le cas de l'analyse de l'adaptation des génotypes à un stress environnemental.....	25
a- Définition des normes de réactions par rapport à un stress environnemental.....	25
b- Le modèle de régression factorielle	26
c- La régression aléatoire modélisant le pedigree entre les individus.....	27
2- Les apports de la génomique pour la prédiction des interactions GxE et des normes de réaction à un stress environnemental	28
2.1- Les avancées génomiques récentes.....	28
2.2- Une extension de la régression factorielle pour la détection de QTL.....	29
2.3- Des extensions genome-wide du modèle de régression factorielle.....	30
3- La caractérisation environnementale permet une meilleure compréhension des stress causant des interactions GxE chez le blé tendre	33
3.1- La stagnation des rendements du blé	33
3.2- La compréhension de l'impact des stress environnementaux affectant le développement du blé	34
a- Comment l'élaboration du rendement est affectée par le climat ?.....	34
b- L'identification des stress impactant le développement du blé et les stress du blé attendus au vu des prévisions climatiques.....	36
3.3- Une caractérisation environnementale adaptée au cycle de développement des cultures	39
3.4- La caractérisation des stress impactant le blé à l'aide des modèles de culture.....	40
4- Problématique	43

Références	44
CHAPITRE 1	52
1.1- Choix des stress environnementaux pour l'étude des normes de réactions.....	55
1.2 - Publication 1.....	59
Abstract.....	60
Introduction.....	61
Materials and Methods	62
Experimental data	62
Statistical models for genomic predictions	63
Importance of the random regression to an environmental covariate effect	66
Assessment of the prediction quality of the models.....	66
Simulation of Experimental data	67
Results.....	67
Different patterns of reaction norms of grain number to nitrogen or to water deficit stress .	67
Modeling the genomic random regression to water deficit stress improve prediction accuracies	69
Influence of the variance of the genomic random regression to the environmental covariate	69
Discussion.....	72
Improving Genomic Predictions by modeling the response to an environmental stress	72
Limits of the FR-g BLUP	73
A step toward exploiting genomic GxE interactions in genomic selection	73
Applications for breeders and advisers.....	74
Conclusions.....	75
References.....	77
1.3- Equivalence entre le modèle FR-gBLUP et le modèle de Jarquin et al. (2014)	84
Conclusion du chapitre 1.....	87
Références complémentaires	88
CHAPITRE 2	90

2.1- Préambule du chapitre 2	94
2.2- Publication 2.....	96
Introduction.....	97
Materials and Methods	99
Core Collection experimental data.....	99
Genotypic data	99
Calibration and testing of crop-model APSIM-Wheat.....	99
Environmental Characterization.....	100
Explanatory power of characterizations of N stress	101
Crop-model characterization of intra-year GxE interactions effects	102
Genomic predictions for the different characterizations of N stress.....	106
Assessment of predictive ability of the different genomic prediction models	109
Results.....	109
Environmental characterization of N stress using a crop-model	109
Model selection criteria favor crop-model based indicators to characterize G x N stress interactions.....	115
Ability of crop-model indicators to capture the differences in GxE interactions variances per year.....	117
Genomic prediction gains by modeling the interaction with N stress using crop-model indicators.....	117
Discussion.....	121
Crop modelling provides N stress indicators well suited for modelling GxE interactions	121
How much gain in prediction accuracy to expect from GxE modelling?.....	121
Genomic predictions of the response to crop-model stress indicators for N stress per year	122
Potential of genomics and crop modeling for predicting GxE interactions	123
Conclusion	124
Supplementary data.....	124
References.....	126

Conclusion du Chapitre 2	131
Références complémentaires	132
DISCUSSION GENERALE	133
1- La prédiction des normes de réaction aux stress environnementaux par une approche éco-physiologique et génétique.....	136
1.1 Le FR-gBLUP : Un modèle de prédiction génomique des réponses à un stress environnemental	136
1.2- La caractérisation environnementale des stress à l'aide de modèle de culture éco-physiologique	138
1.3- Intérêts de la prédiction génomique des normes de réactions aux stress environnementaux par rapport aux autres approches éco-physiologiques et génétiques	140
2- Les conditions d'application de notre approche éco-physiologique et génétique.....	143
3- De nouvelles possibilités pour sélectionner des variétés adaptées ?	145
4- Une recommandation variétale par environnement ?	147
5- Vers quelles perspectives est-ce que les avancées génomiques futures et le développement du phénotypage haut-débit conduiront la prédiction des interactions GxE ?	148
5.1- Vers une prédiction basée sur l'identification des polymorphismes causaux	148
5.2- ... et une amélioration des capacités de prédiction grâce au phénotypage à haut débit	149
Références	150

INTRODUCTION

TABLE DES MATIERES : INTRODUCTION

1- Les interactions entre génotype et environnement en sélection face au changement climatique	19
1.1-De nouvelles conditions agricoles dues aux effets du changement climatique	19
1.2-Les interactions entre génotypes et environnements en sélection	20
1.3-Les modèles classiques d'analyse des interactions entre génotypes et environnements	21
1.4-Le cas de l'analyse de l'adaptation des génotypes à un stress environnemental	25
Définition des normes de réactions par rapport à un stress environnemental.....	25
Le modèle de régression factorielle	26
La régression aléatoire modélisant le pedigree entre les individus.....	27
2- Les apports de la génomique pour la prédiction des interactions GxE et des normes de réaction à un stress environnemental	28
2.1- Les avancées génomiques récentes.....	28
2.2- Une extension de la régression factorielle pour la détection de QTL.....	29
2.3- Des extensions genome-wide du modèle de régression factorielle.....	30
3- La caractérisation environnementale permet une meilleure compréhension des stress causant des interactions GxE chez le blé tendre	33
3.1- La stagnation des rendements du blé	33
3.2- La compréhension de l'impact des stress environnementaux affectant le développement du blé	34
a- Comment l'élaboration du rendement est affectée par le climat ?.....	34
b- L'identification des stress impactant le développement du blé et les stress du blé attendus au vu des prévisions climatiques.....	36
3.3- Une caractérisation environnementale adaptée au cycle de développement des cultures	39
3.4- La caractérisation des stress impactant le blé à l'aide des modèles de culture.....	40

INTRODUCTION

4- Problématique	43
-------------------------------	-----------

1- Les interactions entre génotype et environnement en sélection face au changement climatique

1.1- De nouvelles conditions agricoles dues aux effets du changement climatique

Depuis 1950, de nombreux phénomènes indiquent une tendance vers un réchauffement de l'atmosphère et des océans, à l'échelle mondiale. Récemment, le réchauffement s'est accéléré, avec des vagues de chaleur de plus en plus fréquentes en Europe, Asie et Australie (GIEC, 2014). Ainsi 2015 semble être l'année la plus chaude jamais observée depuis les premiers enregistrements de 1880, et la majorité des années les plus chaudes se concentrent au 21^e siècle (NOAA <http://www.ncdc.noaa.gov/>). Ce réchauffement est attribué principalement à l'accroissement des émissions dans l'atmosphère des gaz à effet de serre d'origine anthropique (dioxyde de carbone CO₂, qui est passé de 320 ppm en 1950 à plus de 400 aujourd'hui ; protoxyde d'azote N₂O ; méthane CH₄ ; ...), et dans une moindre mesure aux cycles solaires ou le volcanisme. Le réchauffement climatique devrait s'accroître avec une intensité qui dépendra du degré de maîtrise des émissions des gaz à effet de serre. C'est l'enjeu des rencontres internationales comme la COP 21 organisée à Paris en décembre 2015.

A l'échelle planétaire, l'agriculture contribue pour 13.5% aux émissions de GES (GIEC, 2007). Les émissions de l'agriculture, des forêts et des pêches ont quasiment doublé au cours des cinquante dernières années. Au cours de la seconde moitié du XX^e siècle, la production agricole mondiale a plus que doublé, essentiellement en raison de l'augmentation des rendements dans les pays industrialisés des grandes cultures, telles que le riz (*Oryza sativa* L.), le maïs (*Zea mays* L.) et le blé (*Triticum aestivum* L.). Le progrès génétique, la mécanisation, les engrais et les traitements phytosanitaires ont tous contribué à la progression des rendements observés. Dans le même temps, les émissions gazeuses d'ammoniac et d'oxydes d'azote à effets polluants dues à l'application d'engrais synthétiques ont augmenté de 55% entre 1960 et 2012 (<http://faostat3.fao.org>). Il a été montré que la surutilisation de fertilisants azotés déséquilibre les écosystèmes (Hirel et al., 2007), et peut entraîner une eutrophisation des systèmes aquatiques environnants (Beman et al., 2005) par exemple.

Les conditions environnementales pour l'agriculture vont être profondément bouleversées dans les prochaines années. Des études prévoient une plus grande fréquence d'épisodes de chaleurs extrêmes et une plus grande variabilité des pluies entraînant des périodes de sécheresse (Battisti and Naylor, 2009; Trnka et al., 2014). Les systèmes de production devront être plus résilients face au risque climatique mais également plus économes en intrants. Le GIEC recommande ainsi de

développer de nouvelles variétés adaptées aux augmentations de concentration atmosphérique de CO₂, des températures, des périodes de sécheresses, et avec une meilleure valorisation de l'azote.

1.2- Les interactions entre génotypes et environnements en sélection

Face à ces bouleversements des conditions agricoles, la sélection variétale doit tenir compte, encore plus qu'auparavant, des interactions entre génotypes et environnements (GxE), c'est-à-dire du fait que les différences phénotypiques entre les génotypes ne sont pas nécessairement constantes d'un environnement à l'autre. Les interactions GxE représentent, plus précisément, une déviation au modèle additif, où le phénotype est expliqué par un effet principal du génotype et un effet principal de l'environnement. Ces interactions sont un enjeu essentiel pour l'amélioration génétique des espèces agricoles (Falconer, 1952, 1990; Kolmodin et al., 2002), les plantes en particulier (Ceccarelli et al., 1994). Ces effets, liés aux conditions climatiques, aux sols, aux pratiques agricoles, contribuent à une instabilité spatiale et inter-annuelle des rendements. En présence de fortes interactions GxE, le classement des variétés peut-être modifié entre deux environnements (Baker, 1988). Ignorer ces interactions pourrait amener à choisir une variété de blé très performante dans un environnement particulier mais qui s'effondrerait par rapport aux autres variétés dans un environnement différent. Du fait de ces possibles inversions de rangs, le choix des environnements pour conduire les essais variétaux doit être représentatif des environnements de production.

Différentes stratégies de sélection existent pour prendre en compte, ou tout du moins, s'accommoder des interactions GxE : les ignorer, les éviter ou les exploiter (Eisemann et al., 1990), en fonction de l'influence de la performance moyenne et du niveau de GxE sur un caractère (Tableau 1).

		Performance moyenne	
		Elevé	Faible
Niveau GxE	Elevé	Potentiel pour des variétés adaptées localement	Potentiel pour des variétés adaptées localement
	Faible	Idéal pour des variétés à large adaptation	Indésirable

Tableau 1 : Combinaisons possibles des performances moyennes et niveau d'interactions GxE et leur implication pour la sélection. A partir de Ceccarelli (1989)

Ignorer les interactions GxE n'est raisonnable qu'à condition que le caractère étudié soit assez héritable. Dans le cas contraire, si les interactions représentent une proportion importante de la

variance phénotypique, le phénotype sera difficilement un bon prédicteur de la valeur génétique de l'individu.

Pour ce qui est d'éviter les interactions GxE, les sélectionneurs peuvent chercher à minimiser ces interactions et rechercher une adaptation large de leurs variétés, c'est-à-dire, sélectionner des variétés avec de bonnes performances dans presque tous les environnements. Cette stratégie a notamment été appliquée au cours de la Révolution Verte. C'est en visant une adaptation large que le Dr Norman Ernst Borlaug propose le shuttle breeding (Ortiz et al., 2007). Il met en place l'évaluation des lignées de blé dans deux environnements très contrastés du Mexique. Les environnements de la vallée Yaqui et de Toluca se distinguent effectivement par 2500 m d'altitude, 10 degrés de latitude, des spectres de maladies, types de sol, et diverses contraintes environnementales différents. Les variétés issues de la sélection dans ces environnements étaient moins sensibles à la longueur du jour, résistantes à la rouille, et adaptées à de nombreuses latitudes et types de sol. En sélectionnant les lignées avec les meilleures performances dans les 2 environnements, les effets négatifs des changements de rangs dus aux interactions GxE sont alors minimisés. Toutefois, dans des régions moins favorables, avec de faibles potentiels écologiques pour la production agricole, où l'insécurité alimentaire est d'autant plus marquée par les fluctuations climatiques, cette stratégie n'a eu que des retours économiques bas et instables (McCown *et al.*, 1992).

En opposition avec la recherche d'une adaptation large, la stratégie d'exploiter les interactions GxE vise une adaptation spécifique, dans le but d'identifier des variétés performantes dans des environnements particuliers. En amélioration variétale, l'utilisation d'un matériel génétique local, par rapport à un matériel choisi pour la stabilité de ses rendements à grande échelle, permet de privilégier une meilleure tolérance aux stress biotiques et abiotiques (Byerlee and Husain, 1993; Byerlee and Morris, 1993; Almekinders et al., 1994; Eyzaguirre and Iwanaga, 1996; Tesemma and Bechere, 1998). De plus, les variétés supérieures dans une population d'environnements représentative des lieux de production ne sont pas nécessairement les meilleures variétés dans un environnement spécifique. La sélection variétale s'intéresse alors également à la compréhension des réponses variétales aux stress environnementaux pour exploiter les interactions GxE et pouvoir développer de variétés adaptées localement. Face au changement climatique, cette compréhension des réactions variétales aux stress environnementaux devient d'autant plus nécessaire.

1.3- Les modèles classiques d'analyse des interactions entre génotypes et environnements

De nombreuses méthodes d'analyse, dites « biadditives » ou « bilinéaires » ont été développées pour modéliser les interactions GxE par un ou un nombre relativement faible de termes(s) multiplicatif(s)

INTRODUCTION

composé(s) de variables caractéristiques des génotypes et des environnements. Cette famille de modèle comprend notamment la régression conjointe (Finlay and Wilkinson, 1963), les modèles AMMI (Gauch and Zobel, 1988) et la régression factorielle (Denis, 1980). Toutefois, ces méthodes sont essentiellement descriptives ou explicatives, car les paramètres sont liés au jeu de données utilisées. Ils ne sont pas utilisables pour prédire le comportement de candidats à la sélection.

L'analyse des interactions GxE pour l'identification de génotypes à adaptation large a été amplement étudiée par des modèles de régression conjointe. A l'origine proposée par Mooers (1921), Stringfield and Salter (1934), et Yates and Cochran (1938), l'utilisation du modèle de régression conjointe a surtout été développée par Finlay and Wilkinson (1963), qui ont suggéré que la qualité d'un environnement pouvait être mesurée par la moyenne des performances de tous les génotypes sur cet environnement.

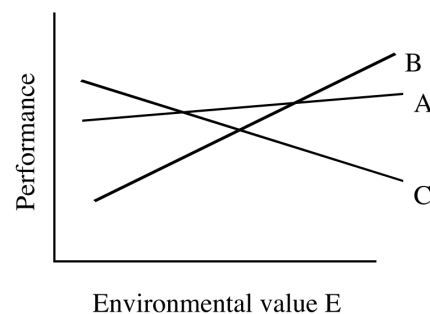


Figure 1 : Exemple de représentation des régressions des génotypes A, B et C selon une régression conjointe de Finlay-Wilkinson.

Chaque environnement est caractérisé par la moyenne des performances des génotypes, de leur rendements par exemple, ce qui permet alors d'apprécier le potentiel de chaque environnement. La régression conjointe représente les interactions GxE par un produit des sensibilités spécifiques à chaque génotype et de la performance moyenne de chaque environnement. La prédiction des performances d'un génotype suppose une réponse linéaire au potentiel de chaque environnement. La régression conjointe de Finlay-Wilkinson permet alors d'identifier des génotypes stables, dans le sens de Römer (1917), c'est-à-dire constants dans les différents environnements.

Toutefois, si on considère qu'idéalement, un génotype serait insensible à un changement environnemental lorsque les conditions sont défavorables, mais réceptif à des conditions favorables, pour exprimer pleinement son potentiel, l'hypothèse de linéarité imposée par la régression de Finlay-Wilkinson s'avère contraignante voire trompeuse. Verma et Chahal (1978) propose une modification de la régression conjointe de Finlay-Wilkinson qui consiste à séparer la régression en 2, sur les environnements défavorables d'une part, et sur les environnements favorables d'autre part.

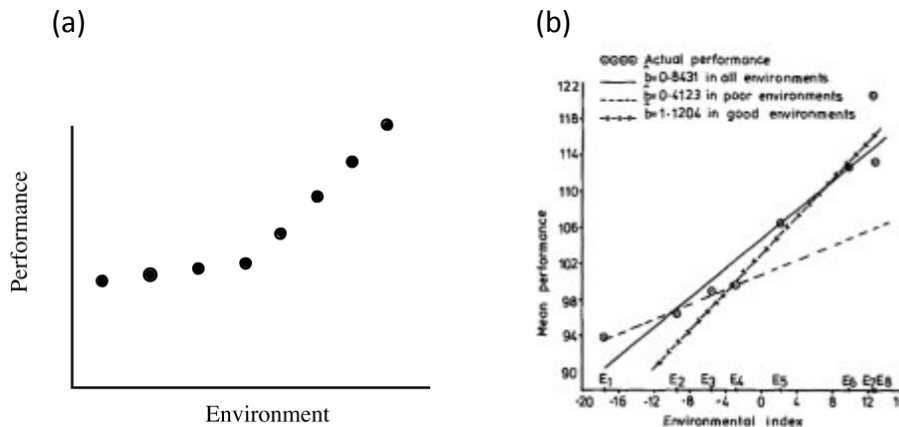


Figure 2 : (2a) Comportement idéal d'un génotype, dont les performances ne seraient pas affectées dans des environnements défavorables, mais valorisées dans des environnements favorables. (2b) Comparaison de la régression conjointe de Finlay-Wilkinson (sur tous les environnements, en trait plein) et de Verma et Chahal (1978) (sur les environnements défavorables et favorables, séparément, en pointillé)

De cette manière, ils identifient parmi 10 variétés de *Nicotiana rustica*, un génotype avec ce type de comportement. Au lieu d'avoir une pente moyenne de 0.84 sur tous les environnements, ils estiment une pente de 0.41 sur les environnements défavorables (en dessous de la moyenne) et une pente de 1.12 dans les environnements favorables. Des indicateurs dérivés de ces régressions, tels que la pente ou le coefficient de détermination permettent de mesurer la stabilité des génotypes (Finlay and Wilkinson, 1963).

Pour ce qui est de la sélection des variétés adaptées à des environnements spécifiques, des modèles de structuration des interactions ont été mis en place. Leur objectif est d'identifier des sets d'environnements où les génotypes obtiennent des performances similaires, et de définir des sets de génotypes qui ont des comportements similaires dans des environnements différents.

Byth et al., (1976) propose un clustering simultané de l'ensemble des génotypes et des environnements de manière à minimiser la variance GxE intra-groupe et maximiser la variance GxE inter-groupe. (Baril et al., 1994) réalise une classification hiérarchique des génotypes et des milieux pour décomposer et interpréter l'interaction GxE avec des données de rendements de blé mesurés au cours d'une période de 8 ans dans 7 lieux, avec deux modes de culture par lieu. Cette approche a mis en évidence des redondances parmi les milieux qui constituent leur réseau expérimental du nord de la France (Baril et al., 1994).

Le modèle AMMI (Additive Main effects and Multiplicative Interaction) est un autre modèle de structuration de l'interaction très répandu pour étudier les interactions GxE. A l'origine développé

INTRODUCTION

par Fisher and Mackenzie (1923), il est repris ensuite par de nombreux généticiens (Williams, 1952; Gollob, 1968; Mandel, 1969; Perkins and Perkins, 1972; Gabriel, 1978; Kempton, 1984), il a été popularisé en sélection végétale par (Gauch and Zobel, 1988). Dans ce modèle, les composantes du signal d'interaction sont identifiées par une Analyse en Composantes Principales pour les distinguer du bruit résiduel. L'approche AMMI considère en fait une famille de modèles AMMI-k, où un modèle AMMI comprend les k termes d'interactions liés aux k premières composantes principales de l'interaction GxE. Les modèles AMMI ont permis de délimiter avec succès des méga-environnements pour la culture du blé dur en Algérie (Annicchiarico, 2002).

La représentation visuelle en « biplot » (Gabriel, 1978) est également un avantage majeur du modèle AMMI, pour aider à l'interprétation des interactions GxE. Les génotypes et les environnements sont projetés sur les 2 axes principaux de l'ACP ; les génotypes étant représentés par des points et les environnements par des axes passant par l'ordonnée à l'origine.

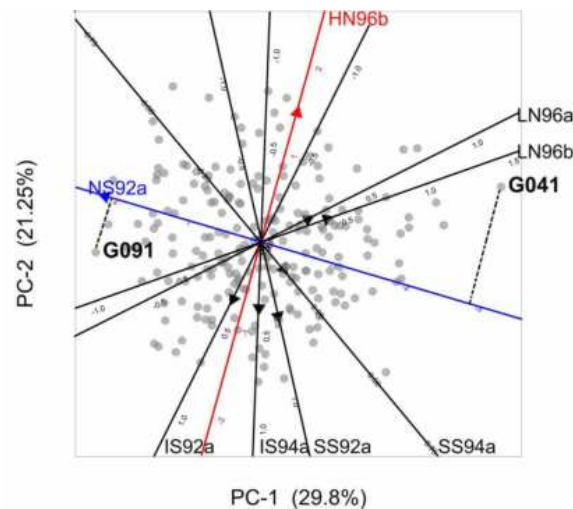


Figure 3 Biplots issus d'analyse AMMI pour décrire les interactions GxE. Les cercles gris représentent les génotypes, et les triangles pleins les environnements. Ces derniers indiquent la direction d'interactions croissantes. La projection du génotype G091 sur l'axe de l'environnement NS92a est représentée en pointillé (Malosetti et al., 2013)

Ainsi, la visualisation de génotypes à hautes performances à proximité d'un environnement (ou axe) particulier permet de déterminer quels génotypes seront meilleurs dans quel environnement (Malosetti et al., 2013).

Les différents modèles d'analyse des interactions GxE présentés ci-dessus considèrent un environnement comme une entité sans chercher à le caractériser en termes de stress. L'environnement peut être défini par la moyenne des performances des génotypes qui y sont cultivés (régression conjointe), ou représenté par des composantes principales en fonction des interactions

qu'ils ont générées (AMMI). Cette décomposition de l'environnement en facteurs environnementaux ouvre le champ à des études des comportements et sensibilités des génotypes en réponse à des stress environnementaux. Rappelons toutefois que ces modèles sont essentiellement descriptifs ou explicatifs, mais ne permettent pas de prédire le comportement de génotypes non évalués, car les paramètres dépendent du jeu de données utilisé.

1.4- Le cas de l'analyse de l'adaptation des génotypes à un stress environnemental

a- Définition des normes de réactions par rapport à un stress environnemental

La notion de plasticité phénotypique, introduite par Nilsson-Ehle (1914), est la capacité d'un génotype à produire différents phénotypes en fonction de son environnement. L'ensemble des différentes réponses exprimées par un seul génotype face à des conditions environnementales variables peut être décrit comme la sensibilité environnementale ou encore la norme de réaction d'un génotype (Falconer, 1990). Woltereck proposait pour la première fois en 1909 le terme norme de réaction pour décrire les variations morphologiques observées chez des clones de daphnies (*Daphnia*). En représentant graphiquement la taille de la tête en fonction d'une mesure de l'environnement pour chaque groupe de clones, Woltereck a démontré que les différences observées pour la taille de la tête étaient reliées à la disponibilité de la nourriture.

Au sein d'une population, en fonction des différents génotypes qui la composent, la variance des normes de réaction est due à la présence d'interactions GxE, c'est-à-dire aux réponses différentielles des génotypes suivant les environnements.

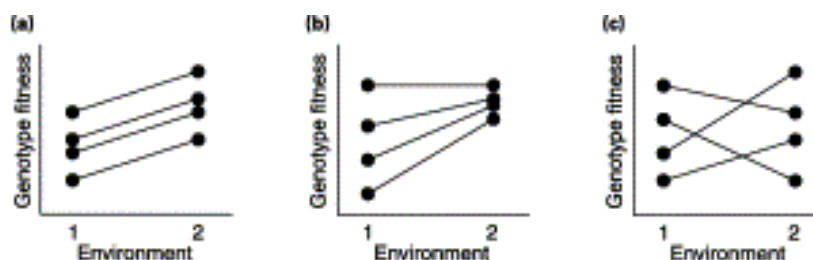


Figure 4 : Représentation des interactions Génotype x Environnement par des normes de réaction.
(a) un cas sans interaction ; (b-c) deux cas avec interaction

Elles sont souvent illustrées par des normes de réaction par rapport aux environnements ou par rapport à un stress environnemental. Dans le cas purement additif (Figure 4 a), tous les génotypes répondent de la même manière à un changement d'environnement. Dans le cas b, l'amplitude des réponses au changement d'environnement est variable selon les génotypes. Dans le cas c, les normes de réaction sont également variables et le classement des génotypes diffère d'un environnement à un autre.

La prédiction des normes de réaction à un stress environnemental peut s'avérer intéressante pour anticiper le comportement de nouvelles variétés dans de nouveaux environnements sans passer par des expérimentations systématiques.

b- Le modèle de régression factorielle

Denis (1980) propose la régression factorielle pour modéliser les normes de réaction. Ce modèle est une régression linéaire des génotypes sur des covariables environnementales. Il modélise ainsi une interaction entre les génotypes et les environnements caractérisés par ces covariables. Le coefficient de régression de chaque génotype représente la sensibilité de chaque génotype i à la/aux covariable(s) choisie(s). Par souci de simplicité, nous ne considérerons qu'une seule covariable environnementale. Le modèle de régression factorielle s'écrit comme suit :

$$y_{ijk} = \mu + g_i + e_j + \beta_i x_j + \varepsilon_{ijk} \quad (\text{Modèle 1})$$

Avec y_{ijk} le phénotype de la répétition k du génotype i dans l'environnement j , μ la moyenne générale, g_i l'effet du génotype i , e_j l'effet de l'environnement j , β_i le coefficient de régression du génotype i ou encore le terme de sensibilité du génotype i , x_j la covariable environnementale de l'environnement j , et ε_{ijk} les résidus supposés indépendants, identiquement distribués et suivant une loi normale $N(0, \sigma^2)$. Il pourrait également y avoir un terme symétrique d'interaction, qui correspondrait à un terme de sensibilité de l'environnement à une covariable génotypique. Par exemple, l'environnement pourrait avoir différents effets en fonction des précocités d'une culture.

En définissant les covariables de stress qui caractérisent un réseau d'environnements, le modèle de régression factorielle est utile pour examiner les causes des interactions GxE. C'est ainsi notamment que (Balfourier et al., 1997) parviennent à expliquer 72.8% des interactions pour la vigueur au printemps du ray-grass, avec 2 covariables, une environnementale (la température minimale du mois le plus froid) et une génotypique (la fréquence allélique de l'allèle PGI2-20). Hebert et al., (1995) quant à eux, comprennent 74 à 98% des interactions GxE des caractères racinaires du maïs avec une seule covariable climatique, liée à la température du sol ou la pluviométrie selon les environnements. (Vargas et al., 1999) montrent effectivement que la régression factorielle permet d'expliquer une large proportion des interactions GxE, et ceci avec un nombre de degrés de liberté relativement faible.

Toutefois, la régression factorielle a été critiquée sur deux points. La sélection des variables environnementale nécessite d'identifier les stress causaux des interactions GxE. Le choix retenu n'explique souvent qu'une petite proportion des interactions GxE. De plus, la régression factorielle

fait l'hypothèse d'une réponse linéaire des génotypes à l'environnement, ce qui est rarement le cas en biologie (Des Marais et al., 2013). Néanmoins, par son hypothèse de linéarité et a fortiori de continuité entre les environnements, le modèle de régression factorielle présente l'avantage de pouvoir prédire des performances génotypiques d'individus préalablement évalués, dans de nouveaux environnements (Smith et al., 2005).

c- La régression aléatoire modélisant le pedigree entre les individus

Une version aléatoire de la régression factorielle a été proposée par Ravagnolo and Misztal (2000). Dans le modèle de régression aléatoire, le coefficient de régression des génotypes à la covariable environnementale est considéré comme aléatoire. La variance covariance de cet effet a une structure bien particulière : elle est proportionnelle à la matrice d'apparentement entre les individus, calculée à partir du pedigree.

Le modèle (simplifié) est le suivant :

$$y_{ijk} = \mu + \beta x_j + g_i + e_j + b_i x_j + \varepsilon_{ijk} \quad (\text{Modèle 2})$$

$$\begin{pmatrix} \mathbf{g} \\ \mathbf{b} \\ \boldsymbol{\varepsilon} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma_G^2 \mathbf{A} & \sigma_{gb} \mathbf{A} & \mathbf{0} \\ \sigma_{gb} \mathbf{A} & \sigma_B^2 \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma^2 \mathbf{I} \end{pmatrix} \right)$$

Où β est le coefficient de régression à la covariable x_j , les b_i représentent l'effet aléatoire des régressions génotypiques à x_j , et plus précisément, les écarts des normes de réactions à la régression moyenne βx_j , et les b_i suivent une distribution $N(0, \sigma_B^2 \mathbf{A})$, où \mathbf{A} est la matrice d'apparentement calculée par les pedigree. Avec cette modélisation, les génotypes ont des normes de réaction plus ou moins parallèles, en fonction de leur degré d'apparentement.

Le modèle de régression aléatoire a été utilisé pour explorer les causes environnementales responsables d'interactions GxE sur la masse corporelle de la truite et leur a permis d'identifier notamment un effet de la photopériode et de la température (Sae-Lim et al., 2014). Ce modèle a également permis chez les bovins, d'estimer pour chaque niveau de stress environnemental différentes héritabilités des caractères et différents classements des individus, (Ravagnolo and Misztal, 2000), et des changements de variances génétiques dans le temps (Lawlor et al., 2002). Kolmodin et al., (2002) détectent dans une population de bovins provenant de différents pays, des variations significatives des pentes des normes de réactions pour la production laitière en réponse au mois de vêlage, et montrent que cette régression aléatoire permet d'éviter les problèmes de

connectivité entre les groupes ou les pays d'origine. Ils mesurent de faibles changements de rangs, excepté dans le cas d'environnements extrêmes.

En opposition avec les modèles de structuration présentés dans la partie précédente, la régression aléatoire, de même que la régression factorielle, modélise l'adaptation des génotypes à un facteur environnemental. Ainsi, ils permettent de prédire les performances d'individus dans de nouveaux environnements, caractérisés par une covariable environnementale. De plus, la modélisation aléatoire proposée par Ravagnolo et Miszta (2000) permet de prédire l'effet des normes de réaction, à partir de l'appareillement, donc pour de nouveaux génotypes non expérimentés dont on connaît l'appareillement avec les génotypes testés.

2- Les apports de la génomique pour la prédiction des interactions GxE et des normes de réaction à un stress environnemental

2.1- Les avancées génomiques récentes

Au cours des dernières décennies, le développement de technologies de séquençage a fourni des informations de marquage moléculaire, sur la présence ou l'absence de sites de restrictions (Restriction Fragment Length Polymorphism, RFLP), sur la taille des fragments amplifiés (Amplified Fragment Length Polymorphism, AFLP), ou encore le nombre de motifs de micro-satellites (Single Sequence Repeat, SSR). L'accès aux marqueurs moléculaires sur l'ensemble du génome est de moins en moins cher et de plus en plus rapide. En effet, les progrès réalisés dans le séquençage des génomes a permis l'identification de nombreux polymorphismes au niveau des nucléotides (Single Nucleotide Polymorphism, SNP). Les marqueurs SNPs se sont rapidement développés, notamment avec la mise en place de puces à ADN, qui fournissent de nombreux marqueurs co-dominants. Chez le blé tendre, l'International Wheat Genome Sequencing Consortium (IWGSC) a entamé en 2005, le séquençage du génome, dans le but d'avoir une séquence de référence du blé pour chacun des 3 génomes (3 génomes x 7 chromosomes = 21 chromosomes, 17Gb). Aujourd'hui de nombreux SNPs sont disponibles, localisés au niveau de gènes, dans des régions inter-géniques, ou encore sur des sites d'insertion d'éléments transposables (Lai et al., 2015). Avec l'accroissement considérable de l'information moléculaire, la sélection commence à avoir accès aux polymorphismes causaux expliquant la variation de caractères d'intérêt, ou à des polymorphismes en déséquilibre de liaison très fort avec ces derniers. La sélection assistée par marqueurs vise alors à identifier et cumuler les allèles favorables aux différents loci contrôlant la variation des caractères d'intérêt.

Toutefois, pour l'étude des caractères agronomiques complexes tels que le rendement ou la qualité du grain, résultant de l'expression de nombreux gènes à effets faibles difficiles à détecter, l'approche

Quantitative Trait Loci (QTL), qui teste les effets de chaque marqueur a été complétée par des approches multi-locus sur l'ensemble du génome qui permettent d'avoir une estimation de la variance génétique totale (Segura et al., 2012) et par la sélection génomique qui calcule simultanément les effets de tous les marqueurs (Goddard & Hayes, 2007; Hayes, Bowman, Chamberlain, & Goddard, 2009; Heffner, Sorrells, & Jannink, 2009; Lorenz et al., 2011; Meuwissen, Hayes, & Goddard, 2001). Le modèle de prédiction génomique le plus communément utilisé est le Genomic Best Linear Unbiased Prediction (gBLUP) (Habier et al., 2013). Il prédit les effets génétiques des individus en utilisant une matrice d'apparentement dite «réalisée», à partir des différents génotypes aux SNPs. Utilisé pour prédire la taille chez les humains (Yang et al., 2010), le gBLUP est également largement utilisé en sélection animale et végétale (VanRaden, 2008).

2.2- Une extension de la régression factorielle pour la détection de QTL

Alors que les modèles de détection de QTL de base prédisent uniquement des effets génétiques additifs, certains modèles incluent les interactions GxE. Malosetti et al., (2004) par exemple ont développé une extension du modèle de régression factorielle qui estime les effets des différents allèles aux QTL en réponse à l'environnement. Ils proposent les deux modèles suivants :

$$y_{ij} = \mu + E_j + x_i\alpha + x_i\alpha_j + \varepsilon_{ij} \quad (\text{Modèle 3})$$

$$y_{ij} = \mu + E_j + x_i\alpha + \beta x_i z_j + \varepsilon_{ij} \quad (\text{Modèle 4})$$

Avec y_{ij} l'observation phénotypique du génotype i dans l'environnement j , μ la moyenne générale, E_j l'effet fixe de l'environnement j , x_i le prédicteur génétique pour l'effet additif du QTL, de valeur 1 ou -1, α l'effet principal du QTL. Dans le modèle (3), α_j représente l'effet fixe du QTL dans l'environnement j . Dans le modèle (4), z_j est la covariable environnementale décrivant l'environnement j . β est le coefficient de régression représentant la sensibilité du QTL à la covariable environnementale z_j .

A partir de ce modèle d'interactions QTL x Environnement, Mathews et al., (2008) identifient chez le blé huit QTLs de rendement, dont deux colocalisent avec des gènes de précocité et de hauteur. Ils remarquent particulièrement pour un gène de précocité que les allèles de la variété Babax retardent la floraison. En conditions irriguées, ces allèles augmentent le rendement, du fait d'un temps plus long d'interception des rayonnements et de production de biomasse. Cependant, dans des environnements à stress hydriques tardifs, ces mêmes allèles sont défavorables. En effet, une floraison précoce permet aux plantes d'échapper aux stress terminaux de fin de cycle. Ainsi, leur étude a révélé des QTL de rendement en interaction avec l'environnement, lié à la précocité,

soulignant le fait que les allèles des marqueurs peuvent avoir des effets différents dans des environnements différents.

Toutefois, du fait d'une architecture génétique complexe des caractères agronomiques et généralement d'un manque de puissance lié aux dispositifs expérimentaux utilisés, la détection des QTL, et les interactions QTL x Environnement a fortiori, a conduit à des approches genome-wide.

2.3- Des extensions genome-wide du modèle de régression factorielle

L'arrivée de la sélection génomique (Meuwissen et al., 2001) a permis d'apporter un nouveau regard sur la prédiction des interactions GxE, incitant à utiliser simultanément toutes les covariables environnementales disponibles, de façon symétrique à l'utilisation de l'ensemble des marqueurs (Crossa et al., 2013, Jarquin, et al., 2014, Heslot et al. 2014). Les interactions GxE peuvent être considérés comme un effet aléatoire, avec une structure de variance covariance qui prend en compte non seulement une matrice d'apparentement, calculée à partir des données génomiques (Meuwissen, et al. 2001), mais aussi une matrice de similarité entre les environnements à partir des covariables disponibles, sur le climat et le sol notamment (Crossa et al., 2010; Burgueño et al., 2012; Jarquín et al., 2014).

Jarquín et al. (2014) proposent ainsi une extension génomique du gBLUP pour la prédiction des interactions GxE. Ils caractérisent chaque environnement avec diverses covariables environnementales. On note q le nombre de covariables environnementales. Chacune de ces covariables environnementales prend des valeurs différentes par génotype, pour modéliser le fait que chaque génotype n'est pas affecté de la même manière par les stress climatiques. Par exemple, dans un même environnement, des précocités différentes génèrent des impacts de stress différents. Jarquín et al. (2014) définissent ainsi une matrice de covariables environnementales centrées et réduites $\mathbf{W}_{n \times q}$ de dimension le nombre d'observations par le nombre de covariables, et avec une valeur de covariable par combinaison génotype-environnement.

$$\mathbf{y} = \mu \mathbf{1}_n + \mathbf{Z}_W \mathbf{w}_{GE} + \mathbf{Z}_G \mathbf{u}_G + \mathbf{Z}_{GE} \mathbf{gw}_{GE} + \mathbf{e} \quad (\text{Modèle 5})$$

$$\left\{ \begin{array}{l} \mathbf{w}_{GE} \sim N(0, \mathbf{\Omega} \sigma_w^2) \\ \mathbf{u}_G \sim N(0, \mathbf{K} \sigma_G^2) \\ \mathbf{gw}_{GE} \sim N(0, [\mathbf{Z}_G \mathbf{K} \mathbf{Z}_G'] \circ \mathbf{\Omega} \sigma_{GE}^2) \\ \mathbf{e} \sim N(0, \sigma_e^2) \end{array} \right.$$

où \mathbf{y} est le vecteur des observations phénotypiques, μ est la moyenne générale, \mathbf{u}_G est l'effet génétique additif, modélisé comme dans un modèle gBLUP (Henderson, 1975; Habier et al., 2013), suivant une loi normale $\mathbf{u}_G \sim N(0, \mathbf{K} \sigma_G^2)$, où \mathbf{K} est calculé à partir de la formule de VanRaden's (2008) tel que $\mathbf{K} = \mathbf{M} \mathbf{M}' / k$, avec \mathbf{M} la matrice de génotypage centrée et k une constante liée aux fréquences alléliques $k = 2 \sum_m p_m (1 - p_m)$, \mathbf{w}_{GE} est l'effet environnement dû aux covariables environnementales, qui suit une loi normale $\mathbf{w}_{GE} \sim N(0, \mathbf{\Omega} \sigma_w^2)$, où $\mathbf{\Omega}_{n \times n}$ est une matrice représentant les ressemblances entre environnements, équivalente à une matrice d'apparement pour les environnements, calculée à partir de la matrice de covariables environnementales $\mathbf{W}_{n \times q}$ qui caractérise chaque environnement, telle que $\mathbf{\Omega}_{n \times n} \propto \mathbf{W}_{n \times q} \mathbf{W}_{q \times n}'$. \mathbf{gw}_{GE} est l'effet d'interaction entre les génotypes caractérisés par leurs marqueurs et les environnements caractérisés par les covariables environnementales, le symbole \circ représente un produit de Hadamard, c'est-à-dire un produit terme à terme entre deux matrices de même dimensions, $\mathbf{Z}_W, \mathbf{Z}_G, \mathbf{Z}_{GE}$ sont les matrices d'incidence qui assurent la correspondance entre les observations et les niveaux des effets, pour les effets environnementaux, génétique et d'interaction, et \mathbf{e} est la résiduelle.

La structure de variance covariance des interactions GxE est modélisée par un produit de Hadamard entre $[\mathbf{Z}_G \mathbf{G} \mathbf{Z}_G']$ et $\mathbf{\Omega}_{n \times n}$. Jarquin et al. (2014) montrent que $\text{cov}(\mathbf{gw}_{ij}, \mathbf{gw}_{ij'}) \propto \mathbf{K}_{ii'} \mathbf{\Omega}_{jj'}$. Ainsi, pour 2 génotypes i et i' dans 2 environnements j et j' , plus les génotypes i et i' ont des marqueurs similaires (ou plus leur apparement $\mathbf{K}_{ii'}$ est élevé), et plus les environnements j et j' sont semblables (ou plus $\mathbf{\Omega}_{jj'}$ est grand), plus leurs interactions GxE seront proches.

Le modèle de Jarquin et al., (2014) (Modèle 5) permet de prédire les performances additives et d'interactions d'un nouveau génotype à partir de ses informations moléculaires. Si les covariables environnementales sont définies par environnement et par précocité par exemple, faire une prédiction pour un nouveau génotype dans un nouvel environnement requiert de connaître les

INTRODUCTION

marqueurs moléculaires des nouveaux individus à tester, les précocités de ces individus et les différents stress en fonction des différentes précocités.

Une autre extension génomique de la régression factorielle a également développée en considérant les covariances entre les covariables environnementales (Heslot et al., 2014).

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}_E \mathbf{u}_E + \mathbf{Z}_G \mathbf{u}_G + \mathbf{Z}_S \mathbf{u}_S + \mathbf{e} \quad (\text{Modèle 6})$$

$$\begin{cases} \mathbf{u}_E \sim N(\mathbf{0}, \mathbf{I}_{n_E} \sigma_E^2) \\ \mathbf{u}_G \sim N(\mathbf{0}, \mathbf{K} \sigma_G^2) \\ \mathbf{u}_S \sim N(\mathbf{0}, \mathbf{H} \otimes \mathbf{K} \sigma_{GE}^2) \\ \mathbf{e} \sim N(\mathbf{0}, \mathbf{I}_e \sigma_e^2) \end{cases}$$

Avec les mêmes notations que précédemment, et où \mathbf{u}_E représente l'effet aléatoire environnemental, \mathbf{u}_S représente le vecteur des sensibilités de chaque génotype à chaque covariable environnementale, \mathbf{u}_S a pour dimension le nombre de génotypes (n_G) multiplié par le nombre de covariables environnementales q , et suit une loi normale $N(\mathbf{0}, \mathbf{H} \otimes \mathbf{K} \sigma_{GE}^2)$ où \otimes désigne un produit de Kronecker, ici entre la matrice de covariance des covariables environnementales \mathbf{H} de dimension $q \times q$, et la matrice d'apparentement génomique \mathbf{K} ($n_G \times n_G$). Dans le cas d'un dispositif expérimental équilibré, les matrices d'incidence pour les effets environnementaux, génétiques, et de sensibilité génétique $\mathbf{Z}_E, \mathbf{Z}_G, \mathbf{Z}_S$ sont définies comme étant : $\mathbf{Z}_E = (\mathbf{I}_{n_E} \otimes \mathbf{1}_{n_G}), \mathbf{Z}_G = (\mathbf{1}_{n_E} \otimes \mathbf{I}_{n_G}), \mathbf{Z}_S = (\mathbf{S}_{n_E \times q} \otimes \mathbf{I}_{n_G})$ où $\mathbf{S}_{n_E \times q}$ est la matrice contenant les scores centrés réduits des q covariables environnementales dans chacun des n_E environnements. Les sensibilités génétiques dans \mathbf{u}_S représentent donc, pour une covariable donnée, les normes de réactions à cette covariable environnementale.

Dans ce modèle l'interaction GxE est modélisée par la somme des sensibilités génétiques à chaque covariable environnementale. Pour un génotype i dans un environnement j , l'interaction ge_{ij} peut s'écrire comme la somme sur toutes les covariables environnementales, des produits des pentes des normes de réaction par les scores environnementaux de chaque covariable : $ge_{ij} = \sum_{c=1}^q \mathbf{u}_{S_{i,c}} \mathbf{S}_{j,c}$.

Pour 2 génotypes i et i' , dans 2 environnements j et j' , caractérisés par les covariables $\mathbf{S}_{j,(1,\dots,q)}$ et $\mathbf{S}_{j',(1,\dots,q)}$, l'interaction GxE est définie à partir des covariances entre les sensibilités génétiques aux covariables environnementales, qui sont modélisées telles que $\text{cov}(\mathbf{u}_{S_{iq}}, \mathbf{u}_{S_{i'q}}) \propto \mathbf{K}_{ii'} \mathbf{H}_{qq'}$. Ainsi,

l'interaction entre ge_{ij} et $ge_{i'j'}$ est définie à partir des covariances entre les individus i et i' , et les covariances entre toutes les covariables environnementales, qui caractérisent les environnements j et j' .

Les modèles de Jarquín et al., (2014) et Heslot et al., (2014) ont montré des gains de précision de prédiction par rapport au modèle additif de 34% et 11.1%, respectivement. Notons que ces modèles font l'hypothèse lourde d'indépendance entre les valeurs génétiques additives et les normes de réactions contenues, et supposent une même variance pour les normes de réactions à des covariables différentes.

Avec les extensions génomiques de la régression factorielle proposées par Jarquín et al., (2014) et Heslot et al., (2014), il serait possible alors de faire des prédictions des interactions GxE pour de nouvelles conditions environnementales, à condition d'avoir leur caractérisation pour toutes les covariables environnementales. Ceci exigerait alors une description exhaustive des conditions futures, ce qui est encore difficilement possible aujourd'hui. Une approche plus ciblée consisterait à identifier et comprendre les stress environnementaux qui constituent les interactions GxE, et à prédire à l'aide de la génomique les normes de réactions à ces stress. La focalisation sur un nombre réduit de stress environnementaux permettrait d'affiner la caractérisation des environnements par rapport à l'impact de leur stress sur le développement d'une culture, et peut-être alors de mieux prédire les différences de normes de réaction entre les individus.

3- La caractérisation environnementale permet une meilleure compréhension des stress causant des interactions GxE chez le blé tendre

Le blé tendre (*Triticum aestivum*), espèce modèle et plante de grande culture, est un sujet d'études exploré dans de nombreux domaines, parmi lesquels on trouve l'écophysiologie, l'agronomie, la génétique et la génomique. Le développement de toutes ces connaissances autour du blé tendre en fait un modèle idéal pour l'étude de la caractérisation des stress environnementaux générant des interactions GxE chez cette espèce, et la prédiction génomique des réponses à ces stress.

3.1- La stagnation des rendements du blé

Le blé (*Triticum*) est la 3^e culture mondiale par l'importance de sa récolte (695 Mt en 2012, probablement 732 en 2015), et avec le riz et le maïs, la céréale la plus consommée par l'homme (Organisation des Nations Unies pour l'alimentation et l'agriculture, <http://faostat.fao.org/>). Au total, 90% de la production mondiale de blé est destinée à l'alimentation, humaine et animale. La France en est le 1^{er} producteur de l'Union Européenne, avec près de 41 Mt en 2015 sur un peu plus de 5 M

INTRODUCTION

d'hectares. Actuellement environ 50% de la production française est destinée à l'exportation : le blé joue donc un rôle majeur dans l'équilibre de notre balance commerciale et la France contribue significativement à la sécurité alimentaire mondiale. Le blé tendre (*T. aestivum* L.) adapté à la culture en milieu tempéré représente environ 95% de la production mondiale de blé, et est utilisé majoritairement pour produire de la farine intervenant dans la fabrication de différents produits (pains, biscuits), mais aussi en alimentation animale et dans l'industrie de l'amidon et de l'éthanol.

Malgré des variations fortes de rendement traduisant des contraintes climatiques et des stratégies de culture très différentes entre les pays, l'importance de la production mondiale du blé tendre s'explique par son fort potentiel de rendement et par sa grande plasticité phénotypique qui permet de la cultiver dans la plupart des régions agricoles mondiales. Durant la 2^e moitié du XX^e siècle, le progrès génétique du blé tendre a permis d'améliorer l'allocation de la biomasse dans les grains (indice de récolte ou harvest index), et ainsi, les rendements. De plus, l'utilisation de la fertilisation, l'irrigation, et les pesticides ont permis de s'affranchir en partie des limites environnementales, et d'exprimer plus fortement le progrès génétique. La sélection du blé était alors principalement axée sur le nanisme des plantes, et leur capacité à valoriser l'azote. Brisson et al., (2010) ont mis en évidence que la stagnation des rendements de blé en Europe au cours des deux dernières décennies n'est pas due à un déclin du progrès génétique, mais au fait que ce progrès est contrecarré principalement par les effets du réchauffement climatique, et aussi par les réductions d'intrants dans une moindre mesure. En comprenant comment ces stress impactent le développement du blé, il serait possible de définir des caractérisations environnementales fines de ces stress, de définir de nouveaux écotypes et d'améliorer le progrès génétique (Donald, 1968).

3.2- La compréhension de l'impact des stress environnementaux affectant le développement du blé

a- Comment l'élaboration du rendement est affectée par le climat ?

L'interaction entre le blé et les stress, entre autres liés au changement climatique, a été beaucoup étudiée. De nombreuses études ont examiné les impacts des stress climatiques survenant à des stades de développement particuliers, notamment au travers des mécanismes physiologiques mis en jeu et des effets sur l'élaboration du rendement.

Le rendement en céréales est déterminé par 2 composantes principales : le nombre de grains par unité de surface (NG/m²), et le poids moyen d'un grain (P1G) (Boiffin et al., 1981). La première peut encore se décomposer ainsi :

$$\text{NG/m}^2 = \text{Nombre plantes/m}^2 \times \text{Nombre épis/plante} \times \text{Nombre grains/épi}$$

INTRODUCTION

Ces composantes du rendement se définissent tout au long du développement du blé, du semis à la maturité du grain. L'aptitude à la compensation du blé lui permet souvent de rattraper sur les composantes les plus tardives du rendement les insuffisances ou pertes dues à des accidents climatiques par exemple.

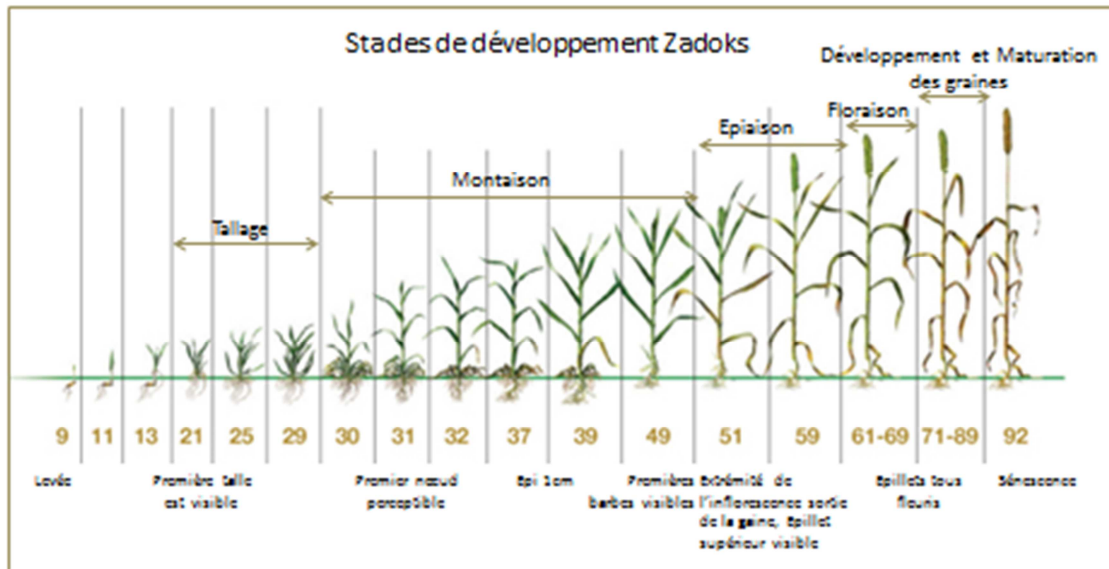


Figure 5 : Le cycle de développement du blé est généralement décrit par les systèmes de Zadoks (Zadoks et al., 1974)

La mise en place du nombre de pieds par unité de surface a lieu pendant la germination (ZC00 à ZC09) (Figure 5). Ce stade correspond à l'apparition d'une tigelle et d'une racine, jusqu'à l'émergence de la tigelle au-dessus de la surface du champ (la levée). La faculté germinative, qui dépend entre autres de l'âge et des conditions de conservation des semences, le taux de levée effective, qui peut être réduit par la présence d'obstacles physiques dans le sol ou un excès d'eau, et les pertes hivernales dues au froid et à l'excès d'eau sont des facteurs influents dans la détermination de la densité de plantes (Kirby, 1981; Gate, 1995).

Le nombre d'épis par plante s'établit suite au tallage (ZC20) et à la montaison (ZC30). Pendant le tallage, les talles sont émises en fonction des sommes de températures (phyllochrone), très souvent en excès. Exceptionnellement, des stress ponctuels peuvent empêcher l'apparition de certaines talles. Puis, pendant la montaison, c'est-à-dire l'élongation des tiges, selon les conditions de croissance et de disponibilité des ressources (nutrition minérale et lumière pour la photosynthèse), la régression des talles est plus ou moins importante.

La fertilité des épis, et donc le nombre de grains par épi s'élabore au milieu du tallage, lorsque l'apex réalise sa transition florale et élabore des ébauches d'épillets (Figure 6) jusqu'à la floraison, c'est-à-

INTRODUCTION

dire la fécondation. Des mécanismes environnementaux régulent la mise en place du nombre d'épillets par épi, du nombre de fleurs par épillet, et du pourcentage de fertilité, donc au final le nombre de « puits » susceptibles de stocker les assimilats sous forme de réserve (amidon et protéines).

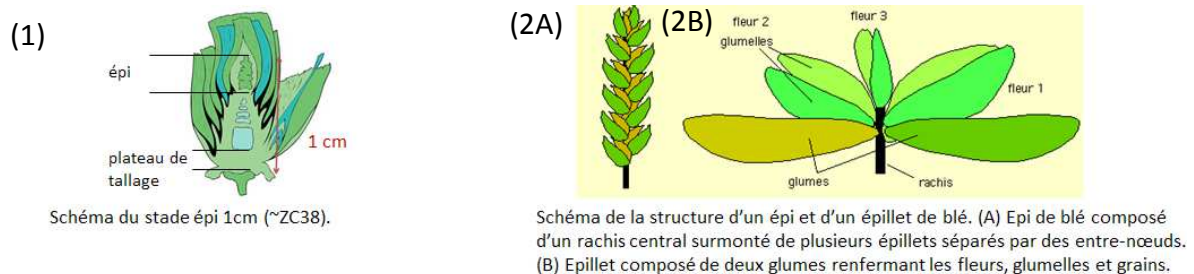


Figure 6 A gauche (1), schéma du stade épi 1cm. A droite (2), schéma d'un épi et d'un épillet de blé.

La formation des grains, et plus précisément leur poids moyen, commence à la fécondation, et peut-être même dès la détermination du poids de l'ovaire (Calderini et al., 2001). Lors de la multiplication cellulaire (floraison ZC61 à stade grain laiteux ZC73), les cellules de l'endosperme se multiplient puis se différencient en cellules de stockage d'amidon. Cette phase est assez cruciale parce que le nombre de cellules peut être limitant. Puis, du stade grain laiteux (ZC73) au stade grain pâteux (ZC83), les cellules formées précédemment se remplissent d'amidon, et dans une moindre proportion, de protéines de réserve synthétisées dans le grain à partir des assimilats apportés par la sève depuis les organes photosynthétiques (limbes, gaines foliaires, glumelles de l'épi...). Enfin, la dernière phase de l'élaboration du poids des grains correspond à la dessiccation des grains, lorsque la maturité physiologique est atteinte (ZC 92). La formation des grains est particulièrement sensible au déficit en eau et aux fortes températures, en particulier pendant la phase de division cellulaire (Farooq et al., 2011).

La possibilité de compensation des effets des stress entre les phases de l'élaboration du rendement est une aptitude du blé qui constitue une des raisons principales de son succès. Une altération du nombre de grains par m², par exemple, peut être partiellement compensée par un plus fort accroissement du poids des grains.

b- L'identification des stress impactant le développement du blé et les stress du blé attendus au vu des prévisions climatiques

Les études agronomiques réalisées par (Meynard and Sebillotte, 1994) ont permis d'identifier les stress impactant les différentes phases de développement du blé (Figure 7).

INTRODUCTION

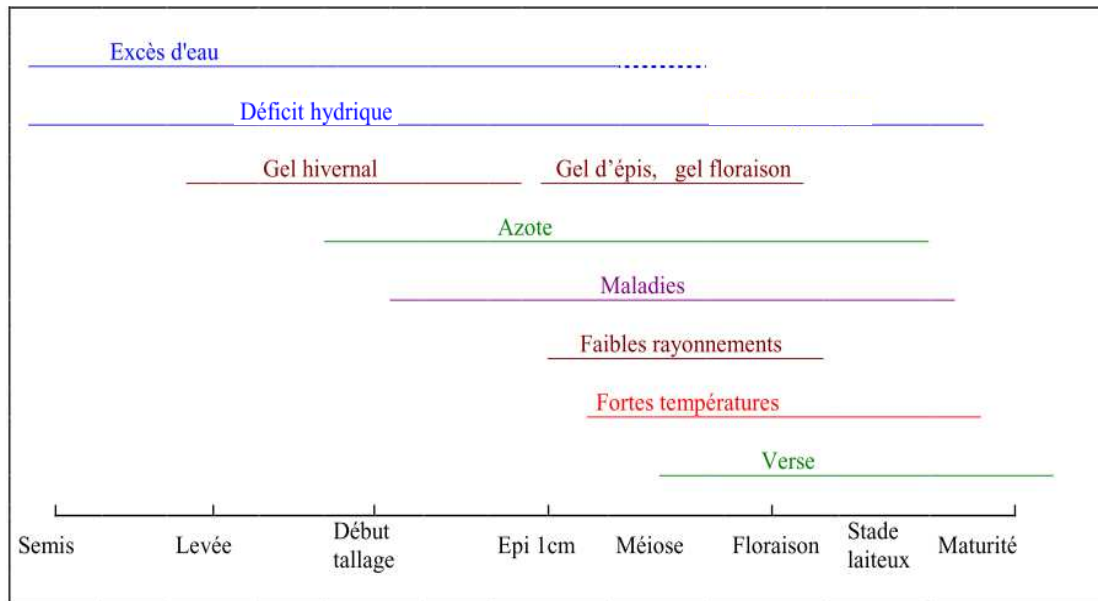


Figure 7 : Principaux facteurs abiotiques limitants qui peuvent intervenir au cours du cycle de développement du blé (adapté de Meynard et Sebillote (1994) et Lecomte (2005))

Les stress hautes températures et stress hydriques

Une hausse globale des températures est attendue, avec des variations régionales, plus importante aux latitudes élevées. Des événements de sécheresses plus fréquents sont également à prévoir mais sont difficiles à modéliser car ils font intervenir des mécanismes complexes comme les circulations atmosphériques et océaniques. Parmi les facteurs climatiques responsables de la stagnation des rendements des blés en Europe, Brisson et al., (2010) ont identifié les épisodes de sécheresse pendant l'élongation de la tige, et les températures échaudantes, c'est-à-dire les températures hautes entraînant une moindre efficacité du remplissage du grain, produisant alors des grains de faible poids, voire des grains ridés et de petite taille.

Les températures élevées affectent à la fois le nombre et le poids des grains. Le nombre de grains peut être réduit suite à des problèmes de fertilité lors de l'initiation des épillets, la sporogénèse, la pollinisation, ou encore la fécondation. De plus, la chaleur accélère le développement de l'épi (Porter and Gawith, 1999), ce qui réduit ainsi le nombre d'épillets et par conséquent le nombre de grains par épi (Saini and Aspinall, 1982). Des températures au-delà de 30°C au moment du développement floral peuvent provoquer une stérilité complète (Saini and Aspinall, 1982), mais il existe toutefois une variabilité génétique pour cette sensibilité (Gibson and Paulsen, 1999). Fischer (1985) a mesuré une réduction du nombre de grains de 4% pour chaque augmentation de degré (entre 15-22°C) dans les 30 jours précédant la floraison. Pour ce qui est du remplissage du grain, les enzymes impliquées dans la synthèse d'amidon sont sensibles aux fortes températures. Une enzyme en particulier, la Soluble Starch Synthase (SSS), est sensible aux températures supérieures à 25°C (Rijven, 1986). En outre, la

INTRODUCTION

chaleur accélère et raccourcit le remplissage (Dias and Lidon, 2009). Dans ces conditions, l'approvisionnement des photo-assimilats vers les grains peut être limité (Calderini et al., 2006). Stone and Nicolas (1994) ont montré dans un matériel génétique diversifié, que des chocs thermiques de 3-4 jours (jusqu'à 40°C) ont réduit les rendements jusqu'à 25%.

Le déficit hydrique, quant à lui, affecte tous les stades de développement du blé. Le tallage et la montaison sont des stades où une sécheresse peut avoir un impact majeur. La comparaison de variétés anciennes et modernes a montré qu'en condition de stress hydrique, les variétés anciennes sur-produisent des talles, dont la plupart ne produiront pas de grains, alors que les variétés modernes tolérantes produisent moins de talles, mais dont la majorité survit. Pendant la montaison, la sécheresse peut induire une plus forte régression des talles et diminuer ainsi le nombre d'épis. De plus, le nombre de grains par épi est également diminué. Lorsque la période de stress hydrique coïncide avec la floraison, la sécheresse affecte tout particulièrement la fertilité, et impacte le nombre de grains, le rendement, et l'indice de récolte (Gupta et al., 2001). En considérant le stress hydrique comme le seul stress limitant le rendement du grain, les physiologistes ont défini le rendement comme :

$$GY = W \times WUE \times HI$$

Avec W l'eau absorbée, la WUE l'habilité de la plante à produire de la biomasse par unité d'eau évapotranspirée et HI l'indice de récolte (Passioura, 1977).

A partir de cette équation, les physiologistes préconisent de sélectionner des variétés capables d'absorber plus d'eau, de produire plus de biomasse par unité d'eau absorbée, et d'allouer plus de biomasse vers les grains. Il s'agit alors de sélectionner des caractères phénotypiques tels que des racines profondes par exemple (Richard et al., 2015).

Les stress azotés

La nécessité de réduire les apports d'intrants azotés pour répondre aux enjeux de durabilité environnementale peut amener les cultures à subir des déficits d'alimentation azotée plus ou moins longs. Or, la figure 7 montre que le stress azoté joue un rôle majeur dès le tallage, sur la formation des épis puis du nombre de grains, et également sur le remplissage des grains. Alcoz et al., (1993) a démontré les effets du timing de la fertilisation azotée sur la production de blé en interaction avec la quantité d'azote (N) apportée, et a également mis en évidence une augmentation de l'efficacité d'utilisation de l'azote due au fractionnement des apports. Il a été montré qu'un déficit précoce en N freine l'apparition et la croissance des talles, pouvant ainsi diminuer le nombre d'épis (Masle, 1985). Des stress plus tardifs n'affectent pas le nombre d'épis mais le nombre de grains par épi, du fait

INTRODUCTION

d'avortements floraux, de réductions des initiations des primordia floraux, ou encore des réductions du nombre d'épillets (Langer and Liew, 1973; Whingwiri and Kemp, 1980). Le nombre de grains montre effectivement une corrélation linéaire forte avec la quantité de N dans les épis à floraison. La durée et l'intensité du stress azoté influent sur la constitution du nombre de grains (Jeuffroy and Bouchard, 1999).

La remobilisation de l'azote vers le grain dépend des conditions pendant le remplissage et des variétés. L'absorption et l'assimilation de N pendant la phase de remplissage de la graine sont généralement insuffisantes pour satisfaire la forte demande de la graine. La remobilisation progressive qui se produit dans les différents organes de la plante est nécessaire pour acheminer l'azote vers la graine. La contribution de l'azote des feuilles remobilisé vers l'N des grains représente entre 50 à 90% chez les céréales. L'initiation du processus de sénescence est dépendant de la disponibilité en N, car le maintien de l'absorption retarde la remobilisation. En conditions non limitantes, la quantité remobilisée est directement liée à la quantité absorbée à floraison (Barbottin et al., 2005) (Przulj and Momcilovic 2001 chez l'orge). Le N absorbé après floraison contente les besoins azotés des grains en formation, limitant l'appel de N exercé par les grains sur les feuilles. Lorsque les conditions environnementales sont peu favorables, on observe de meilleures remobilisations (Barbottin et al., 2005). En fonction de la disponibilité en N, l'absorption de N peut être maintenue après floraison. Le contentement des besoins en N des grains limite l'appel de N des grains sur les feuilles, et retarde alors la sénescence foliaire (Thomas and Howarth, 2000). Le contrôle de la sénescence influence alors le rendement et la concentration en protéines des grains. Les variétés à sénescence retardée (de type « stay-green ») ont une durée de photosynthèse plus longue et un rendement plus élevé (Thomas and Howarth, 2000) tandis que celles qui présentent une sénescence précoce ont une bonne efficacité de remobilisation associée à une forte concentration en protéines.

Parmi tous les stress influençant le développement du blé, et notamment, dans les exemples présentés ci-dessus concernant les stress hydriques thermiques et azotés, le timing des stress est essentiel pour comprendre comment l'environnement a impacté les composantes du rendement du blé. La caractérisation d'un stress est alors plus pertinente en prenant en compte le stade de développement auquel il survient.

3.3- Une caractérisation environnementale adaptée au cycle de développement des cultures

Les premières caractérisations environnementales ont utilisé des variables géographiques et climatiques. La pluie et la température ont été des indicateurs fréquemment utilisés, chez le maïs (Hebert et al., 1995) ou le blé dur (Annicchiarico, 2002). L'altitude ainsi que d'autres facteurs

géographiques peuvent également influencer la variation des caractères agronomiques (Balfourier and Charmet, 1991; Annicchiarico, 2002). Brandle and Arthur (1992) ont utilisé des variables météorologiques telles que le nombre de degrés jours (base 5°C), le nombre de jours moyens sans gel, et le déficit hydrique relatif pour caractériser des lieux d'essai. Toutefois, ils ont montré qu'un groupement a priori des environnements à partir des informations géographiques ou climatiques seules aboutissait à une variabilité de rendement intra-groupe légèrement inférieure à celle inter-groupes.

La prise en compte de la phénologie, c'est-à-dire, la chronologie des relations entre les phénomènes climatiques et les stades de développement des plantes, a été surtout développée par Lecomte (2005). Il définit les variables de stress à partir des variables climatiques et pédologiques. L'aspect agronomique dans la définition de ces variables réside dans un premier temps dans la considération des seuils de variables au-delà desquels la plante est considérée en stress. Par exemple, pour caractériser le stress thermique, il considère la somme des températures au-dessus de 25°C. De plus, il décompose ces variables par stade phénologique, afin de pouvoir identifier quel stress à quel moment du développement a impacté la plante. Le défi majeur de cette approche est alors d'identifier parmi la multitude de variables ainsi générée le ou les stress limitant le développement du blé. Lecomte (2005) propose un diagnostic agronomique pour réaliser cette identification. Cette approche a été utilisée par Brancourt-Hulmel (2011) pour expliquer les interactions GxE dans des essais variétaux de blé. Voltas et al. (2005) souligne tout de même la complexité de la sélection de variables représentatives de l'environnement.

Ainsi, l'utilisation de ces variables agronomiques peut être délicate, du fait de la difficulté à prendre en compte de la multiplicité des stress, mais aussi de leur intégration limitée des interactions entre les différents compartiments de l'environnement (plante-sol-atmosphère), ou entre les stress.

3.4- La caractérisation des stress impactant le blé à l'aide des modèles de culture

Des travaux sur l'utilisation de modèle de culture pour la caractérisation environnementale, notamment dans le cas du stress hydrique en Australie (Chenu et al., 2013), démontrent la possibilité de simuler les stress hydriques subis par la plante dans des conditions environnementales données. Les modèles de culture simulent le développement journalier d'un couvert végétal, à partir de données climatiques, pédologiques et de conduite technique. Les modèles de culture sont une représentation simplifiée d'un système, constituée d'un ensemble de relations mathématiques décrivant le fonctionnement du système sol-plante en interaction avec le climat et les opérations techniques. Les modèles mécanistes, en particulier, décrivent et quantifient les mécanismes et les

INTRODUCTION

processus qui interviennent au niveau de la plante et du sol en se basant sur la formalisation d'hypothèses concernant le fonctionnement physiologique et physique du système étudié.

Des modèles ont été développés pour simuler la croissance et le développement d'une culture donnée. Le modèle ARCWHEAT (Weir et al., 1984) a été développé pour fournir un outil d'identification des facteurs limitant la production de blé tendre au Royaume Uni. Il a ensuite évolué pour intégrer un module de bilan hydrique et azoté (Porter, 1984 in Porter et al., 1993). Le modèle SWHEAT a été développé pour simuler la croissance et le développement du blé tendre de printemps en conditions semi-arides (Van Keulen and Seligman, 1987 in Porter et al., 1993). Le modèle SiriusQuality1 (Martre et al., 2006), dérivé de SIRIUS (Jamieson et al., 1998) est spécialisé dans la simulation des différentes fractions protéiques du blé tendre. Le modèle AZODYN (Jeuffroy and Recous, 1999) a été développé pour évaluer différentes options de fertilisation sur le rendement et la teneur en protéines du blé tendre. APSIM (Keating et al., 2003) est un modèle générique, développé pour prendre en compte différentes espèces, en traitant la spécificité de chaque espèce au travers d'options de simulations et paramètres différents pour chaque espèce (Wang et Engel, 2002). Ce modèle a été développé dans l'objectif notamment d'évaluer, non pas uniquement les performances d'un itinéraire technique, mais aussi les performances des systèmes de culture à long terme.

L'utilisation de modèle de culture pour la caractérisation environnementale a été développée par (Muchow et al., 1996; Chapman et al., 2000b). Ils utilisent un modèle chez le sorgho (Hammer and Muchow, 1994) pour définir des indices de sécheresse, qui prennent en compte le timing du stress par rapport à la phénologie de la plante, de manière semblable à Gate (1995) et Lecomte (2005). Une différence majeure réside dans le fait que les indices fournis par les modèles de culture sont régulés par l'évolution de la biomasse de la plante. Ainsi, le calcul de la demande variétale journalière dépend de l'état phénologique et physiologique de la plante. La caractérisation du stress hydrique développée par Muchow et al., (1996) cherche à évaluer le ratio entre la demande et la disponibilité en eau, et s'intéresse à l'évolution de cet index au cours du développement de la plante. Elle reflète ainsi le degré de satisfaction de la demande de la culture en eau. Dans le modèle APSIM, la disponibilité en eau est calculée à partir du volume et de l'humidité du sol explorée par les racines, et la demande de la plante dépend des rayonnements, de la température, et des conditions d'humidité de l'air. Chenu et al. (2013) a ainsi caractérisé les zones de cultures de blé en Australie, en distinguant 3 types de pattern du stress hydrique.

Les modèles de culture constituent des outils de caractérisation environnementale intégratifs du système plante-sol en relation avec l'atmosphère. De plus, parce qu'ils simulent le développement du couvert jour après jour, les modèles de culture fournissent des indicateurs de stress plus

INTRODUCTION

intégratifs dans le temps, et ainsi plus susceptibles de cumuler les effets de différents stress, abiotiques en particulier, au cours de la saison.

Des études similaires ont été menées chez le maïs aux Etats-Unis (Löffler et al., 2005), utilisant un modèle de culture pour définir des macro-environnements. Chez le blé dur, Lacaze and Roumet (2004) ont identifié, à l'aide d'un modèle de culture, deux facteurs limitants de la teneur en protéines dans des conditions agro-climatiques françaises : la sécheresse et la disponibilité en azote entre la montaison et l'épiaison. Dans le cas particulier du stress azoté, la satisfaction d'une plante en azote peut être mesurée à floraison par son indice de nutrition azoté (INN) (Lemaire and Gastal, 1997; Lemaire and Meynard, 1997). Toutefois, la mesure au champ de cet indicateur nécessite un gros investissement de terrain et est coûteuse à analyser. Un autre intérêt des modèles de culture serait de pouvoir modéliser des indicateurs difficiles à mesurer. Une approche qui combine l'utilisation des modèles de cultures et des modèles de sélection génomique a été développée par Heslot et al. (2014). Ils utilisent le modèle de culture SIRIUS pour prédire la phénologie de ses groupes de précocité dans les conditions environnementales testées. Toutefois, les covariables agro-climatiques (Lecomte, 2008) ainsi définies par groupe de précocité, ne sont pas directement générées par le modèle de culture. L'intérêt de développer des indicateurs de stress intégratifs, à partir de simulation des modèles de culture reste à explorer.

Ainsi dans un contexte de changement climatique, la culture du blé va devoir s'adapter au réchauffement et à la sécheresse, tout en limitant les apports d'intrant azotés. La caractérisation de ces stress, qui influent sur différentes composantes du rendement, peut être intégrative du système plante-sol dans son environnement, et aussi, des différents stress survenant au cours de la saison. A partir d'une caractérisation fine des environnements, il serait alors possible de chercher à prédire les réponses des variétés de blé à ces stress. La caractérisation environnementale et la prédiction des réponses aux stress constituent un moyen de s'affranchir de l'impossibilité d'évaluer tout le matériel génétique dans toutes les combinaisons de stress.

4- Problématique

La prédiction génomique des interactions GxE à l'aide d'indicateurs agro-climatiques peut être vue à la fois comme un problème de caractérisation des stress environnementaux pour lesquels on cherche à faire des prédictions, et aussi comme un problème de la grande dimension des données moléculaires que l'on multiplie par le nombre de covariables environnementales.

Pour simplifier la prise en compte de la multiplicité de covariables environnementales, les deux modèles génomiques développés par Jarquín et al. (2014) et Heslot et al. (2014) émettent certaines hypothèses fortes, qui ne sont pas forcément justifiées d'un point de vue biologique, à savoir que les réponses aux covariables environnementales sont modélisées indépendantes de l'effet génétique additif, et avec un poids équivalent quelle que soit la covariable de stress. Avec la modélisation de (Jarquín et al., 2014) en particulier, l'intégration de toutes les covariables environnementales dans une matrice de similarité entre les environnements pour caractériser les interactions GxE n'est pas complètement exploitable dans la mesure où les prédictions sont difficiles à interpréter au sens agronomique. Outre le problème d'interprétation des réponses à plusieurs stress, il faut également tenir compte de la difficulté de prédiction de ces multiples covariables environnementales dans un environnement futur.

Pour ces diverses raisons, l'ensemble de notre travail se déploie dans une approche à l'interface entre la prédiction et la compréhension des interactions.

Nous nous intéresserons dans un premier chapitre à la mise en place d'un modèle de prédiction génomique des interactions GxE, en nous focalisant sur la réponse à un stress environnemental particulier : Après une démarche d'identification des stress environnementaux qui auraient causé des interactions GxE, nous chercherons à prédire à partir de la génomique, les normes de réactions des génotypes à un stress.

Nous étudierons ensuite, dans le deuxième chapitre, la possibilité de caractériser les stress environnementaux, à l'aide d'un modèle de culture, APSIM, afin d'investiguer la possibilité de prendre en compte l'impact de stress multiples et leurs interactions. Notre étude aura pour but d'évaluer (i) la caractérisation des interactions GxE à l'aide du modèle de culture, et (ii) aussi la qualité des prédictions génomiques des normes de réactions à des stress caractérisées par le modèle de culture.

Références

- Alcoz, M.M., F.M. Hons, and V. a. Haby. 1993. Nitrogen Fertilization Timing Effect on Wheat Production, Nitrogen Uptake Efficiency, and Residual Soil Nitrogen. *Agron. J.* 85: 1198–1203.
- Almekinders, C.J.M., N.P. Louwaars, and G.H. Bruijn. 1994. Local seed systems and their importance for an improved seed supply in developing countries. *Euphytica* 78(3): 207–216.
- Annicchiarico, P. 2002. Case study : durum wheat in Algeria. p. 89–104. *In* Genotype X Environment Interactions: Challenges and Opportunities for Plant Breeding and Cultivar Recommendations. Food and Agriculture Organization of the United Nations, Rome.
- Baker, R.J. 1988. Tests for crossover genotype-environmental interactions. *Can. Journal Plant Sci.* 68: 405–410.
- Balfourier, F., and G. Charmet. 1991. Relationships between agronomic characters and ecogeographical factors in a collection of French perennial ryegrass populations. *Agronomie* 11: 645–657 Available at <http://www.agronomy-journal.org/10.1051/agro:19910802>.
- Balfourier, F., J.A. Oliveira, G. Charmet, and E. Arbones. 1997. Factorial regression analysis of genotype by environment interaction in ryegrass populations, using both isozyme and climatic data as covariates. 98: 37–46.
- Barbottin, A., C. Lecomte, C. Bouchard, and M.H. Jeuffroy. 2005. Nitrogen remobilization during grain filling in wheat: Genotypic and environmental effects. *Crop Sci.* 45: 1141–1150.
- Baril, C.P., J. Denis, and P. Brabant. 1994. Selection of environments using simultaneous clustering based on genotype x environment interaction. *Can. J. Plant Sci.* 74: 311–317.
- Battisti, D.S., and R.L. Naylor. 2009. Historical Warnings of Future Food Insecurity with Unprecedented Seasonal Heat. *Science*. 323(5911): 240–244 Available at <http://www.sciencemag.org/cgi/doi/10.1126/science.1164363>.
- Beman, J., K. Arrigo, and P. Matson. 2005. Agricultural runoff fuels large phytoplankton blooms in vulnerable areas of the ocean. *Nature* 434: 211–214.
- Boiffin, J., J. Caneill, J.-M. Meynard, and M. Sebillotte. 1981. Élaboration du rendement et fertilisation azotée du blé d'hiver en Champagne crayeuse I. Protocole et méthode d'étude d'un problème technique régional. *Agronomie* 1: 549–558.
- Brancourt-Hulmel, M. 2011. De l'étude des interactions génotype x milieu chez le blé tendre d'hiver vers la génétique de la production de biomasse par le miscanthus. Habilitation à Diriger des Recherches, INRA, FRA.
- Brandle, J.E., and L.M. Arthur. 1992. A comparative evaluation of two methods of selecting locations used for testing spring wheat cultivars. *Theor. Appl. Genet.* (83): 301–304.
- Brisson, N., P. Gate, D. Gouache, G. Charmet, F.-X. Oury, and F. Huard. 2010. Why are wheat yields stagnating in Europe? A comprehensive data analysis for France. *F. Crop. Res.* 119(1): 201–212 Available at <http://linkinghub.elsevier.com/retrieve/pii/S0378429010001929>.

INTRODUCTION

- Burgueño, J., G. de los Campos, K. Weigel, and J. Crossa. 2012. Genomic Prediction of Breeding Values when Modeling Genotype \times Environment Interaction using Pedigree and Dense Molecular Markers. *Crop Sci.* 52(2): 707 Available at <https://www.crops.org/publications/cs/abstracts/52/2/707>.
- Byerlee, D., and T. Husain. 1993. Agricultural research strategies for favoured and marginal areas - the experience of farming system research in Pakistan. *Expl. Agric* 29: 155–171.
- Byerlee, D., and M. Morris. 1993. Research for marginal environments: are we underinvested? *Food Policy* 18(5): 381–393.
- Byth, D.E., R.L. Eisemann and, and I.H. De Lacy. 1976. Two-way pattern analysis of a large data set to evaluate genotype adaptation. *Heredity (Edinb)*. 37: 189–201.
- Calderini, D.F., M.P. Reynolds, and G.A. Slafer. 2006. Source–sink effects on grain weight of bread wheat, durum wheat, and triticale at different locations. *Crop Pasture Sci.* 57(2): 227–233.
- Calderini, D.F., R. Savin, L.G. Abeledo, M.P. Reynolds, and G.A. Slafer. 2001. The importance of the period immediately preceding anthesis for grain weight determination in wheat. *Euphytica* 119(1): 199–204.
- Ceccarelli, S. 1989. Wide adaptation: How wide? *Euphytica* 40(3): 197–205.
- Ceccarelli, S., W. Erskine, J. Hamblin, and S. Grando. 1994. Genotype by Environment Interaction and International Breeding Programmes. *Exp. Agric.* 30(2): 177–187.
- Chapman, S.C., M. Cooper, G.L. Hammer, and D.G. Butler. 2000a. Genotype by environment interactions affecting grain sorghum. II. Frequencies of different seasonal patterns of drought stress are related to location effects on hybrid yields. *Aust. J. Agric. Res.* 51(2): 209 Available at <http://www.publish.csiro.au/journals/abstractHTML.cfm?J=AR&V=51&I=2&F=AR99021abs.XML>.
- Chapman, S.C., G.L. Hammer, D.G. Butler, and M. Cooper. 2000b. Genotype by environment interactions affecting grain sorghum. Temporal sequences and spatial patterns in the target population of environments. *Aust. J. Agric. Res.* 51(2): 223 Available at <http://www.publish.csiro.au/journals/abstractHTML.cfm?J=AR&V=51&I=2&F=AR99022abs.XML>.
- Chenu, K., R. Dehifard, and S.C. Chapman. 2013. Large-scale characterization of drought pattern: a continent-wide modelling approach applied to the Australian wheatbelt--spatial and temporal trends. *New Phytol.* 198(3): 801–20 Available at <http://www.ncbi.nlm.nih.gov/pubmed/23425331>.
- Crossa, J., G.D.L. Campos, P. Pérez, D. Gianola, J. Burgueño, J.L. Araus, D. Makumbi, R.P. Singh, S. Dreisigacker, J. Yan, V. Arief, M. Banziger, and H.-J. Braun. 2010. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186(2): 713–24 Available at <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2954475&tool=pmcentrez&render type=abstract>.
- Denis, J.B. 1980. Analyse de la régression factorielle. *Biom Praxim* 20: 1–34.

INTRODUCTION

- Dias, A.S., and F.C. Lidon. 2009. Evaluation of Grain Filling Rate and Duration in Bread and Durum Wheat, under Heat Stress after Anthesis. *J. Agron. Crop Sci.* 195: 137–147.
- Donald, C.M. 1968. The breeding of crop ideotypes. *Euphytica* 17(3): 385–403.
- Eisemann, R.L., M. Cooper, and D.R. Woodruff. 1990. Beyond the analytic methodology: better interpretation and exploitation of GE interactions in plant breeding. p. 108–117. *In* Kang, M.S. (ed.), *Genotype- by-environment interactions and plant breeding*. Louisiana State University Agri. Center, Baton Rouge, LA.
- Eyzaguirre, P., and M. Iwanaga. 1996. Participatory plant breeding. *In* Proceedings of a workshop on participatory plant breeding. Rome, IPGRI, Wageningen, the Netherlands.
- Falconer, D.S. 1952. The Problem of Environment and Selection. *Am. Nat.* 86(830): 293–298 Available at <http://www.journals.uchicago.edu/doi/10.1086/281736>.
- Falconer, D.S. 1990. Selection in different environments: effects on environmental sensitivity (reaction norm) and on mean performance. *Genet. Res.* 56(1): 57–70.
- Farooq, M., H. Bramley, J.A. Palta, and K.H.M. Siddique. 2011. Heat Stress in Wheat during Reproductive and Grain-Filling Phases. *CRC. Crit. Rev. Plant Sci.* 30(6): 491–507.
- Finlay, K.W., and G.N. Wilkinson. 1963. The analysis of adaptation in a plant-breeding program. *Aust. J. Agric. Res.* 14(6): 742–754.
- Fischer, R.A. 1985. Number of kernels in wheat crops and the influence of solar radiation and temperature. *The J. Agric. Sci.* 105: 447–461.
- Fisher, R.A., and W.A. Mackenzie. 1923. Studies in crop variation. II. The manurial response of different potato varieties. *J. Agric. Sci.* 13(3): 311–320.
- Gabriel, K.R. 1978. Least Squares Approximation of Matrices by Additive and Multiplicative Models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 40(2): 186–196.
- Gate, P. 1995. *Ecophysiologie du blé. Technique et Documentation Lavoisier*.
- Gauch, H.G., and R.W. Zobel. 1988. Predictive and postdictive success of statistical analyses of yield trials. *Theor. Appl. Genet.* 76(1): 1–10.
- Gibson, L.R., and G.M. Paulsen. 1999. Yield Components of Wheat Grown under High Temperature Stress. 39: 1841–1846.
- GIEC. 2014. *Climate Change 2014 Synthesis Report*.
- Gollob, H.F. 1968. A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika* 33(1): 73–115.
- Gupta, N., S. Gupta, and A. Kumar. 2001. Effect of water stress on physiological attributes and their relationship with growth and yield of wheat cultivars at different stages. *J. Agron. Crop Sci.* 186: 55–62 Available at <http://onlinelibrary.wiley.com/doi/10.1046/j.1439-037x.2001.00457.x/full>.

INTRODUCTION

- Habier, D., R.L. Fernando, and D.J. Garrick. 2013. Genomic BLUP Decoded: A Look into the Black Box of Genomic Prediction. *Genetics* 194(3): 597–607 Available at <http://www.ncbi.nlm.nih.gov/pubmed/23640517>.
- Hammer, G.L., and R.C. Muchow. 1994. Assessing climatic risk to sorghum production in water-limited subtropical environments I. Development and testing of a simulation model. *F. Crop. Res.* 36(3): 221–234.
- Hebert, Y., C. Plomion, and N. Harzic. 1995. Genotype x environment interaction for root traits in maize, as analysed with factorial regression models. *Euphytica* 81: 85–92.
- Henderson, C.R. 1975. Best Linear Unbiased Estimation and Prediction under a Selection Model. *Biometrics* 31(2): 423 Available at <http://www.ncbi.nlm.nih.gov/pubmed/1174616>.
- Heslot, N., D. Akdemir, M.E. Sorrells, and J.-L. Jannink. 2014. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theor. Appl. Genet.* 127(2): 463–480 Available at <http://link.springer.com/10.1007/s00122-013-2231-5>.
- Hirel, B., J. Le Gouis, B. Ney, and A. Gallais. 2007. The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches. *J. Exp. Bot.* 58(9): 2369–87 Available at <http://www.ncbi.nlm.nih.gov/pubmed/17556767>.
- Jamieson, P.D., M.A. Semenov, I.R. Brooking, and G.S. Francis. 1998. Sirius: a mechanistic model of wheat response to environmental variation. *Eur. J. Agron.* 8: 161–179.
- Jarquín, D., J. Crossa, X. Lacaze, P. Du Cheyron, J. Daucourt, J. Lorgeou, F. Piraux, L. Guerreiro, P. Pérez, M. Calus, J. Burgueño, and G. de los Campos. 2014. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* 127(3): 595–607 Available at <http://link.springer.com/10.1007/s00122-013-2243-1>.
- Jeuffroy, M.-H., and C. Bouchard. 1999. Intensity and Duration of Nitrogen Deficiency on Wheat Grain Number. *Crop Sci.* 39: 1385–1393.
- Jeuffroy, M.H., and S. Recous. 1999. Azodyn: A simple model simulating the date of nitrogen deficiency for decision support in wheat fertilization. *Eur. J. Agron.* 10(2): 129–144.
- Keating, B. a., P.S. Carberry, G.L. Hammer, M.E. Probert, M.J. Robertson, D. Holzworth, N.I. Huth, J.N.G. Hargreaves, H. Meinke, Z. Hochman, G. McLean, K. Verburg, V. Snow, J.P. Dimes, M. Silburn, E. Wang, S. Brown, K.L. Bristow, S. Asseng, S. Chapman, R.L. McCown, D.M. Freebairn, and C.J. Smith. 2003. An overview of APSIM, a model designed for farming systems simulation. *Eur. J. Agron.* 18(3-4): 267–288.
- Kempton, R.A. 1984. The use of biplots in interpreting variety by environment interactions. *J. Agric. Sci.* 103(1): 123–135.
- Kirby, E. 1981. Cereal Development Guide (A Barron, Ed.). National Agriculture Center - Cereal Unit.
- Kolmodin, R., E. Strandberg, P. Madsen, J. Jensen, and H. Jorjani. 2002. Genotype by Environment Interaction in Nordic Dairy Cattle Studied Using Reaction Norms. *Acta Agric. Scand. Sect. A* -

INTRODUCTION

- Anim. Sci. 52(1): 11–24 Available at
<http://www.tandfonline.com/doi/abs/10.1080/09064700252806380>.
- Lacaze, X., and P. Roumet. 2004. Environment characterisation for the interpretation of environmental effect and genotype x environment interaction. *Theor. Appl. Genet.* 109(8): 1632–40 Available at <http://www.ncbi.nlm.nih.gov/pubmed/15372155>.
- Lai, K., M.T. Lorenc, H.C. Lee, P.J. Berkman, P.E. Bayer, P. Visendi, P. Ruperao, T.L. Fitzgerald, M. Zander, C.-K.K. Chan, S. Manoli, J. Stiller, J. Batley, and D. Edwards. 2015. Identification and characterization of more than 4 million intervarietal SNPs across the group 7 chromosomes of bread wheat. *Plant Biotechnol. J.* 13(1): 97–104 Available at
<http://doi.wiley.com/10.1111/pbi.12240>.
- Langer, R.H.M., and F.K.Y. Liew. 1973. Effects of varying nitrogen supply at different stages of the reproductive phase on spikelet and grain production and on grain nitrogen in wheat. *Aust. J. Agric. Res.* 24: 647–656.
- Lawlor, T.J., S. Tsuruta, L. Klei, and I. Misztal. 2002. Use of a random regression model to investigate changes in genetic parameters over time. p. 1–4. *In* Proceedings of the 7th World Congress on Genetics Applied to Livestock Production. Montpellier, France.
- Lecomte, C. 2008. L'évaluation expérimentale des innovations variétales Proposition d'outils d'analyse de l'interaction génotype-milieu adaptés à la diversité des besoins et des contraintes des acteurs de la filière semences. Available at <http://prodinra.inra.fr/record/244938>.
- Lemaire, G., and F. Gastal. 1997. N Uptake and Distribution in Plant Canopies. p. 3–43. *In* Diagnosis of the Nitrogen Status in Crops. Springer Berlin Heidelberg, Heidelberg, Germany.
- Lemaire, G., and J.-M. Meynard. 1997. Use of the Nitrogen Nutrition Index for the Analysis of Agronomical Data. p. 44–55. *In* Lemaire, G. (ed.), Diagnosis of the Nitrogen Status in Crops. Springer Berlin Heidelberg.
- Löffler, C.M., J. Wei, T. Fast, J. Gogerty, S. Langton, M. Bergman, B. Merrill, and M. Cooper. 2005. Classification of Maize Environments Using Crop Simulation and Geographic Information Systems. *Crop Sci.* 45(5): 1708 Available at
<https://www.crops.org/publications/cs/abstracts/45/5/1708>.
- Malosetti, M., J.M. Ribaut, and F.A. van Eeuwijk. 2013. The statistical analysis of multi-environment data: Modeling genotype-by-environment interaction and its genetic basis. *Front. Physiol.* 4: 44 Available at
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3594989&tool=pmcentrez&render type=abstract>.
- Malosetti, M., J. Voltas, I. Romagosa, S.E. Ullrich, and F.A. Van Eeuwijk. 2004. Mixed models including environmental covariables for studying QTL by environment interaction. *Euphytica* 137: 139–145.
- Mandel, J. 1969. The partitioning of interaction in analysis of variance. *J. Res. Natl. Bur. Stand. Ser. B* 73(4): 309–328.

INTRODUCTION

- Des Marais, D.L., K.M. Hernandez, and T.E. Juenger. 2013. Genotype-by-Environment Interaction and Plasticity: Exploring Genomic Responses of Plants to the Abiotic Environment. *Annu. Rev. Ecol. Evol. Syst.* 44(1): 5–29 Available at <http://www.annualreviews.org/doi/abs/10.1146/annurev-ecolsys-110512-135806>.
- Martre, P., P.D. Jamieson, M.A. Semenov, R.F. Zyskowski, J.R. Porter, and E. Triboi. 2006. Modelling protein content and composition in relation to crop nitrogen dynamics for wheat. *Eur. J. Agron.* 25(2): 138–154 Available at <http://linkinghub.elsevier.com/retrieve/pii/S1161030106000451>.
- Masle, J. 1985. Elaboration du nombre de grains potentiel d'un peuplement de blé d'hiver. p. 71:857–869. *In* CR Académie Agriculture. France.
- Meuwissen, T.H., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157(4): 1819–1829 Available at http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1461589&tool=pmcentrez&render_type=abstract.
- Meynard, J.M., and M. Sebillotte. 1994. L'élaboration du rendement du blé, base pour l'étude des autres céréales à talles. p. 31–51. *In* Combes, L., Picard, D. (eds.), *Elaboration du rendement des principales cultures annuelles*. INRA Editi. Paris, FRA.
- Mooers, C.A. 1921. The agronomic placement of varieties. *J. Am. Soc. Agron.* 13(9): 337–352.
- Muchow, R., M. Cooper, and G. Hammer. 1996. Characterizing environmental challenges using models. p. 349–364. *In* Cooper, M., Hammer, G. (eds.), *Plant adaptation and crop improvement*. CAB International: Wallingford, UK.
- Nilsson-Ehle, H. 1914. Vilka erfarenheter hava hittills vunnits rörande möjligheten av växters acklimatisering. *Kgl. Landtbruks-Akad. Handl. Tidskr* 53: 537–572.
- Ortiz, R., R. Trethowan, G.O. Ferrara, M. Iwanaga, J.H. Dodds, J.H. Crouch, J. Crossa, and H.-J. Braun. 2007. High yield potential, shuttle breeding, genetic diversity, and a new international wheat improvement strategy. *Euphytica* 157(3): 365–384 Available at <http://link.springer.com/10.1007/s10681-007-9375-9>.
- Passioura, J.B. 1977. Grain yield, harvest index, and water use of wheat. *J. Aust. Inst. Agric. Sci.* 43: 117–120.
- Perkins, J.M. 1972. The principal component analysis of genotype-environmental interactions and physical measures of the environment. *Heredity (Edinb.)* 29(1): 51–70 Available at <http://www.nature.com/doi/10.1038/hdy.1972.64>.
- Porter, J.R., and M. Gawith. 1999. Temperatures and the growth and development of wheat: a review. *Eur. J. Agron.* 10(1): 23–36 Available at <http://linkinghub.elsevier.com/retrieve/pii/S1161030198000471>.
- Porter, J.R., P.D. Jamieson, and D.R. Wilson. 1993. Comparison of the wheat simulation models Afrcwheat2, Ceres-wheat and Swheat for non-limiting conditions of crop growth. *F. Crop. Res.* 33(1-2): 131–157 Available at <http://linkinghub.elsevier.com/retrieve/pii/0378429093900988>.

INTRODUCTION

- Ravagnolo, O., and I. Misztal. 2000. Genetic component of heat stress in dairy cattle, parameter estimation. *J. Dairy Sci.* 83(9): 2126–30 Available at <http://www.ncbi.nlm.nih.gov/pubmed/11003247>.
- Richard, C.A., L.T. Hickey, S. Fletcher, R. Jennings, K. Chenu, and J.T. Christopher. 2015. High-throughput phenotyping of seminal root traits in wheat. *Plant Methods* 11: 1–11 Available at <http://www.plantmethods.com/content/11/1/13>.
- Rijven, a H. 1986. Heat inactivation of starch synthase in wheat endosperm tissue. *Plant Physiol.* 81(2): 448–453.
- Römer, T. 1917. Sind die ertragreichen Sorten ertragssicherer? *Mitteilungen der DLG* 32: 87–89.
- Sae-Lim, P., H. Komen, A. Kause, and H. a Mulder. 2014. Identifying environmental variables explaining genotype-by-environment interaction for body weight of rainbow trout (*Onchorynchus mykiss*): reaction norm and factor analytic models. *Genet. Sel. Evol.* 46(1): 16 Available at <http://www.gsejournal.org/content/46/1/16>.
- Saini, H.S., and D. Aspinall. 1982. Abnormal Sporogenesis in Wheat (*Triticum aestivum* L.) Induced by Short Periods of High Temperature. *Ann. Bot.* 49: 835–846.
- Segura, V., B.J. Vilhjálmsson, A. Platt, A. Korte, Ü. Seren, Q. Long, and M. Nordborg. 2012. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat. Genet.* 44(7): 825–830 Available at <http://www.nature.com/doifinder/10.1038/ng.2314>.
- Smith, A.B., B.R. Cullis, and R. Thompson. 2005. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. *J. Agric. Sci.* 143(6): 449 Available at http://www.journals.cambridge.org/abstract_S0021859605005587.
- Stone, P., and M. Nicolas. 1994. Wheat Cultivars Vary Widely in Their Responses of Grain Yield and Quality to Short Periods of Post-Anthesis Heat Stress. *Funct. plant Biol.* 21(6): 887–900.
- Stringfield, G.H., and R.M. Salter. 1934. Differential response of corn varieties to fertility levels and to seasons. *J. Agric. Res.* 49(11): 991–1000.
- Tesemma, T., and E. Bechere. 1998. Developing elite durum wheat landrace selections (composites) for Ethiopian peasant farm use: raising productivity while keeping diversity alive. *Euphytica* 102: 323–328 Available at <http://link.springer.com/article/10.1023/A:1018360432426>.
- Thomas, H., and C.J. Howarth. 2000. Five ways to stay green. *J. Exp. Bot.* 51 Spec No(February): 329–337.
- Trnka, M., R.P. Rötter, M. Ruiz-Ramos, K.C. Kersebaum, J.E. Olesen, Z. Žalud, and M.A. Semenov. 2014. Adverse weather conditions for European wheat production will become more frequent with climate change. *Nat. Clim. Chang.* 4(7): 637–643 Available at <http://www.nature.com/doifinder/10.1038/nclimate2242>.
- VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91(11): 4414–23 Available at <http://www.ncbi.nlm.nih.gov/pubmed/18946147>.

INTRODUCTION

- Vargas, M., J. Crossa, F.A. van Eeuwijk, M.E. Ramírez, and K. Sayre. 1999. Using Partial Least Squares Regression, Factorial Regression, and AMMI Models for Interpreting Genotype \times Environment Interaction. *Crop Sci.* 39(4): 955 Available at <https://www.crops.org/publications/cs/abstracts/39/4/CS0390040955>.
- Voltas, J., H. López-Córcoles, and G. Borrás. 2005. Use of biplot analysis and factorial regression for the investigation of superior genotypes in multi-environment trials. *Eur. J. Agron.* 22(3): 309–324.
- Weir, A.H., P.L. Bragg, J.R. Porter, and J.H. Rayner. 1984. A winter wheat crop simulation model without water or nutrient limitations. *J. Agric. Sci.* 102(2): 371–382.
- Whingwiri, E., and D. Kemp. 1980. Spikelet development and grain yield of the wheat ear in response to applied nitrogen. *Aust. J. Agric. Res.* 31(4): 637.
- Williams, E.J. 1952. The Interpretation of Interactions in Factorial Experiments. *Biometrika* 39(1/2): 65.
- Yang, J., B. Benyamin, B.P. McEvoy, S. Gordon, A.K. Henders, D.R. Nyholt, P. a Madden, A.C. Heath, N.G. Martin, G.W. Montgomery, M.E. Goddard, and P.M. Visscher. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42(7): 565–569 Available at <http://www.nature.com/doifinder/10.1038/ng.608>.
- Yates, F., and W.G. Cochran. 1938. The analysis of groups of experiments. *J. Agric. Sci.* 28(4): 556–580.

CHAPITRE 1

**Prédictions génomiques des normes de réactions à un stress
environnemental**

TABLE DES MATIERES : CHAPITRE 1

1.1- Choix des stress environnementaux pour l'étude des normes de réactions.....	55
1.2 - Publication 1	59
Abstract.....	60
Introduction.....	61
Materials and Methods	62
Experimental data	62
Statistical models for genomic predictions	63
Importance of the random regression to an environmental covariate effect.....	66
Assessment of the prediction quality of the models.....	66
Simulation of Experimental data	67
Results.....	67
Different patterns of reaction norms of grain number to nitrogen or to water deficit stress .	67
Modeling the genomic random regression to water deficit stress improve prediction accuracies	69
Influence of the variance of the genomic random regression to the environmental covariate	69
Discussion.....	72
Improving Genomic Predictions by modeling the response to an environmental stress	72
Limits of the FR-g BLUP	73
A step toward exploiting genomic GxE interactions in genomic selection	73
Applications for breeders and advisers.....	74
Conclusions.....	75
References.....	77
1.3- Equivalence entre le modèle FR-gBLUP et le modèle de Jarquin et al. (2014)	84
Conclusion du chapitre 1.....	87
Références complémentaires.....	88

1.1- Choix des stress environnementaux pour l'étude des normes de réactions

La prédiction génomique des normes de réaction à un stress environnemental nécessite d'avoir identifié au préalable les stress qui ont non seulement impacté l'élaboration du rendement dans les essais étudiés, et qui ont également généré de l'interaction GxE.

Dans notre étude, nous avons choisi de simplifier le problème de prédiction en nous intéressant plus particulièrement au nombre de grains par m². Plutôt que de chercher à prédire le rendement en lui-même, qui est un caractère beaucoup plus complexe, nous nous sommes focalisés sur les phases pré-floraison de l'élaboration du rendement. De plus, dans les essais étudiés, deux traitements azotés sont comparés. La différence entre ces traitements, notamment pour ce qui est du moment des applications, impacterait a priori plutôt le nombre de grains.

2008						2009				
Lieu	Villiers-le-Bâcle			Estrées-Mons			Vraux		Estrées-Mons	
Traitement										
Azoté	HN	LN	HN+	HN	LN	HN	LN	LN-	HN	LN
(kg/ha)										
Stade application										
Sortie Hiver	-	-	-	50	70	60	60	-	50	50
Epi 1cm	66.5	44	69	70	-	100	60	-	50	-
2 nœuds	60	-	-	50	-	60	-	-	50	-
Floraison	-	-	60	-	-	-	-	-	-	-

Table 1 : Plan expérimental (Apport HN : Fort apport d'intrants, LN : Faibles apport d'intrants)

Pour discerner les stress à examiner, nous avons ainsi considéré a priori le stress azoté et nous avons réalisé une identification des facteurs limitants à partir de 2 géotypes révélateurs, Soissons et Camp-Rémy (Brancourt-hulmel et al., 1999).

Nous avons réalisé une régression du nombre de grains au nombre de grains seuil par rapport aux covariables environnementales décrites en Table (2).

$$y_{jk} = \mu + w_j + \varepsilon_{jk} \quad (\text{Modèle 1})$$

Où y_{jk} représente le nombre de grains, de la répétition k dans l'environnement j , μ la moyenne générale, w_j la covariable environnementale dans l'environnement j , et ε_{jk} la résiduelle.

Covariable environnementale	R2 Soissons	R2 Camp Remy
ExcesH2O_Jun	0.39	0.18
Sum_PETP_Avr	0.39	0.15
Sum_PETP_May	0.33	0.12
Sum_PETP_Jun	0	0
Deficit_Rg_Avr	0.31	0.13
Deficit_Rg_May	0.21	0.07
Deficit_Rg_June	0.45	0.18
Sum_Rg_Avr	0.23	0.09
Sum_Rg_May	0.18	0.06
Sum_Rg_June	0.43	0.16
Sum_Tmax25_May	0.17	0.06
Sum_Tmax25_June	0.3	0.14

Table 2 : Coefficients de détermination (R2) des régressions des écarts au nombre de grain seuil en fonction des covariables environnementales une à une. Les covariables ont été déclinées par mois en faisant l'approximation grossière que chaque mois correspond à un stade de développement (Avril : montaison, Mai: méiose-floraison, Juin: remplissage). Les stress environnementaux ont été définis par : ExcesH2O, excès d'eau ; Sum_PETP, somme des différences entre les apports d'eau par la pluie (P) et la demande climatique définie par l'évapo-transpiration potentielle (ETP) ; Deficit_Rg, nombre de jours où le rayonnement global a été inférieur à 1045J/cm², Sum_Rg, somme des rayonnements, Sum_Tmax25, somme des températures maximales supérieures à 25°C.

Toutefois, une covariable environnementale expliquant fortement les écarts au nombre de grain seuil, chez Soissons, et/ou chez Camp Remy, ne génère pas forcément d'interaction GxE. L'importance des interactions entre les génotypes et une covariable environnementale en particulier a été évaluée sur l'ensemble des 159 génotypes à partir du modèle suivant :

$$y_{ijk} = \mu + \beta w_j + g_i + e_j + b_i w_j + ge_{ij} + \varepsilon_{ijk} \quad (\text{Modèle 2})$$

Où y_{ijk} représente le nombre de grains, β est la régression à la covariable environnementale w_j , g_i est l'effet aléatoire du génotype, et suit une distribution normale $N(0, \sigma_g^2)$, e_j l'effet aléatoire de l'environnement, non expliqué par la régression à w_j et suit une loi $N(0, \sigma_e^2)$. b_i est un effet aléatoire qui représente les écarts des régressions génotypiques à la covariable w_j par rapport à la pente de régression moyenne β . Les b_i suivent une distribution $N(0, \sigma_b^2)$. ge_{ij} l'effet aléatoire des interactions GxE non expliquées par les régressions $b_i w_j$, et suivant une loi $N(0, \sigma_{ge}^2)$, et ε_{ijk} représente le terme résiduel de loi $N(0, \sigma^2)$.

Les composantes de la variance de ce modèle sont détaillées, pour chaque covariable environnementale, ci-dessous (Table 3).

Covariable environnementale	σ_g^2	σ_e^2	σ_b^2	σ_{ge}^2	σ^2
Quotient_thrmq	2.98 +/- 0.35	8.21 +/- 4.11	0.44 +/- 0.06	0.44 +/- 0.06	1.38 +/- 0.05
ExcesH2O_Jun	2.94 +/- 0.34	8.61 +/- 4.31	0.4 +/- 0.06	0.51 +/- 0.06	1.39 +/- 0.05
Sum_PETP_Avr	2.98 +/- 0.34	9.19 +/- 4.6	0.81 +/- 0.11	0.42 +/- 0.06	1.35 +/- 0.05
Sum_PETP_May	2.97 +/- 0.34	9.5 +/- 4.76	0.82 +/- 0.12	0.49 +/- 0.06	1.34 +/- 0.05
Sum_PETP_Jun	2.9 +/- 0.34	10.41 +/- 5.21	0 +/- 0	1.01 +/- 0.07	1.34 +/- 0.05
Deficit_Rg_Avr	2.96 +/- 0.34	8.31 +/- 4.16	0.48 +/- 0.07	0.42 +/- 0.06	1.38 +/- 0.05
Deficit_Rg_May	2.9 +/- 0.34	10.54 +/- 5.27	0 +/- 0	1.01 +/- 0.07	1.34 +/- 0.05
Deficit_Rg_June	2.9 +/- 0.34	10.21 +/- 5.11	0 +/- 0	1.01 +/- 0.07	1.34 +/- 0.05
Sum_Rg_Avr	2.93 +/- 0.34	8.9 +/- 4.45	0.23 +/- 0.05	0.72 +/- 0.07	1.37 +/- 0.05
Sum_Rg_May	2.94 +/- 0.34	10.11 +/- 5.06	0.55 +/- 0.1	0.73 +/- 0.07	1.32 +/- 0.05
Sum_Rg_June	3 +/- 0.35	8.77 +/- 4.39	0.68 +/- 0.09	0.41 +/- 0.06	1.36 +/- 0.05
Sum_Tmax25_May	2.86 +/- 0.34	10.34 +/- 5.17	0.49 +/- 0.09	0.77 +/- 0.07	1.32 +/- 0.05
Sum_Tmax25_June	2.82 +/- 0.33	9.72 +/- 4.86	0.47 +/- 0.08	0.61 +/- 0.06	1.35 +/- 0.05

Table 3 : Composantes de la variance pour le nombre de grains et de la réponse à chaque covariable à partir du Modèle 2.

Le choix de la covariable environnementale d'étude pour les normes de réaction prend en compte plusieurs facteurs. Il est basé i) sur les R2 des régressions des écarts au nombre de grains (Table 2), ii) sur la possibilité éco-physiologique de l'impact du stress climatique sur le caractère considéré en fonction des dates de stress, par rapport aux facteurs limitants connus au cours du développement du blé (Figure 7- Introduction générale), iii) sur la vérification du sens biologique de la régression (plus la covariable environnementale indique un stress, plus la composante du rendement diminue), et iv) également sur la présence d'interaction entre les génotypes et la covariable (Table 3).

CHAPITRE 1

Du fait de la connaissance a priori que le stress hydrique au cours du tallage et au début de la montaison est connu pour affecter la mise en place des talles et leur régression (Figure 7-Introduction générale), et au vu des résultats des régressions (Tables 2 et 3), notre étude des normes de réactions portera sur la somme des déficits hydriques en avril ($\text{Sum PETP} < 0$ April)

1.2 - Publication 1

Genomic random regression for adaptation breeding: Whole-genome prediction of the reaction norms to environmental stress in bread wheat (*Triticum aestivum* L.)

Delphine Ly*, Sylvie Huet[†], Arnaud Gauffreteau[‡], Fabien Cormier^{*,§}, Renaud Rincant*, Jean-Luc Jannink** and Gilles Charmet^{*,1}

*Institut National de la Recherche Agronomique - Blaise Pascal University, UMR1095, Genetics, Diversity and Ecophysiology of Cereals, 63039 Clermont-Ferrand, France

[†] Institut National de la Recherche Agronomique, MAIAGE – Applied Mathematics for Computer Sciences, Jouy-en-Josas, France

[‡] Institut National de la Recherche Agronomique, UMR Agronomy, Thiverval Grignon, France

[§] Biogemma, Research center, Route d'Ennezat, Chappes, France

* USDA-ARS, R.W. Holley Center, 538 Tower Rd, and Cornell University, Section of Plant Breeding and Genetics, Ithaca, NY

¹**Corresponding author: Gilles Charmet**, INRA UMR GDEC, 234 av du Brézat, 63 100 Clermont-Ferrand, France, 33 +4 73 62 43 09, Gilles.Charmet@clermont.inra.fr

Accepted in *Theoretical and Applied Genetics*

Abstract

In a climate change context, breeding for adaptation to environmental stresses has emerged as a major priority. Factorial Regression models the genotypic response to environmental stress covariates and seems appropriate to predict adaptation. The contribution of genomics could enhance the predictive ability of the factorial regression model, allowing predictions of untested individuals' adaptation to untested environments. We propose a genomic random regression as a genomic extension of the factorial regression model: the Factorial Regression genomic Best Linear Unbiased Predictor (FR-gBLUP) models the genomic linear reaction norms to an environmental stress. The predictive ability of the standard additive genomic BLUP (gBLUP) and the FR-gBLUP for the response of grain number to nitrogen stress or to water deficit stress models were compared in 4 cross-validation schemes, which simulated different combinations of tested or untested individuals in tested or untested environments. Comparing prediction accuracies of the additive gBLUP and the FR-gBLUP models for the genomic regression to water stress deficit, we showed accuracy gains from +3.5 to 15.4% on real data. To predict performances of individuals in new environments, the FR-gBLUP model was consistently more accurate than the additive gBLUP. Furthermore, we showed by simulation the variance of the genomic regression to the environmental covariate strongly affected prediction accuracies of the FR-gBLUP. Simulation results showed indeed that greater gains of accuracy were achieved when the variance component of the genomic regression to the stress covariate was high. Our study broadens the possibilities of adaptation breeding to genome-wide selection for adaptation.

Key words: genotype x environment interaction, Factorial Regression, Genomic Prediction, Random regression, Reaction norm, Water deficit and Nitrogen stresses

Key message: We developed a genomic random regression model to predict the adaptation to an environmental stress. This model showed better prediction accuracies than the genomic additive model.

Introduction

From a plant evolutionary perspective, the multiplicity of phenotypes a single plant genotype can produce in response to the environment is dazzling. Phenotypic plasticity reflects the plant ability to sense, respond to and survive a diversity of abiotic stresses (Bohnert et al., 1995; Des Marais et al., 2013). One way to assess the importance of genetic variation in plasticity is to measure the genotype-by-environment interaction (G×E). In selection, the existence of large G×E interactions is noteworthy particularly when it results in a change in the ranking of genotype across environments (Haldane, 1946). The definition of a superior individual then becomes conditional upon the environment in which the individuals are tested, and thus, conditional upon the stresses to which the genotypes have to respond. Particularly in a context of changing climate and nitrogen fertilizer reduction policies – causing, among others, wheat yields stagnation in Europe (Brisson et al., 2010), we need greater insight into the genetics of adaptation to the environment (Via et al., 1995). For breeding purposes, indicators have been developed to integrate the stress response information, either to select the most stable genotypes or to differentiate the most adapted ones (Finlay and Wilkinson, 1963; Fischer and Maurer, 1978; Lin and Binns, 1988). To dissect the plastic responses to climate change, studies have identified functional traits and explored the genetics underlying the response to a particular stimulus (Violle et al., 2007; Nicotra et al., 2010; Parchman et al., 2012; Weinig et al., 2014), for instance in wheat physiology and genetics of nitrogen use efficiency (Hirel et al., 2007; Habash et al., 2007).

A common statistical analysis to model and predict response to an environmental stress is the factorial regression analysis (Denis, 1980; Denis et al., 1997), which also provides insights into G×E interactions. The factorial regression model consists of a linear regression of the genotypes on environmental covariates. It represents the genotype-by-environment interaction term as sensitivities of each genotype to environmental stress covariates. In factorial regression, the genotypic sensitivities depict linear reaction norms to the environmental covariates. Factorial regression presents advantages concerning its potential to predict untested environments and to predict local adaptation (Smith et al., 2005). Ravagnolo and Misztal (2000) proposed a random regression model of the factorial regression in dairy cattle. In their random regression, they modeled the genotypic sensitivities to heat stress as a random effect with a covariance proportional to the pedigree-based relationship. With the recent development of molecular markers, factorial regression studies have been applied to predict untested genotypes by modeling different effects of alleles at a QTL when regressed on environmental covariates (Malosetti et al. 2004). In wheat, a multi-environment quantitative trait loci (QTL) model has identified QTL by environment interaction for anthesis in drought stress trials, emphasizing the fact that marker alleles may have different effects

in different environments (Mathews et al., 2008). With faster, cheaper and more considerable genome-wide marker information available, genomic selection has been developed to use simultaneously all markers across the genome to predict quantitative traits (Meuwissen et al., 2001). Currently, genomic selection models have mostly been developed to predict additive genetic values, thus ignoring interactions with environments. Some genomic models have been proposed to take into account GxE interaction (Jarquín et al., 2014b; Heslot et al., 2014). They modeled GxE interaction as a random effect with a covariance structure that accounts for not only the relationship between individuals based on their markers, but also for the covariance between environments (Burgueño et al., 2012b; Crossa et al., 2013). Models accounting for marker-by-environment interactions (Schulz-Streeck et al., 2013) with a pre-selection of the markers have performed with predictive abilities similar to genotype-by-environment models. Nevertheless, genomic predictions of GxE or marker-by-environment interactions prediction models have not exploited yet the predictive ability of factorial regression to model the genomic random regression to a particular environmental stress.

In this paper we develop a genomic model to predict for the reaction norms to a specific environmental factor. Such predictions can then be used to select for specific adaptation to environments where this factor varies. We illustrate the usefulness of the model in wheat with the response of grain number to water deficit stress. We explore the influence of the variance explained by the regression to the environmental covariate; with the example of nitrogen stress and using simulated datasets. Finally, we discuss the use of this model for adaptation breeding.

Materials and Methods

Experimental data

The phenotypic trials have been described in (Cormier et al., 2013). Briefly, one hundred fifty nine elite European wheat elite varieties were evaluated for 2 years (2008 and 2009), in 2 French locations each year, managed by INRA (Estrés Mons (EM) in 2008 and 2009) and by Arvalis-Institut du végétal (Villiers-le-Bâcle (VB) in 2008 and Vraux (VR) in 2009). In each year-location combination, at least two nitrogen treatments were applied: low nitrogen (LN) and high nitrogen (HN). In 2008 in Villiers-le-Bâcle, there was an additional nitrogen treatment that consisted of a high N treatment with an additional application of nitrogen at flowering (HN+). In 2009 in Vraux, a third treatment consisted of no nitrogen application (LN-). In total, 10 combinations of year, location and N treatment were evaluated in this experiment designed to explore different levels of N stress in the experiment. Grain number per square meter was measured for the 159 varieties. The block designs were different between the INRA and the Arvalis experiments and had either all genotypes in common between blocks or only the check lines. For each trial, we used the corrections of (Cormier et al., 2013), in

which the block effect was estimated based on the common genotypes using a linear model with block and genotype as fixed effects. Climatic data was also available on each trial. We analyzed two stresses, nitrogen stress and water deficit stress. The influence of nitrogen stress (Mascianica and Walden, Frederick and Marshall, 1985; Jeuffroy and Bouchard, 1999) and also the effect of water deficit (Barlow, 1986; Armstrong et al., 1996; Gupta et al., 2001) have been shown to strongly impact wheat grain number and thus grain yield. Water deficit was measured as the sum of the negative terms of the difference between rainfall (P) and potential evapotranspiration (PET) in April and called Sum P-PET<0 (Doorenbos and Pruitt, 1977; Allen et al., 1998). For nitrogen stress, we considered the average nitrogen nutrition index (NNI) (Justes et al., 1994; Lemaire and Meynard, 1997) at flowering for the variety Soissons as a plant-based indicator of the environmental stress.

The 159 varieties have been genotyped with the 90K Illumina chip (Wang et al., 2014) and private Single Nucleotide Polymorphism (SNP) markers developed by Biogemma (Cormier et al., 2013). Markers with a minor allele frequency below 0.025 were removed, which resulted in 24 835 remaining polymorphic markers. Missing values represented 6% of the data and were imputed by the observed frequency of the corresponding SNP (Rutkoski et al., 2013).

Following the method of Patterson et al. (2006), Cormier et al. (2013) did not find any strong structure in the same panel. Although the elite lines of the panel were obtained by different breeders, there seems to be enough gene flow between breeding programs to prevent strong structure to appear.

Statistical models for genomic predictions

We describe here genomic predictions models: an additive gBLUP model (Model 1) (Hayes et al., 2009; Habier et al., 2013), and a genomic random regression to an environmental stress covariate model, that we called a Factorial Regression genomic BLUP model or FR-gBLUP (Model 2).

Additive genomic BLUP model: The additive gBLUP model consisted of modeling the genetic additive effects of the genotypes as a random effect with a variance covariance structured by a genomic estimate of kinship between the individuals.

$$y = \mu \mathbf{1}_n + \beta \mathbf{Z}_E \mathbf{x}_E + \mathbf{Z}_G \mathbf{u}_G + \mathbf{Z}_E \mathbf{u}_E + \mathbf{Z}_{GE} \mathbf{u}_{GE} + \boldsymbol{\varepsilon} \quad \left\{ \begin{array}{l} \mathbf{u}_G \sim N(0, \sigma_G^2 \mathbf{K}) \\ \mathbf{u}_E \sim N(0, \sigma_E^2 \mathbf{I}_{nE}) \\ \mathbf{u}_{GE} \sim N(0, \sigma_{GE}^2 \mathbf{I}_{nGE}) \\ \boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I}_n) \end{array} \right. \quad (\text{Model 1})$$

In which \mathbf{y} is the adjusted mean for grain number at the plot level corrected by the block effect, μ is the overall mean, \mathbf{x}_E is the vector of the environmental covariate, with one unique value of the covariate per environment, and which is centered to 0 and scaled to a variance of 1, β is the coefficient of the fixed regression to \mathbf{x}_E , \mathbf{u}_G is the genetic effects following a normal distribution $\mathbf{u}_G \sim N(0, \sigma_G^2 \mathbf{K})$ with \mathbf{K} the genomic relationship matrix calculated using VanRaden's (2008) formula based on the centered genotypic data \mathbf{M} and a allelic frequencies constant $k = 2 \sum_m p_m (1 - p_m)$, so that $\mathbf{K} = \mathbf{M} \mathbf{M}' / k$, \mathbf{u}_E is the random effect of the environments, \mathbf{u}_{GE} is the random effect of the genotype-by-environment interactions following a normal distribution, $\mathbf{Z}_G, \mathbf{Z}_E, \mathbf{Z}_{GE}$ are the incidence matrices for the genotypic, environment and GxE interaction effects (respectively), \mathbf{I} represents the identity matrix, nE, the number of environments, nGE the number of Genotype-Environment combinations, n the number of observations and ϵ represents the residuals. The absence of structure of the standardized residuals suggested no need for more complex model, such as a heteroscedastic model with a residual variance by environment.

According to the equation of prediction of random effects in (Foulley, 2003), the equation of the predictions of the genetic additive effects can be written

$$\hat{\mathbf{u}}_G = \sigma_G^2 \mathbf{K} \mathbf{Z}_G' \mathbf{V}^{-1} (\mathbf{y} - \hat{\mu} \mathbf{1}_n - \hat{\beta} \mathbf{Z}_E \mathbf{x}_E) \quad (\text{Equation A.1})$$

$$\text{with } \mathbf{V} = \sigma_G^2 \mathbf{Z}_G \mathbf{K} \mathbf{Z}_G' + \sigma_E^2 \mathbf{Z}_E \mathbf{Z}_E' + \sigma_{GE}^2 \mathbf{Z}_{GE} \mathbf{Z}_{GE}' + \sigma^2 \mathbf{I}$$

To predict the performance $\hat{\mathbf{u}}_{G_new}$ of a new genotype, its marker information \mathbf{M}_{G_new} was used to calculate its relatedness to the other genotypes. Thus the prediction of additive effect for a new genotype was

$$\hat{\mathbf{u}}_{G_new} = \sigma_G^2 \left(\frac{\mathbf{M}_{G_new} \mathbf{M}'}{k} \right) \mathbf{Z}_G' \mathbf{V}^{-1} (\mathbf{y} - \hat{\mu} \mathbf{1}_n - \hat{\beta} \mathbf{Z}_E \mathbf{x}_E) \quad (\text{Equation A.2})$$

BLUP predictions of a genomic random regression to an environmental stress - Mixed-model Factorial Regression genomic BLUP (FR-gBLUP): We modeled here the genomic random regression (\mathbf{b}_G) to the environmental covariate \mathbf{x}_E , so that the reaction norms to \mathbf{x}_E were more similar for related than unrelated genotypes.

$$\mathbf{y} = \mu \mathbf{1}_n + \beta \mathbf{Z}_E \mathbf{x}_E + \mathbf{Z}_G \mathbf{u}_G + (\mathbf{Z}_G \mathbf{b}_G) \circ (\mathbf{Z}_E \mathbf{x}_E) + \mathbf{Z}_E \mathbf{u}_E + \mathbf{Z}_{GE} \mathbf{u}_{GE} + \epsilon \quad (\text{Model 2})$$

$$\text{with } \begin{pmatrix} \mathbf{u}_G \\ \mathbf{b}_G \\ \mathbf{u}_E \\ \mathbf{u}_{GE'} \\ \boldsymbol{\varepsilon} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma_G^2 \mathbf{K} & \sigma_{gb} \mathbf{K} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \sigma_{gb} \mathbf{K} & \sigma_B^2 \mathbf{K} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_E^2 \mathbf{I}_{nE} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_{GE'}^2 \mathbf{I}_{nGE} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma^2 \mathbf{I} \end{pmatrix} \right)$$

In which the notations are the same as in Model 2, and \mathbf{b}_G is the coefficient of the random regression to the environmental covariate \mathbf{x}_E for each genotype and represents the specific genotypic linear reaction norm to the environmental covariate \mathbf{x}_E following the normal distribution $\mathbf{b}_G \sim N(\mathbf{0}, \sigma_B^2 \mathbf{K})$, the symbol \circ denotes the Hadamard product, and σ_{gb} the covariance between the random effects of the intercepts (\mathbf{u}_G) and the slopes (\mathbf{b}_G).

The phenotypic trait \mathbf{y} is thus modeled as a linear function of the environmental covariate \mathbf{x}_E , with \mathbf{u}_G the intercept, \mathbf{b}_G a random deviation to the average slope β depending on genotypes. The different slopes \mathbf{b}_G of the genotypes are modeled considering their genomic relationship. An example of Model 2 is described in Appendix 1.

We calculated the BLUP predictions of the genomic additive effects and regression to \mathbf{x}_E as:

$$\begin{pmatrix} \hat{\mathbf{u}}_G \\ \hat{\mathbf{b}}_G \end{pmatrix} = \begin{pmatrix} \sigma_G^2 \mathbf{K} & \sigma_{gb} \mathbf{K} \\ \sigma_{gb} \mathbf{K} & \sigma_B^2 \mathbf{K} \end{pmatrix} \begin{pmatrix} \mathbf{Z}'_G \\ \mathbf{Z}'_G \end{pmatrix} \mathbf{V}^{-1} (\mathbf{y} - \hat{\mu} \mathbf{1}_n - \hat{\beta} \mathbf{Z}_E \mathbf{x}_E) \quad (\text{Equation B.1})$$

$$\text{with } \mathbf{V} = \sigma_G^2 \mathbf{Z}_G \mathbf{K} \mathbf{Z}'_G + \sigma_{gb} (\mathbf{D} \mathbf{Z}_G \mathbf{K} \mathbf{Z}'_G + \mathbf{Z}_G \mathbf{K} \mathbf{Z}'_G \mathbf{D}') + \sigma_B^2 \mathbf{D} \mathbf{Z}_G \mathbf{K} \mathbf{Z}'_G \mathbf{D}' + \sigma_E^2 \mathbf{Z}_E \mathbf{Z}'_E + \sigma_{GE'}^2 \mathbf{Z}_{GE'} \mathbf{Z}'_{GE'} + \sigma^2 \mathbf{I}$$

where \mathbf{D} is a diagonal matrix of dimension the number of observations, and which contains on the diagonal the elements of $\mathbf{Z}_E \mathbf{u}_E$.

Similarly, we wrote the predictions for genomic additive and interactive effects in model (4) for a new genotype as:

$$\begin{pmatrix} \hat{\mathbf{u}}_{G_new} \\ \hat{\mathbf{b}}_{G_new} \end{pmatrix} = \begin{pmatrix} \sigma_G^2 \left(\mathbf{M}_{G_new} \mathbf{M}' / k \right) & \sigma_{gb} \left(\mathbf{M}_{G_new} \mathbf{M}' / k \right) \\ \sigma_{gb} \left(\mathbf{M}_{G_new} \mathbf{M}' / k \right) & \sigma_B^2 \left(\mathbf{M}_{G_new} \mathbf{M}' / k \right) \end{pmatrix} \begin{pmatrix} \mathbf{Z}'_G \\ \mathbf{Z}'_G \end{pmatrix} \mathbf{V}^{-1} (\mathbf{y} - \hat{\mu} \mathbf{1}_n - \hat{\beta} \mathbf{Z}_E \mathbf{x}_E)$$

$$(\text{Equation B.2})$$

The variances were estimated using REML. We computed these models using ASReml in R (Gilmour et al., 2006).

Importance of the random regression to an environmental covariate effect

We evaluated the part of GxE interactions relatively to the genetic additive effects as the $\frac{\sigma_{GE}^2}{(\sigma_G^2 + \sigma_{GE}^2)}$ ratio, in a model similar to Model 1, but assuming independence and identical distribution of the genetic additive effects instead of the genomic kinship structure of variance covariance.

Likewise, we measured the part of GxE interactions explained by the regression \mathbf{b}_G to the environmental covariate \mathbf{x}_E as the $\frac{\sigma_B^2}{(\sigma_B^2 + \sigma_{GE^*}^2)}$ ratio, in a model similar to Model 2, but without genomic relationships and without covariance between additive values and random regression coefficients.

Assessment of the prediction quality of the models

The prediction quality of the different models was assessed by cross validations. We used 4 different cross validation schemes: CVrandom, CVnewG, and CVnewE, and CVnewGE. All cross validations consisted of a 4-fold cross validation (Kohavi, 1995), using 75% of the data in the training set and 25% in the validation set. The variance components were estimated for each training set. In CVrandom, the folds were built randomly from the matrix of observations of all individuals in different environments. CVnewG consisted of building the folds so that there were no common genotypes between the training set and the validation set. Thus, CVnewG simulated situations where a model is built using some genotypes to predict the performance of new lines in the same set of environments. Similarly, CVnewE had no common environments *i.e.* location x year climatic conditions for water deficit, and location x year climatic condition x nitrogen treatment for nitrogen stress, between training and validation sets. Thus CVnewE simulated situations to predict the performances of tested genotypes in new environments. Finally, CVnewGE split the data according to the genotypes and the environment, so that predictions were made for new genotypes in new environmental conditions. The cross-validations were repeated 10 times, except for water stress with CVnewE which was repeated only once since the experiment only explored 4 location-year climatic conditions.

To evaluate the interest of modeling an interaction term, we calculated, within location-year-N treatment environments, the Pearson correlation $\text{cor}(\mathbf{y}, \hat{\mathbf{y}})$ where $\hat{\mathbf{y}} = \hat{\mathbf{u}}_G$ without interaction (Model 1), and $\hat{\mathbf{y}} = \hat{\mathbf{u}}_G + \hat{\mathbf{b}}_G \mathbf{x}_E$ with interaction (Model 2).

A Wilcoxon rank sum test was performed to assess the significance of the difference of accuracies per environment between the additive gBLUP and the FR-gBLUP.

Simulation of Experimental data

To assess the effect of variance σ_B^2 on the prediction gain in accuracy, we simulated data according to the FR-gBLUP model, so that they followed the normal distribution $N(\mu \mathbf{1}_n + \beta \mathbf{Z}_E \mathbf{x}_E, \mathbf{V})$ with \mathbf{V} defined above (Equation B.1), with $\mu, \beta, \sigma_G^2, \sigma^2$ the parameters estimated by the FR-gBLUP (Model 2) for NNI, the environmental covariate which displayed the smallest variability in genomic random regression, with $\sigma_{gb} = 0$ for the sake of simplicity, and with σ_B^2 defined to be a multiple of σ_G^2 (x0.15, x0.30, x0.5). For each value of σ_B^2 , we simulated 20 times virtual phenotypes from the given distribution. For each sampling, we run the cross validation analysis to measure the prediction accuracy gains in the different cross-validation cases. 20 simulation runs for each ratio were sufficient for stable prediction accuracies.

Results

Different patterns of reaction norms of grain number to nitrogen or to water deficit stress

The measure of heritability calculated using the Cullis et al. (2006) formula showed that grain number has high heritability ($h^2=0.94$). The $\frac{\sigma_{GE}^2}{(\sigma_G^2 + \sigma_{GE}^2)}$ ratio was estimated at 27%. Within this part of the GxE interaction for grain number, the $\frac{\sigma_B^2}{(\sigma_B^2 + \sigma_{GE'}^2)}$ ratio showed that the genomic random regression term to nitrogen stress represented only 4.2% of the variance of GxE interactions while it represented 66% in the case of water deficit stress.

The covariance term between slope and intercept has been estimated to be close to 0 for the response to the Sum of P-PET<0, indicating independence between additive values and genomic regression to water deficit stress (Table 2). For NNI, the covariance was estimated as positive, and the correlation was 0.99, indicating that for related genotypes, high deviations from the average intercept had high slope deviations from the average slope. Thus the high positive covariance suggests that varieties with high grain number in low NNI conditions better use nitrogen when they

CHAPITRE 1

are in high NNI conditions, compared to varieties with low grain number in stressed conditions. However, the variance explained by the regressions to NNI was so small that the genotypic reaction norms to NNI are almost parallel and do not display much cross over interactions. Also, the Spearman correlations of the genotypic values between the 10 environments (Table 1) confirms that the level of GxE interaction within a location x year, *i.e.* the G x N stress, is globally much lower than the GxE interaction between locations x years.

	VB08_HN	VB08_LN	VB08_HN+	VR09_HN	VR09_LN	VR09_LN-	EM08_HN	EM08_LN	EM09_HN	EM09_LN
VB08_HN	1	0.76	0.81	0.61	0.64	0.54	0.67	0.65	0.65	0.56
VB08_LN	-	1	0.73	0.49	0.51	0.44	0.56	0.58	0.57	0.5
VB08_HN+	-	-	1	0.5	0.56	0.5	0.69	0.68	0.59	0.5
VR09_HN	-	-	-	1	0.93	0.78	0.32	0.23	0.75	0.68
VR09_LN	-	-	-	-	1	0.81	0.39	0.32	0.76	0.66
VR09_LN-	-	-	-	-	-	1	0.37	0.34	0.61	0.54
EM08_HN	-	-	-	-	-	-	1	0.66	0.45	0.35
EM08_LN	-	-	-	-	-	-	-	1	0.36	0.37
EM09_HN	-	-	-	-	-	-	-	-	1	0.66
EM09_LN	-	-	-	-	-	-	-	-	-	1

Table 1 Pairwise Spearman correlation of the genotypic values between environments

	$\hat{\sigma}_G^2$	$\hat{\sigma}_{gb}$	$\hat{\sigma}_B^2$	$\hat{\sigma}_E^2$	$\hat{\sigma}_{GE'}^2$	$\hat{\sigma}^2$	$\hat{\sigma}_B^2 / \hat{\sigma}_G^2$
Grain							
Number and	1.74+/-	0.03+/-	0.44+/-	9.18+/-	0.43+/-	1.34+/-	
Sum of P-	0.22	0.09	0.07	4.59	0.06	0.05	0.25
PET<0							
Grain							
Number and	1.73+/-	0.37+/-	0.08+/-	1.94+/-	0.87+/-	1.33+/-	
NNI	0.22	0.06	0.02	0.97	0.07	0.05	0.05

Table 2 Variance components of Grain Number with Model 2 for nitrogen and water deficit stresses (With the environmental covariates NNI for nitrogen stress and Sum of P-PET<0 for water deficit stress)

Modeling the genomic random regression to water deficit stress improve prediction accuracies

With both additive and FR-gBLUP model, CVrandom and CVnewE cross-validations showed the highest prediction accuracies while CVnewG and CVnewGE cross-validations showed lower prediction accuracies (Table 3). For water deficit stress, prediction accuracies per environment with the interaction FR-gBLUP model were generally higher than those obtained with the additive model (Fig. 1). The highest gains in accuracies of the FR-gBLUP relatively to the additive model were obtained when predicting for new environments, *i.e* under CVnewE (+15.4%) and CVnewGE (14.0%). The FR-gBLUP also benefits the average prediction accuracies under CVrandom (+12.3%) with significant gains in 9 environments out of 10. Under CVnewG, the FR-gBLUP model for the reaction norms to Sum P<PET in April increased the average accuracy by 3.5%.

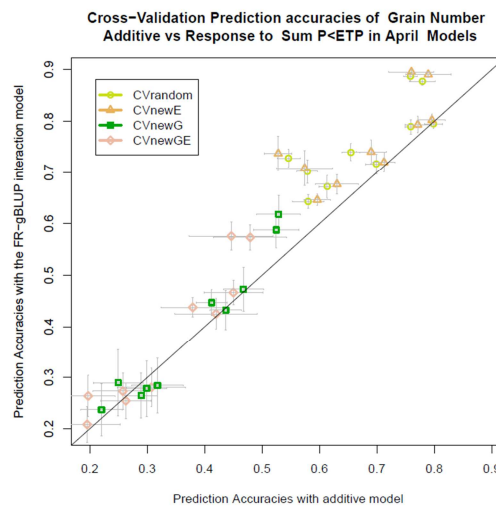


Figure 1 Prediction accuracy of FR-gBLUP vs Additive model for Sum of P-PET<0. The light grey crosses represent the prediction accuracy standard deviations, measured on the 10 repetitions of the cross-validations.

Influence of the variance of the genomic random regression to the environmental covariate

The genomic regression to NNI at flowering did not really affect prediction accuracies compared to an additive model (Table 3). This interaction between genotypes and nitrogen stress displayed indeed very low variance (Table 2). Simulations showed that the σ_B^2 / σ_G^2 ratio strongly affected the prediction accuracy gains from modeling the variability of the genomic regression to the stress covariate, in all cross validations (Fig. 2). The gains were three times higher when σ_B^2 / σ_G^2 ratio was increased from 0.15 to 0.50. In comparison with the observed data, for the response to Sum of P-PET<0 which displays a σ_B^2 / σ_G^2 ratio of 0.25 (Table 2), the observed gains (from +3.5% to 15.4%) globally sit within the range of the simulated gains for ratios between 0.15 and 0.30.

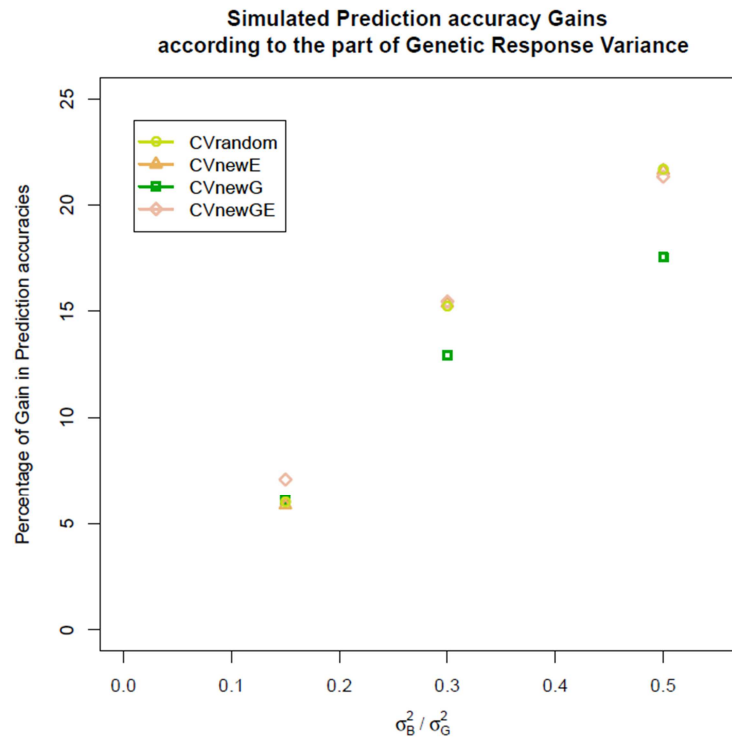


Figure 2 Prediction accuracy gains ([Accuracy in FR-gBLUP model - Accuracy in additive model] x 100 / Accuracy in additive model) simulated with different ratios of σ_B^2 / σ_G^2 in all cross validations.

CHAPITRE 1

Table 3: Table of Prediction Accuracies per environment with the additive gBLUP and the FR-gBLUP models for the reaction norms to water and nitrogen stresses Signif. codes: 0.001 '***' 0.01 '**' 0.05 '*'

Sum P<ETP in April	Cross- Validation	VB08_LN	VB08_HN	VB08_HN+	VR09_HN	VR09_LN-	VR09_LN	EM08_HN	EM08_LN	EM09_LN	EM 09_HN
Additive model	CVrandom	0.61+/-0.01	0.76+/-0.01	0.65+/-0.02	0.76+/-0.01	0.58+/-0.03	0.78+/-0.02	0.58+/-0.02	0.55+/-0.02	0.7+/-0.02	0.8+/-0.02
	CVnewE	0.57	0.72	0.61	0.64	0.49	0.67	0.55	0.51	0.70	0.8
	CVnewG	0.29+/-0.03	0.32+/-0.04	0.3+/-0.04	0.53+/-0.04	0.41+/-0.03	0.52+/-0.04	0.22+/-0.04	0.25+/-0.04	0.44+/-0.03	0.47+/-0.03
	CVnewGE	0.26+/-0.03	0.30+/-0.04	0.27+/-0.03	0.33+/-0.05	0.29+/-0.03	0.36+/-0.05	0.22+/-0.04	0.21+/-0.05	0.41+/-0.04	0.44+/-0.04
FR-gBLUP model	CVrandom	0.67+/-0.02***	0.79+/-0.01***	0.74+/-0.02***	0.89+/-0.01***	0.64+/-0.01***	0.88+/-0.01***	0.7+/-0.02***	0.73+/-0.02***	0.72+/-0.02*	0.79+/-0.01
	CVnewE	0.63	0.75	0.70	0.80	0.57	0.79	0.70	0.72	0.70	0.79
	CVnewG	0.26+/-0.04	0.28+/-0.05	0.28+/-0.05	0.62+/-0.04***	0.45+/-0.02**	0.59+/-0.04***	0.24+/-0.05	0.29+/-0.06*	0.43+/-0.04	0.47+/-0.04
	CVnewGE	0.24+/-0.05	0.26+/-0.05	0.29+/-0.06	0.47+/-0.04***	0.37+/-0.03***	0.47+/-0.04***	0.23+/-0.04	0.3+/-0.04***	0.4+/-0.05	0.45+/-0.04

NNI	Cross- Validation	VB08_LN	VB08_HN	VB08_HN+	VR09_HN	VR09_LN-	VR09_LN	EM08_HN	EM08_LN	EM09_LN	EM 09_HN
Additive model	CVrandom	0.61+/-0.02	0.76+/-0.01	0.67+/-0.01	0.76+/-0.01	0.58+/-0.01	0.79+/-0.02	0.56+/-0.02	0.53+/-0.03	0.71+/-0.02	0.79+/-0.01
	CVnewE	0.62+/-0.03	0.78+/-0.01	0.66+/-0.03	0.76+/-0.05	0.59+/-0.04	0.78+/-0.04	0.58+/-0.06	0.55+/-0.05	0.71+/-0.02	0.8+/-0.02
	CVnewG	0.3+/-0.04	0.33+/-0.04	0.31+/-0.04	0.53+/-0.04	0.41+/-0.03	0.53+/-0.04	0.25+/-0.05	0.24+/-0.05	0.44+/-0.04	0.48+/-0.04
	CVnewGE	0.25+/-0.04	0.31+/-0.03	0.27+/-0.04	0.47+/-0.02	0.39+/-0.03	0.51+/-0.02	0.21+/-0.04	0.19+/-0.04	0.43+/-0.02	0.46+/-0.02
FR-gBLUP model	CVrandom	0.62+/-0.02	0.75+/-0.02	0.66+/-0.02	0.74+/-0.02	0.59+/-0.02	0.78+/-0.02	0.56+/-0.02	0.55+/-0.02	0.69+/-0.02	0.79+/-0.01
	CVnewE	0.62+/-0.04	0.76+/-0.03	0.66+/-0.04	0.74+/-0.04	0.57+/-0.03	0.8+/-0.02	0.59+/-0.06	0.53+/-0.05	0.7+/-0.03	0.8+/-0.03
	CVnewG	0.28+/-0.03	0.31+/-0.05	0.28+/-0.03	0.52+/-0.04	0.4+/-0.03	0.52+/-0.04	0.23+/-0.04	0.22+/-0.05	0.43+/-0.04	0.46+/-0.05
	CVnewGE	0.28+/-0.04	0.31+/-0.04	0.29+/-0.05	0.39+/-0.03	0.38+/-0.03	0.48+/-0.04	0.23+/-0.06*	0.24+/-0.07**	0.42+/-0.04	0.44+/-0.04

Discussion

Improving Genomic Predictions by modeling the response to an environmental stress

Since grain number presents high heritability (0.94), its prediction accuracy was expected to be high with the additive model (Hayes et al., 2009; Jannink, 2010). Based on the simulation results (Fig. 2), modeling the genomic interaction with an environmental stress covariate is only interesting when the reaction norms to the stress covariate show large enough variability. Even though the experiment aimed at exploring highly diverse nitrogen stress conditions, the very low variance of the grain number reaction norms to the NNI of Soissons, reveals a low genetic variability of nitrogen response in this panel composed of varieties developed in the last 20 years of wheat breeding. The high correlation between genetic additive values and reaction norms to NNI shows that varieties that valorize N the most are also the ones with high performance. Because this population comes from elite material, it results from breeding programs which targeted for varieties valorizing nitrogen the best. Indeed, the reaction norms to NNI colored according to the year of release (Appendix 2) supports that breeding has historically favored varieties with the highest potential, and that indirectly, the effect on tolerance to nitrogen stress (or valorization of the available N) has not generated crossover interactions: the best varieties in high N conditions are also performing the best in stressed conditions. Thus, for NNI, the standard additive gBLUP model was sufficient to achieve high prediction accuracies, and the FR-gBLUP did not show substantial gains, as expected considering the low proportion of variance explained by this covariate. We also looked into a quadratic and a piecewise linear response to NNI but none of them explained more GxE interaction than the linear response (results not shown). In contrast, the response to the Sum of P-PET<0 in April showed a larger variance, as high as 25% of the genetic additive variance, indicating that there is an interaction signal that FR-gBLUP model was able to capture and predict. The predictions for the same cross validations displayed indeed increased gains of accuracy (from +3.5% to 15.4%) compared to the additive model.

For genomic models, one of the most difficult challenges is to predict the performance of new lines, which is indeed a major objective for breeders. This ability depends on the strength of the relationship shared between genotypes in the training population and testing population (Habier et al., 2007, 2010a; Rincet et al., 2012; Ly et al., 2013). In our study, we found a consistent decrease of accuracy from CVrandom to CVnewG and CVnewGE, and also observed larger variations of prediction accuracies, despite the absence of population structure in this panel (Cormier et al., 2013), confirming the difficulty to predict new genotypes depending on the relationship between reference individuals and predicted individuals. However, in average, for the response to water deficit stress, the FR-gBLUP model displayed gains of prediction accuracies for CVnewG (3.5%) and CVnewGE

(14.0%). Thus, when predicting the performance of untested individuals, information from relatives on the reaction norms to an environmental stress covariate still contributes to accuracy.

The interest of the FR-gBLUP model is that it also addresses the challenge of predicting for new environments. Under CVnewE and CVnewGE, that require predictions for environments which have not been evaluated in the training set, FR-gBLUP exhibited high gains in accuracy compared to the additive gBLUP model (+15.4% in CVnewE and +14.0% in CVnewGE for water deficit stress). Consequently, while modeling a genomic kinship for the responses to the environmental stress covariate allows predicting untested genotypes, the factorial regression model performs well in conferring the FR-gBLUP model the ability to predict for untested environments. In addition, similar trends with important prediction accuracy gains were obtained with the radiation stress in May, affecting meiosis with variability in genotypic sensibilities. These results also support the predictive ability of FR-gBLUP (data not shown).

Limits of the FR-g BLUP

The model we proposed for the genotypic reaction norm to an environmental stress has limitations. First, regarding the issue of selecting environmental variables, only a limited number of environmental stress variables can be modeled by factorial regression (Denis, 1980; Denis et al., 1997). Nevertheless, more environmental covariates could be added, as long as they are not too correlated to each other and that the degrees of freedom allowed by the dataset are not exceeded. Otherwise raw covariates can be summarized by principal components or other combinations. Secondly, factorial regression hypothesizes that the response to environmental covariates is linear, but this linearity of the response would have to be verified (Des Marais et al., 2013).

A step toward exploiting genomic GxE interactions in genomic selection

In our study, the FR-gBLUP model's purpose was to dissect the response to a single stress. We therefore did not aim at matching the prediction accuracy gains of models whose purpose is to predict overall GxE and which may include many environmental covariates. Rather we aimed at evaluating how much the FR-gBLUP can bring to the prediction and understanding of GxE interactions. Nevertheless, it is instructive to compare FR-gBLUP to models seeking an overall prediction of GxE. To compare our study to Jarquín et al. (2014) and Heslot et al. (2014), we consider accuracy gains relatively to the part of genetic variance explained by GxE. In our study, the FR-gBLUP model was evaluated for grain number where GxE interactions accounted for 27% of the total genetic variance (additive and interactive). Jarquín et al. (2014) analyzed yield, for which GxE interactions accounted for 30% of the total genetic variance. They showed a 34% gain in prediction accuracies with a cross-validation similar to our CVrandom, and a 17% gain with a cross-validation

similar to CVnewG. Our study did not achieve such gains. Because we focused on modeling the response to a single environmental factor, which represented 66% (Sum of P-PET<0) of the GxE, we aimed at predicting $66\% \times 27\% = 17.8\%$ only of the total genetic variance. By considering an interaction which variability represented about half of the GxE interaction in Jarquín et al.(2013) 's study, the FR-gBLUP model displayed lower prediction accuracy gains for CVrandom and CVnewG (12.3 and 3.5% for Sum of P-PET<0).

Heslot et al. (2014) also modeled genomic GxE interactions to predict yield, for which GxE variance was 63% of the total genetic variance. With a cross-validation similar to CVnewGE, they measured a 11.1% gain in accuracies. In our dataset we obtained with CVnewGE, a prediction accuracy gain of 14.0% for the genomic regression to water deficit stress. The FR-gBLUP model thus proved its ability to improve prediction accuracy of untested genotypes in untested environments. Some uncertainty still remains in the comparison of these 3 different studies because (i) the part of cross-over in the GxE interactions and its impact on the varieties ranking influences the extent to which a GxE model can improve an additive model, also because (ii) the strength of the additive relationships in the population determines how well we can borrow information from genotypes to predict new ones, (iii) besides, the genotyping method influences how well the genomic relationship matrix is estimated, and finally (iv) the calculation of the prediction accuracy gain is relative to the additive model prediction accuracy and thus when the additive model already performs well, as in our examples,, it makes the gains look small. It also has to be pointed out that according to the experimental designs of the different studies' trials and the precision of measurement of the trait considered, the precision of the phenotypic data was likely to be different. On top of that, the methods to evaluate prediction accuracies, with cross-validations, used different numbers of folds. Despite all these caveats, the FR-gBLUP model seems to be as good as other GxE models to predict GxE interactions.

Applications for breeders and advisers

In the situation of predicting for known genotypes in untested environments (CVnewE), the FR-gBLUP model for the genomic regressions to Sum of P-PET<0 was better than the additive model (+15.4% accuracy). Indeed, factorial regression is an appropriate model for prediction of environments which have not been tested but have been characterized (Malosetti et al., 2013), as long as the new environmental conditions sit within the range of the tested conditions. This may allow breeders to use fewer environments during their selection process, enabling them to screen more candidates, thereby increasing selection intensity and genetic progress rate. Moreover, predicting for a characterized environment also implies predicting future environmental trends or future environmental objectives that we want to target, such as future water deficit stresses, or reduced

applications of nitrogen. The factorial regression model can also help advisers to guide farmers for the choice of the most appropriate varieties, given the risk for a particular environmental stress.

Furthermore, as an extension of the genomic BLUP, the FR-gBLUP allows predictions for new individuals based on their genotypic data. Our study aims at expanding the possibilities of breeding for stress adaptation. Nowadays the need to accelerate breeding for yield potential and better adaptation to the changing climate is a critical issue. The main approach developed in this way is to give more insight into the physiology of the tolerance to environmental stress to define physiological traits and use genetics to explain and predict these traits. Adaptation is a complex trait to predict, determined by many genes and their interactions. The FR-gBLUP encompasses the big amount of marker data available to predict the response to the environmental stress as the trait of selection interest. Also, in consideration of the potentialities of genomic selection to accelerate selection cycles, our study brings up breeding for adaptation beyond breeding for optimizing available resource. One might aim for better use of the environmental resources when crops are in not stressed conditions, or one may rather intend to have a low decrease of the trait of interest when the environmental stress increases.

Conclusions

We have developed a genomic random regression extension of the factorial regression model to model and predict for the reaction norms to a stress environmental covariate. In this wheat case study, as long as the variance of the genomic regression to the stress covariate is sufficient, this model proved to perform better than an additive model. The FR-gBLUP model also provides insights into the understanding of GxE and broadens the possibilities for varietal recommendation to particular environmental stresses and for adaptation breeding.

List of abbreviations

gBLUP, best linear unbiased predictor; FR-gBLUP, factorial regression best linear unbiased predictor; GxE, genotype-by-environment; Sum of P-PET<0, hydric balance = sum of rainfall – potential evapotranspiration; NNI : nitrogen nutrition index; SNP, single nucleotide polymorphism; CVrandom, random cross-validation; CVnewE, cross-validation new environments; CVnewG, cross-validation new genotypes; CVnewGE, cross-validation new genotypes and new environments.

Competing interest

The authors declare they have no competing interests.

Authors' contributions

DL, SH, AG and GC designed the analysis. DL and SH developed the models and the equations. DL, SH, AG FC, RR, JLJ and GC interpreted and discussed the results. DL performed the statistical analyses and wrote the manuscript.

Acknowledgements

The authors thank the staff at Estrées-Mons experimental unit (INRA) and at Villiers-le-Bâcle and Vraux experimental farms (Arvalis) for setting the experiments and producing the data. The authors also thank Biogemma for providing the genotyping information. This work was supported by the INRA metaprogram SELGEN and the Auvergne Region, through the funding of DL's PhD thesis. We are grateful to Sue Welham and the ASReml support for their help. We also thank Jacques Legouis and David Gouache for the constructive and enriching discussions. The authors thank the reviewers for their comments, which improved the manuscript.

References

- Allen R, Pereira L, Raes D, Smith M (1998) Crop evapotranspiration: Guidelines for computing crop water requirements. FAO Irrig. Drain.
- Armstrong L, Abrecht D, Anderson W, Belford R (1996) The effect of non-lethal water deficits during establishment on the growth of wheat crops. 8th Aust. Agron. Conf.
- Barlow E (1986) Water relations of expanding leaves. *Funct plant Biol* 13:45–58. doi: 10.1071/PP9860045
- Benmahammed a., Kribaa M, Bouzerzour H, Djekoun a. (2009) Assessment of stress tolerance in barley (*Hordeum vulgare* L.) advanced breeding lines under semi arid conditions of the eastern high plateaus of Algeria. *Euphytica* 172:383–394. doi: 10.1007/s10681-009-0046-x
- Bohnert H, Nelson D, Jensen R (1995) Adaptations to Environmental Stresses. *Plant Cell* 7:1099–1111. doi: 10.1105/tpc.7.7.1099
- Brisson N, Gate P, Gouache D, et al. (2010) Why are wheat yields stagnating in Europe? A comprehensive data analysis for France. *F Crop Res* 119:201–212. doi: 10.1016/j.fcr.2010.07.012
- Burgueño J, de los Campos G, Weigel K, Crossa J (2012) Genomic Prediction of Breeding Values when Modeling Genotype × Environment Interaction using Pedigree and Dense Molecular Markers. *Crop Sci* 52:707. doi: 10.2135/cropsci2011.06.0299
- Cormier F, Faure S, Dubreuil P, et al. (2013) A multi-environmental study of recent breeding progress on nitrogen use efficiency in wheat (*Triticum aestivum* L.). *Theor Appl Genet* 126:3035–48. doi: 10.1007/s00122-013-2191-9
- Crossa J, Vargas M, Cossani CM, et al. (2013) Evaluation and Interpretation of Interactions. *Agron J*. doi: 10.2134/agronj2012.0491
- Denis JB (1980) Analyse de la régression factorielle. *Biom Praxim* 20:1–34.
- Denis JB, Piepho HP, van Eeuwijk F. (1997) Modelling expectation and variance for genotype by environment data. *Heredity (Edinb)* 79:162–171.
- Des Marais DL, Hernandez KM, Juenger TE (2013) Genotype-by-Environment Interaction and Plasticity: Exploring Genomic Responses of Plants to the Abiotic Environment. *Annu Rev Ecol Evol Syst* 44:5–29. doi: 10.1146/annurev-ecolsys-110512-135806
- Doorenbos J, Pruitt WO (1977) Crop water requirements. FAO Irrig. Drain.
- Finlay KW, Wilkinson GN (1963) The analysis of adaptation in a plant-breeding programme. *Aust J Agric Res* 14:742–754. doi: 10.1071/AR9630742
- Fischer RA, Maurer R (1978) Drought resistance in spring wheat cultivars. I. Grain yield responses. *Aust J Agric Res* 29:897–912. doi: 10.1071/AR9780897

- Foulley J (2003) Le modèle linéaire mixte. 1–139.
- Frederick JR, Marshall HG (1985) Grain Yield and Yield Components of Soft Red Winter Wheat as Affected by Management Practices1. *Agron J* 77:495. doi: 10.2134/agronj1985.00021962007700030030x
- Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2006) ASReml User Guide Release 2.0. Hemel Hempstead, HP1 1ES, UK
- Gupta N, Gupta S, Kumar A (2001) Effect of water stress on physiological attributes and their relationship with growth and yield of wheat cultivars at different stages. *J Agron Crop ...* 186:55–62.
- Habash DZ, Bernard S, Schondelmaier J, et al. (2007) The genetics of nitrogen use in hexaploid wheat: N utilisation, development and yield. *Theor Appl Genet* 114:403–19. doi: 10.1007/s00122-006-0429-5
- Habier D, Fernando R, Dekkers J (2007) The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177:2389–97. doi: 10.1534/genetics.107.081190
- Habier D, Fernando RL, Garrick DJ (2013) Genomic-BLUP Decoded: A Look into the Black Box of Genomic Prediction. *Genetics* 1–37. doi: 10.1534/genetics.113.152207
- Habier D, Tetens J, Seefried F, et al. (2010) The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genet Sel Evol* 42:5. doi: 10.1186/1297-9686-42-5
- Haldane JBS (1946) The interaction of nature ad nurture. *Ann Hum Genet* 13:197–205. doi: 10.1111/j.1469-1809.1946.tb02358.x
- Hayes BJ, Bowman PJ, Chamberlain a J, Goddard ME (2009) Invited review: Genomic selection in dairy cattle: progress and challenges. *J Dairy Sci* 92:433–43. doi: 10.3168/jds.2008-1646
- Heslot N, Akdemir D, Sorrells ME, Jannink J-L (2014) Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theor Appl Genet* 127:463–480. doi: 10.1007/s00122-013-2231-5
- Hirel B, Le Gouis J, Ney B, Gallais A (2007) The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches. *J Exp Bot* 58:2369–87. doi: 10.1093/jxb/erm097
- Jannink J-L (2010) Dynamics of long-term genomic selection. *Genet Sel Evol* 42:35. doi: 10.1186/1297-9686-42-35
- Jarquín D, Crossa J, Lacaze X, et al. (2014) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor Appl Genet* 127:595–607. doi: 10.1007/s00122-013-2243-1
- Jeuffroy M-H, Bouchard C (1999) Intensity and Duration of Nitrogen Deficiency on Wheat Grain Number. *Crop Sci* 39:1385–1393.

CHAPITRE 1

- Justes E, Mary B, Meynard J-M, et al. (1994) Determination of a critical nitrogen dilution curve for winter wheat crops. *Ann Bot* 74:397–4007.
- Kohavi R (1995) A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Int. Jt. Conf. Artif. Intell.*
- Lemaire G, Meynard J-M (1997) Use of the Nitrogen Nutrition Index for the Analysis of Agronomical Data. In: Lemaire G (ed) *Diagnosis Nitrogen Status Crop*. Springer Berlin Heidelberg, pp 44–55
- Lin CS, Binns MR (1988) A superiority measure of cultivar performance for cultivar x location data. *Can J Plant Sci* 68:193–198.
- Ly D, Hamblin M, Rabbi I, et al. (2013) Relatedness and Genotype × Environment Interaction Affect Prediction Accuracies in Genomic Selection: A Study in Cassava. *Crop Sci* 53:1312. doi: 10.2135/cropsci2012.11.0653
- Malosetti M, Ribaut J, van Eeuwijk F (2013) The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. *Front Physiol* 4:44. doi: 10.3389/fphys.2013.00044
- Malosetti M, Voltas J, Romagosa I, et al. (2004) Mixed models including environmental covariables for studying QTL by environment interaction. *Euphytica* 137:139–145.
- Mascianica MP, Walden RF Performance of winter wheat under conventional and intensive N management. *Appl Agric Res* 1:32–36.
- Mathews KL, Malosetti M, Chapman S, et al. (2008) Multi-environment QTL mixed models for drought stress adaptation in wheat. *Theor Appl Genet* 117:1077–91. doi: 10.1007/s00122-008-0846-8
- Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–29.
- Nicotra a B, Atkin OK, Bonser SP, et al. (2010) Plant phenotypic plasticity in a changing climate. *Trends Plant Sci* 15:684–92. doi: 10.1016/j.tplants.2010.09.008
- Parchman TL, Gompert Z, Mudge J, et al. (2012) Genome-wide association genetics of an adaptive trait in lodgepole pine. *Mol Ecol* 21:2991–3005. doi: 10.1111/j.1365-294X.2012.05513.x
- Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* 2:e190. doi: 10.1371/journal.pgen.0020190
- Ravagnolo O, Misztal I (2000) Genetic component of heat stress in dairy cattle, parameter estimation. *J Dairy Sci* 83:2126–30. doi: 10.3168/jds.S0022-0302(00)75095-8
- Rincent R, Laloë D, Nicolas S, et al. (2012) Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.). *Genetics* 192:715–28. doi: 10.1534/genetics.112.141473
- Rutkoski JE, Poland J, Jannink J-L, Sorrells ME (2013) Imputation of unordered markers and the impact on genomic selection accuracy. *G3 (Bethesda)* 3:427–39. doi: 10.1534/g3.112.005363

CHAPITRE 1

- Schulz-Streeck T, Ogutu JO, Gordillo A, et al. (2013) Genomic selection allowing for marker-by-environment interaction. *Plant Breed* 132:532–538. doi: 10.1111/pbr.12105
- Smith a. B, Cullis BR, Thompson R (2005) The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. *J Agric Sci* 143:449. doi: 10.1017/S0021859605005587
- Stinchcombe JR, Kirkpatrick M (2012) Genetics and evolution of function-valued traits: understanding environmentally responsive phenotypes. *Trends Ecol Evol* 27:637–47. doi: 10.1016/j.tree.2012.07.002
- VanRaden PM (2008) Efficient methods to compute genomic predictions. *J Dairy Sci* 91:4414–23. doi: 10.3168/jds.2007-0980
- Via S, Gomulkiewicz R, De Jong G, et al. (1995) Adaptive phenotypic plasticity: consensus and controversy. *Trends Ecol Evol* 10:212–7.
- Violle C, Navas M-L, Vile D, et al. (2007) Let the concept of trait be functional! *Oikos* 116:882–892. doi: 10.1111/j.2007.0030-1299.15559.x
- Wang S, Wong D, Forrest K, et al. (2014) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnol J* 12:787–796. doi: 10.1111/pbi.12183
- Weinig C, Ewers BE, Welch SM (2014) Ecological genomics and process modeling of local adaptation to climate. *Curr Opin Plant Biol* 18:66–72. doi: 10.1016/j.pbi.2014.02.007

Appendix 1: Example of the FR-gBLUP (Model 2)

$$\mathbf{y} = \mu \mathbf{1}_n + \beta \mathbf{Z}_E \mathbf{x}_E + \mathbf{Z}_G \mathbf{u}_G + (\mathbf{Z}_G \mathbf{b}_G) \circ (\mathbf{Z}_E \mathbf{x}_E) + \mathbf{Z}_E \mathbf{u}_E + \mathbf{Z}_{GE'} \mathbf{u}_{GE'} + \boldsymbol{\varepsilon} \quad (\text{Model 2})$$

$$\text{with } \begin{pmatrix} \mathbf{u}_G \\ \mathbf{b}_G \\ \mathbf{u}_E \\ \mathbf{u}_{GE'} \\ \boldsymbol{\varepsilon} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma_G^2 \mathbf{K} & \sigma_{gb} \mathbf{K} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \sigma_{gb} \mathbf{K} & \sigma_B^2 \mathbf{K} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_E^2 \mathbf{I}_{nE} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_{GE'}^2 \mathbf{I}_{nGE} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma^2 \mathbf{I} \end{pmatrix} \right)$$

Let's consider 3 genotypes G1, G2 and G3, evaluated in 2 environments E1 and E2. Their phenotypic observations are noted $y_{Gi,Ej}$. The environments E1 and E2 are characterized by an environmental stress covariate x_{Ej} .

$$\text{Let's } \mathbf{y} = \begin{pmatrix} y_{G1,E1} \\ y_{G2,E1} \\ y_{G3,E1} \\ y_{G1,E2} \\ y_{G2,E2} \\ y_{G3,E2} \end{pmatrix} \text{ be the vector of the phenotypic observations. In this case, } \mathbf{Z}_E \mathbf{x}_E \text{ is } \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_{E1} \\ x_{E2} \end{pmatrix},$$

$$\mathbf{Z}_G \mathbf{u}_G \text{ is equal to } \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_{G1} \\ u_{G2} \\ u_{G3} \end{pmatrix}, \text{ and similarly, the genotypic response term can be written as}$$

$$\mathbf{Z}_G \mathbf{b}_G = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} b_{G1} \\ b_{G2} \\ b_{G3} \end{pmatrix}$$

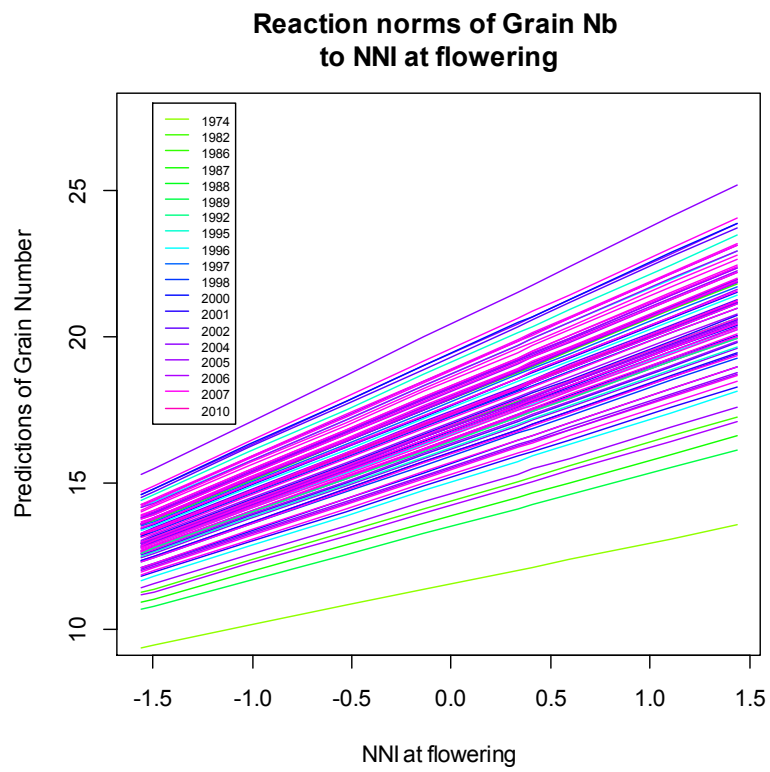
With these notations, Model 2 becomes:

$$\begin{pmatrix} y_{G1,E1} \\ y_{G2,E1} \\ y_{G3,E1} \\ y_{G1,E2} \\ y_{G2,E2} \\ y_{G3,E2} \end{pmatrix} = \mu \mathbf{1}_n + \beta \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_{E1} \\ x_{E2} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_{G1} \\ u_{G2} \\ u_{G3} \end{pmatrix} + \left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} b_{G1} \\ b_{G2} \\ b_{G3} \end{pmatrix} \right) \circ \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_{E1} \\ x_{E2} \end{pmatrix} + \mathbf{Z}_E \mathbf{u}_E + \mathbf{Z}_{GE'} \mathbf{u}_{GE'} + \boldsymbol{\varepsilon}$$

$$\Leftrightarrow \begin{pmatrix} y_{G1,E1} \\ y_{G2,E1} \\ y_{G3,E1} \\ y_{G1,E2} \\ y_{G2,E2} \\ y_{G3,E2} \end{pmatrix} = \mu \mathbf{1}_n + \beta \begin{pmatrix} x_{E1} \\ x_{E1} \\ x_{E1} \\ x_{E2} \\ x_{E2} \\ x_{E2} \end{pmatrix} + \begin{pmatrix} u_{G1} \\ u_{G2} \\ u_{G3} \\ u_{G1} \\ u_{G2} \\ u_{G3} \end{pmatrix} + \begin{pmatrix} b_{G1} \\ b_{G2} \\ b_{G3} \\ b_{G1} \\ b_{G2} \\ b_{G3} \end{pmatrix} \circ \begin{pmatrix} x_{E1} \\ x_{E1} \\ x_{E1} \\ x_{E2} \\ x_{E2} \\ x_{E2} \end{pmatrix} + \mathbf{Z}_E \mathbf{u}_E + \mathbf{Z}_{GE'} \mathbf{u}_{GE'} + \boldsymbol{\varepsilon}$$

$$\Leftrightarrow \begin{pmatrix} y_{G1,E1} \\ y_{G2,E1} \\ y_{G3,E1} \\ y_{G1,E2} \\ y_{G2,E2} \\ y_{G3,E2} \end{pmatrix} = \mu \mathbf{1}_n + \beta \begin{pmatrix} x_{E1} \\ x_{E1} \\ x_{E1} \\ x_{E2} \\ x_{E2} \\ x_{E2} \end{pmatrix} + \begin{pmatrix} u_{G1} \\ u_{G2} \\ u_{G3} \\ u_{G1} \\ u_{G2} \\ u_{G3} \end{pmatrix} + \begin{pmatrix} b_{G1}x_{E1} \\ b_{G2}x_{E1} \\ b_{G3}x_{E1} \\ b_{G1}x_{E2} \\ b_{G2}x_{E2} \\ b_{G3}x_{E2} \end{pmatrix} + \mathbf{Z}_E \mathbf{u}_E + \mathbf{Z}_{GE'} \mathbf{u}_{GE'} + \boldsymbol{\varepsilon}$$

Appendix 2: Reaction norms of Grain Number to NNI at flowering



1.3- Equivalence entre le modèle FR-gBLUP et le modèle de Jarquin et al. (2014)

De même qu'il existe une équivalence entre le RR-BLUP et le gBLUP (Ruppert et al., 2003; Lorenz et al., 2011), nous nous sommes penchés sur la question d'une équivalence entre le FR-gBLUP et le modèle de Jarquín et al., (2014).

En simplifiant les modèles pour ne garder que les effets génétiques et d'interaction :

$$\mathbf{y} = \mu \mathbf{1}_n + \mathbf{Z}_w \mathbf{w}_{GE} + \mathbf{Z}_G \mathbf{u}_G + \mathbf{Z}_{GE} \mathbf{g} \mathbf{w}_{GE} + \mathbf{e} \quad (\text{Modèle 1})$$

$$\text{avec} \begin{cases} \mathbf{w}_{GE} \sim N(0, \sigma_E^2) \\ \mathbf{u}_G \sim N(0, \sigma_G^2) \\ \mathbf{g} \mathbf{w}_{GE} \sim N(0, [\mathbf{Z}_G \mathbf{G} \mathbf{Z}_G'] \circ \Omega \sigma_{GE}^2) \\ \mathbf{e} \sim N(0, \sigma_e^2) \end{cases}$$

où $\Omega_{nE, nE} = \mathbf{W}_{nE, q} \mathbf{W}_{nE, q}'$, q étant le nombre de covariables environnementales.

Dans le modèle de Jarquín et al., (2014), ils définissent donc la covariance entre y_{ij} et $y_{i'j'}$ comme :

$$\text{cov}(y_{ij}, y_{i'j'}) = k_{ii'} \sigma_G^2 + \Omega_{ij, i'j'} \sigma_w^2 + k_{ii'} \Omega_{ij, i'j'} \sigma_{Gw}^2$$

$$\begin{aligned} \text{où} \quad \Omega_{ij, i'j'} &= \sum_l w_{l,ij} w_{l,i'j'} \\ \Omega_{ij, i'j'} &= w_{1,ij} w_{1,i'j'} + w_{2,ij} w_{2,i'j'} + \dots + w_{q,ij} w_{q,i'j'} \end{aligned}$$

Avec le FR-gBLUP, en considérant plusieurs covariables environnementales, le modèle s'écrit :

$$\mathbf{y} = \mu \mathbf{1}_n + \beta_1 \mathbf{w}_{1E} + \dots + \beta_c \mathbf{w}_{qE} + \mathbf{Z}_G \mathbf{u}_G + (\mathbf{Z}_G \mathbf{b}_{1G}) \circ \mathbf{w}_{1E} + \dots + (\mathbf{Z}_G \mathbf{b}_{qG}) \circ \mathbf{w}_{qE} + \mathbf{Z}_E \mathbf{u}_E + \mathbf{Z}_{GE'} \mathbf{u}_{GE'} + \mathbf{e}$$

$$\begin{pmatrix} \mathbf{u}_G \\ \mathbf{b}_{1G} \\ \dots \\ \mathbf{b}_{qG} \\ \mathbf{u}_E \\ \mathbf{u}_{GE'} \\ \boldsymbol{\varepsilon} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_G^2 \mathbf{K} & \sigma_{gb_1} \mathbf{K} & \dots & \sigma_{gb_q} \mathbf{K} & 0 & 0 & 0 \\ 0 & \sigma_{gb_1} \mathbf{K} & \sigma_{B_1}^2 \mathbf{K} & \dots & \sigma_{b_1 b_q} \mathbf{K} & 0 & 0 \\ 0 & \dots & \dots & \dots & \dots & 0 & 0 \\ 0 & \sigma_{gb_q} \mathbf{K} & \sigma_{b_1 b_q} \mathbf{K} & \dots & \sigma_{B_q}^2 \mathbf{K} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_E^2 \mathbf{I}_{nE} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{GE'}^2 \mathbf{I}_{nGE} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \right)$$

(Modèle 2 FR-gBLUP à plusieurs covariables)

La covariance entre y_{ij} et $y_{i'j'}$ est alors :

$$\begin{aligned} \text{cov}(gge_{ij}, gge_{i'j'}) &= \text{cov}(u_{Gi}, u_{Gi'}) + \text{cov}(b_{G1i} w_{1ij}, b_{1Gi'} w_{1i'j'}) + \dots + \text{cov}(b_{Gqi} w_{qij}, b_{Gqi'} w_{qi'j'}) \\ &\quad + \left(\text{cov}(u_{Gi}, b_{1Gi'} w_{1i'j'}) + \text{cov}(u_{Gi'}, b_{1Gi} w_{1ij}) \right) + \dots + \left(\text{cov}(u_{Gi}, b_{Gqi'} w_{qi'j'}) + \text{cov}(u_{Gi'}, b_{Gqi} w_{qij}) \right) \\ &\quad + 2 \text{cov}(b_{G1i} w_{1ij}, b_{G2i'} w_{2i'j'}) + \dots + 2 \text{cov}(b_{G1i} w_{1ij}, b_{Gqi'} w_{qi'j'}) + \dots + 2 \text{cov}(b_{G(q-1)i} w_{(q-1)ij}, b_{Gqi'} w_{qi'j'}) \\ \text{cov}(gge_{ij}, gge_{i'j'}) &= \text{cov}(u_{Gi}, u_{Gi'}) + \text{cov}(b_{G1i} w_{1ij}, b_{1Gi'} w_{1i'j'}) + \dots + \text{cov}(b_{Gqi} w_{qij}, b_{Gqi'} w_{qi'j'}) \\ &\quad + (w_{1i'j'} + w_{1ij}) \text{cov}(u_{Gi}, b_{1Gi'}) + \dots + (w_{qi'j'} + w_{qij}) \left(\text{cov}(u_{Gi}, b_{Gqi'}) + \text{cov}(u_{Gi'}, b_{Gqi}) \right) \\ &\quad + 2 \text{cov}(b_{G1i} w_{1ij}, b_{G2i'} w_{2i'j'}) + \dots + 2 \text{cov}(b_{G1i} w_{1ij}, b_{Gqi'} w_{qi'j'}) + \dots + 2 \text{cov}(b_{G(q-1)i} w_{(q-1)ij}, b_{Gqi'} w_{qi'j'}) \end{aligned}$$

En considérant que :

- 1- les covariances entre l'intercept (u_G) et les pentes (b_{1G}, \dots, b_{qG}) et entre les pentes de réponse entre elles sont nulles, soit :

$$2 w_{1j'} w_{1j} k_{ii'} \sigma_{gb_1} + \dots + 2 w_{qj'} w_{qj} k_{ii'} \sigma_{gb_q} + 2 w_{1j} w_{2j'} k_{ii'} \sigma_{b_1 b_2} + \dots + 2 w_{1j} w_{qj'} k_{ii'} \sigma_{b_1 b_q} + \dots + 2 w_{(q-1)j} w_{qj'} k_{ii'} \sigma_{b_{q-1} b_q} = 0$$

- 2- et que les variances associées aux pentes (b_{1G}, \dots, b_{qG}) sont égales :

$$\sigma_{B_1}^2 = \dots = \sigma_{B_q}^2$$

La covariance $\text{cov}(gge_{ij}, gge_{i'j'})$ du modèle FR-gBLUP devient égale à la covariance $\text{cov}(y_{ij}, y_{i'j'})$ du modèle de Jarquín et al., (2014):

$$\begin{aligned} \text{cov}(gge_{ij}, gge_{i'j'}) &= k_{ii'} \sigma_G^2 + k_{ii'} w_{1ij} w_{1i'j'} \sigma_B^2 + \dots + k_{ii'} w_{qij} w_{qi'j'} \sigma_B^2 \\ \text{cov}(gge_{ij}, gge_{i'j'}) &= k_{ii'} \sigma_G^2 + k_{ii'} \sigma_B^2 \left(w_{1ij} w_{1i'j'} + \dots + w_{qij} w_{qi'j'} \right) \\ \text{cov}(gge_{ij}, gge_{i'j'}) &= k_{ii'} \sigma_G^2 + k_{ii'} \sigma_B^2 \Omega_{ij, i'j'} \end{aligned}$$

Ainsi, il existe donc une équivalence entre le modèle de Jarquín et al., (2014) et le FR-gBLUP à plusieurs covariables, à condition que les covariances entre la valeur génétique additive et les normes de réaction soient nulles, que les covariances entre les normes de réaction entre elles soient nulles également, et que les variances associées aux normes de réaction à chaque covariable environnementale soient égales.

Par rapport à la réalité biologique, il est très peu probable que les valeurs génétiques additives soient indépendantes des normes de réactions, que les normes de réaction soient indépendantes entre elles, et qu'elles suivent la même loi de distribution. Ces hypothèses du modèle Jarquín et al. (2014) seraient à vérifier. L'extension du FR-gBLUP à plusieurs covariables permettrait d'avoir plus de flexibilité par rapport à ces hypothèses.

Cependant, cette extension du FR-gBLUP (Modèle 2) est plus couteûse en paramètres. Elle nécessite d'estimer $q(q+1)/2$ covariances et q variances de plus que le modèle de Jarquín et al., (2014) (q étant le nombre de covariables environnementales considérées).

Dans un objectif de prédiction des normes de réaction à différents stress, et pour contrôler l'effet GxE non dû aux régressions linéaires aléatoires à ces stress, il peut s'avérer intéressant de mettre en place un modèle qui combine les avantages des 2 modèles : d'une part, en explicitant les réponses aux covariables de stress préalablement identifiées, et d'autre part, en contrôlant l'effet GxE restant avec une modélisation similaire à celle de Jarquín et al., (2014), basée sur le reste des covariables environnementales.

Dans le modèle (3) ci-dessous, on considère ainsi toutes les q covariables environnementales disponibles et mesurées. Un sous ensemble de q' covariables parmi les q a été préalablement identifié comme jouant un rôle sur l'interaction GxE et est explicité dans le modèle.

$$\mathbf{y} = \mu \mathbf{1}_n + \beta_1 \mathbf{w}_{1_E} + \dots + \beta_q \mathbf{w}_{q'_E} + \mathbf{Z}_G \mathbf{u}_G + (\mathbf{Z}_G \mathbf{b}_{1_G}) \circ \mathbf{w}_{1_E} + \dots + (\mathbf{Z}_G \mathbf{b}_{q'_G}) \circ \mathbf{w}_{q'_E} + \mathbf{Z}_E \mathbf{u}_E + \mathbf{Z}_{GE'} \mathbf{u}_{GE'} + \mathbf{e}$$

$$\begin{pmatrix} \mathbf{u}_G \\ \mathbf{b}_{1_G} \\ \dots \\ \mathbf{b}_{q'_G} \\ \mathbf{u}_E \\ \mathbf{u}_{GE'} \\ \boldsymbol{\varepsilon} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_G^2 \mathbf{K} & \sigma_{gb_1} \mathbf{K} & \dots & \sigma_{gb_{q'}} \mathbf{K} & 0 & 0 & 0 \\ \sigma_{gb_1} \mathbf{K} & \sigma_{b_1}^2 \mathbf{K} & \dots & \sigma_{b_1 b_{q'}} \mathbf{K} & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & 0 & 0 & 0 \\ \sigma_{gb_{q'}} \mathbf{K} & \sigma_{b_1 b_{q'}} \mathbf{K} & \dots & \sigma_{b_{q'}}^2 \mathbf{K} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_E^2 \mathbf{I}_{nE} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{GE'}^2 [\mathbf{Z}_G \mathbf{K} \mathbf{Z}_G'] \circ \boldsymbol{\Omega} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2 \mathbf{I} \end{pmatrix} \right)$$

(Modèle 3)

Avec les mêmes notations que précédemment, et où la matrice de ressemblance environnementale est calculée à partir des $(q-q')$ covariables restantes : $\boldsymbol{\Omega}_{nE, nE} = \mathbf{W}_{nE, (q-q')} \mathbf{W}_{nE, (q-q')}'$.

Le modèle (7) associerait ainsi les avantages du FR-gBLUP à plusieurs covariables (6) et du modèle de (Jarquín et al., 2014), à savoir la possibilité de variances différentes pour les réponses aux q' stress identifiés, d'une covariance entre les réponses à ces stress, et un nombre de paramètres à estimer réduit. Ce modèle permettrait alors de prédire les performances de nouveaux individus génotypés, dans de nouveaux environnements, caractérisés pour les q covariables environnementales.

Conclusion du chapitre 1

Ce chapitre montre les possibilités de la sélection génomique pour la prédiction des valeurs génétiques additives et des normes de réaction à un stress environnemental défini. La modélisation génomique des normes de réaction ressemble à celle des valeurs génétiques additives : elle considère une matrice d'apparementement génomique entre les individus, de sorte que des individus apparentés aient des normes de réaction plus proches que des individus non apparentés. Dans notre étude, la modélisation génomique des normes de réaction au stress hydrique a permis d'avoir de meilleures précisions de prédiction, dans les différents schémas de validation croisée testés. Toutefois, notre approche requiert d'avoir identifié préalablement un stress majeur explicatif des interactions GxE. Les niveaux de gains en précision de prédictions par rapport au modèle additif obtenus avec le FR-gBLUP en modélisant une réponse au stress hydrique sont comparables à ceux obtenus par (Jarquín et al., 2014b; Heslot et al., 2014) en modélisant l'interaction GxE. Nous avons obtenu des gains de précision de 3.5% à 15.4%, toutes validations croisées confondues. Les différences de populations, d'environnements, d'architectures génétiques des caractères, de choix de prédiction de interactions complètes ou d'une part lié à un stress environnemental particulier, et également, par conséquent, les précisions de prédiction du modèle additif utilisés comme référence pour mesurer les gains, nous incitent à nuancer les comparaisons entre les études. A titre indicatif seulement, dans les autres études de prédiction génomique des interactions GxE, les gains de précision sont de l'ordre de 17% (Burgueno), 11.1% (Heslot et al., 2014), et de 17 à 34% (Jarquín et al., 2014). Comme le montrent les résultats de simulations, la part de variance des régressions génétiques à un stress environnemental joue un rôle majeur dans la capacité à prédire les normes de réactions au stress considéré, et peut dans notre étude expliquer les gains relativement faibles, car nos jeux de données contenaient au final une variabilité des interactions GxE assez faible.

Dans le cas où la diversité environnementale des stress responsables de l'interaction et la diversité génétique des normes de réaction seraient plus grandes, la modélisation des normes de réactions à plusieurs stress avec l'extension du modèle FR-gBLUP proposée (6) permettrait de capturer plus de signal d'interaction GxE. (Jarquín et al., 2014) modélise les normes de réactions à de nombreux stress et gère la grande dimension des covariables en émettant l'hypothèse que toutes les normes de réaction aux différents stress suivent une même loi de distribution, et qu'ils sont indépendants entre stress. Nous proposons alors une modélisation (Modèle 7) qui explicite les normes de réaction aux stress identifiés avec une forte variance des réponses génétiques, tient compte d'une covariance entre les réponses aux différents stress, et modélise les normes de réaction à toutes les autres covariables de stress de la manière proposée par (Jarquín et al., 2014).

Malgré des améliorations possibles des modèles de prédiction génomique des interactions GxE, les différentes extensions génomiques de la régression factorielle permettent la prédiction des réponses génétiques aux environnements, pour de nouveaux génotypes, à partir de l'information de génotypage seulement. Néanmoins, quel que soit le modèle génétique utilisé, la prédiction génomique des interactions GxE nécessite d'avoir une caractérisation environnementale de l'ensemble de l'expérimentation.

Les caractérisations reposant sur des critères indépendants de la plante et de son développement (climatiques et pédologiques notamment) s'avèrent peu efficaces pour expliquer les interactions GxE (Brandle and Arthur, 1992). Les mesures sur la plante s'avèrent coûteuses en temps, généralement destructives et donc peu adaptées aux essais variétaux multi-locaux. De plus, ces mesures donnent une image très ponctuelle du statut de la plante (hydrique ou azoté). Pour pallier ces problèmes, il peut être intéressant de caractériser les stress environnementaux à partir de modèles de culture. En simulant des indicateurs de stress qui prennent en compte le développement de la plante au cours de la saison, les modèles de culture pourraient être de bons outils de caractérisation environnementale, pour caractériser les stress et comprendre les interactions GxE, comme chez le sorgho et le blé, dans les conditions australiennes (Chapman, 2008; Chenu et al., 2011).

Références complémentaires

- Brandle, J.E., and L.M. Arthur. 1992. A comparative evaluation of two methods of selecting locations used for testing spring wheat cultivars. *Theor. Appl. Genet.* (83): 301–304.
- Burgueño, J., G. de los Campos, K. Weigel, and J. Crossa. 2012. Genomic Prediction of Breeding Values when Modeling Genotype \times Environment Interaction using Pedigree and Dense Molecular Markers. *Crop Sci.* 52(2): 707 Available at <https://www.crops.org/publications/cs/abstracts/52/2/707>.
- Heslot, N., D. Akdemir, M.E. Sorrells, and J.-L. Jannink. 2014. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theor. Appl. Genet.* 127(2): 463–480 Available at <http://link.springer.com/10.1007/s00122-013-2231-5>.
- Jarquín, D., J. Crossa, X. Lacaze, P. Du Cheyron, J. Daucourt, J. Lorgeou, F. Piraux, L. Guerreiro, P. Pérez, M. Calus, J. Burgueño, and G. de los Campos. 2014. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* 127(3): 595–607 Available at <http://link.springer.com/10.1007/s00122-013-2243-1>.
- Lorenz, A.J., S. Chao, F.G. Asoro, E.L. Heffner, T. Hayashi, H. Iwata, K.P. Smith, M.E. Sorrells, and J. Jannink. 2011. *Genomic Selection in Plant Breeding : Knowledge and Prospects* (DL Sparks, Ed.). Elsevier Inc.

CHAPITRE 1

Ruppert, D., M. Wand, and R. Carroll. 2003. *Semiparametric Regression*. Cambridge University Press, New York, NY.

CHAPITRE 2

Compréhension des réponses à un stress et prédictions génomiques à l'aide de caractérisation environnementale par les modèles de culture

TABLE DES MATIERES : CHAPITRE 2

2.1- Préambule du chapitre 2	94
2.2- Publication 2.....	96
Introduction.....	97
Materials and Methods	99
Core Collection experimental data.....	99
Genotypic data	99
Calibration and testing of crop-model APSIM-Wheat.....	99
Environmental Characterization.....	100
Explanatory power of characterizations of N stress	101
Crop-model characterization of intra-year GxE interactions effects	102
Genomic predictions for the different characterizations of N stress.....	106
Assessment of predictive ability of the different genomic prediction models	109
Results.....	109
Environmental characterization of N stress using a crop-model	109
Model selection criteria favor crop-model based indicators to characterize G x N stress interactions.....	115
Ability of crop-model indicators to capture the differences in GxE interactions variances per year.....	117
Genomic prediction gains by modeling the interaction with N stress using crop-model indicators.....	117
Discussion.....	121
Crop modelling provides N stress indicators well suited for modelling GxE interactions	121
How much gain in prediction accuracy to expect from GxE modelling?.....	121
Genomic predictions of the response to crop-model stress indicators for N stress per year	122
Potential of genomics and crop modeling for predicting GxE interactions	123
Conclusion	124
Supplementary data.....	124

CHAPITRE 2

References.....	126
Conclusion du Chapitre 2	131
Références complémentaires.....	132

2.1- Préambule du chapitre 2

Souvent, les généticiens sélectionneurs souhaiteraient pouvoir utiliser les données issues des essais de sélection accumulées au cours du temps pour constituer une population d'entraînement. Cependant, par rapport à la volonté de prédire les interactions GxE, notamment dans les essais historiques, l'absence de caractérisations représentatives des stress, telles que des mesures éco-physiologiques, pose problème. A partir de données climatiques et pédologiques, plus facilement disponibles, les modèles de culture permettent alors de simuler les variables nécessaires à la caractérisation des stress environnementaux.

De plus, certaines études ont montré l'apport des modèles de culture pour une meilleure compréhension des interactions GxE. Muchow et al., (1996) ont montré dans le cas de la sécheresse que des indicateurs issus de modèle de culture donnent une meilleure caractérisation des interactions GxE que des indicateurs basés uniquement sur données climatiques. Les modèles de culture permettent en effet d'estimer des variables environnementales localement, en prenant en compte les interactions entre la plante et l'environnement, ce qui permet d'avoir une interprétation physiologique des interactions (Chenu et al., 2008). L'identification de patterns de stress simulé par les modèles de culture, et l'étude de la fréquence des patterns sur une longue période, dans plusieurs lieux, peut servir pour regrouper les environnements en environnement type (ET). Des ET ont été définis pour caractériser les environnements dans les zones de culture de sorgho pendant le 20^e siècle (Chapman et al., 2000b). Dans la caractérisation du stress hydrique chez le blé en Australie, Chenu et al. (2011) montrent que considérer des ET réduit l'interaction GxE au sein des ET relativement aux effets génétiques additifs intra-ET. De plus, Löffler et al. (2005) expliquent jusqu'à 30% des interactions GxE grâce aux ET définis par les patterns de stress hydriques. Par ailleurs, la pondération des essais multi-environnementaux, sur la base de leur fréquence dans la population d'environnement cible (target population of environments : TPE) permettrait d'améliorer les gains de rendement (Podlich and Cooper, 1998). (Chapman, 2008) a montré qu'il était possible de réduire les effets néfastes des interactions GxE sur les mauvais classements de variétés, en classant les essais en fonction de leur pattern de réponse au stress hydrique.

L'utilisation des modèles de culture pour le stress hydrique, stress majeur en Australie, a été largement étudiée. Cependant, dans les conditions environnementales françaises, les interactions GxE sont dues à plusieurs stress. Un des stress principaux en France, du fait de la volonté de réduire les intrants azotés, est le stress azoté.

Notre 2^e chapitre cherche à savoir à quel point la caractérisation environnementale par les modèles de culture permet de comprendre les interactions GxE, dans les conditions environnementales

CHAPITRE 2

françaises. De plus, est-ce que la caractérisation par modèle de culture permet d'avoir une meilleure compréhension et d'affiner la caractérisation des stress azotés ? Dans ce cas, peut-elle être utile à la prédiction ? Pourrait-on envisager une approche plus globale de prédiction de réponses de génotypes encore non évalués, grâce à la génomique, pour des conditions environnementales caractérisées par un modèle de culture ?

2.2- Publication 2

How can crop modeling improve genomic predictions? Illustration for N deficiency in a wheat Core Collection

Delphine Ly*, Karine Chenu[§], Arnaud Gauffreteau[‡], Renaud Rincant*, Sylvie Huet[†], David Gouache**,
and Gilles Charmet

*Institut National de la Recherche Agronomique - Blaise Pascal University, UMR1095, Genetics, Diversity and Ecophysiology of Cereals, 63039 Clermont-Ferrand, France

[§] The University of Queensland; QAAFI, 203 Tor Street, Toowoomba, QLD 4350, Australia

[‡] Institut National de la Recherche Agronomique, UMR Agronomy, Thiverval Grignon, France

[†] Institut National de la Recherche Agronomique, MAIAGE – Applied Mathematics for Computer Sciences, Jouy-en-Josas, France

** ARVALIS – Institut du Végétal, Station Expérimentale, 91720 Boigneville, France

Introduction

From 1990 onwards, wheat yields have been stagnating in Europe despite the constant genetic progress. This is partly due to climate change counteracting the genetic improvement, causing more frequent heat and drought stresses during critical phases of the crop development (Brisson et al., 2010). Furthermore, fertilizer policies aiming at reducing nitrate pollution are also impeding the expression of yield potential. In address these challenges, plant breeding increasingly tends to account for the genetic responses to these environmental components (Ortiz et al., 2008; Cormier et al., 2013).

Both breeding and varietal recommendation require an ability to predict the differential responses of the genotypes to their environments, and more specifically to the type of environments that varieties experience in farming systems. Characterizing relevant environmental stresses in particular in regards to their severity and their timing in the plant development is essential (Chenu, 2015). While environmental characterization of early studies used climatic, geographical and/or pedo-climatic variables (Balfourier and Charmet, 1991; Pollak and Corbett, 1993; Berger and Turner, 2007), new stress indices have been considered to account for crop phenology and comprehend stress timing during the crop growth cycle (Chapman et al., 2000c; Lecomte, 2008; Chenu et al., 2011), as the substantial impact of an abiotic stress such as drought depends on the developmental stage of the crop (Denmead and Shaw, 1960).

Crop models offer the possibility to refine environmental characterization in particular reflecting impacts of various environmental factors up to the stress of interest, and the interaction between stresses. Chapman et al., (2000a) used a sorghum crop-model (Hammer and Muchow, 1994) in this way to characterize environments based on a drought stress index, which integrates climatic, soil, management, crop development and also accounts for the timing and severity of the stress. Their stress index was used to distinguish three patterns of drought stress in sorghum and show that long-term frequency of severe terminal-stress differentiated differently cultivars than mid-season or mid-terminal stress. The approach was extended in wheat to characterize stress patterns in representative growing conditions in the Australian wheat belt (Chenu et al., 2013). Overall, environmental characterization for drought has been deployed for several crops including wheat, and regions including Europe (Chapman et al., 2000a; Chenu et al., 2011; Chauhan et al., 2013; Chauhan and Rachaputi, 2014; Harrison et al., 2014). Crop-modeling has also been used in the US to define maize major target populations of environments in terms of photoperiod, maximum temperature and average radiation (Löffler et al., 2005), and in Australian wheat to look at impact of frost or heat (Zheng et al., 2012, 2015; Lobell et al., 2015). Indexes relative to water deficit and nitrogen availability between booting and heading, two main limiting factors for grain protein content, were

also computed with crop simulation model in durum wheat to characterize French agro-climatic conditions (Lacaze and Roumet, 2004).

To characterize nitrogen (N) stress, 'N satisfaction indicators' have been defined accounting for supply and demand (Lecomte, 2008), such as the ratio between absorbed N at maturity and the number of grains (Meynard et Limaux, 1987), or the ratio between N concentration in the biomass and the N concentration required for optimal growth, called Nitrogen Nutrition Index (NNI) (Lemaire et Gastal, 1997; Lemaire et Meynard, 1997). However the evaluation of those indices is destructive and requires non-standard measurements and thus extra-resources. The use of crop-modelling could thus be advantageous for other stresses than water deficit, in particular for N stress. Crop-modelling could be applied to simulate patterns of N deficiency over time, or estimate indices such as NNI without destructive and time consuming sampling.

Regarding the genetics of N responses, quantitative trait loci (QTL) for grain yield, kernel number have been identified specifically at low or high N levels (Laperche et al., 2007). QTL for traits relative to N use are also known (Hirel et al., 2007; Habash et al., 2007; Bordes et al., 2013). However in such studies, the characterization of panels of genotypes in controlled stress conditions is difficult and expensive. Genomic selection allows making predictions for these yield-related highly polygenic traits (Meuwissen et al., 2001), and now aims at improving N use efficiency (Cormier et al. submitted). Recently, Ly et al. (in press) developed a genomic factorial regression prediction model (FR-gBLUP) to focus on response to particular environmental covariates, and illustrated its use on grain number in response to N stress, using NNI at anthesis (Lemaire et Gastal, 1997; Lemaire et Meynard, 1997). The FR-gBLUP model has been tested in different cross-validation schemes to represent different breeding or varietal recommendation case scenarios and was shown to improve grain number prediction compared to the classical additive genomic model.

The objective of this study was to assess the ability of crop modelling to explain Genotype x Environment (GxE) interactions. In particular, we aimed to evaluate the explanatory and predictive ability of FR-gBLUP to capture the GxE interactions due to N stress, in a core-collection, where N indices were not measured during the experiments. Crop modelling was used to estimate indicators for N stress, first in a set of experiments where N indices were measured to validate the model, and then in multi-environment trials (MET) of the Core Collection. Different stress indices were computed and compared to investigate the GxE interactions in the MET. In these trials, N treatments were applied to induce stresses before flowering, so that grain number had a high heritability and was the main driver of yield. Genomic predictions of grain number were made with the FR-gBLUP model, considering the response to N stress, with or without the N indices.

Materials and Methods

Core Collection experimental data

194 bread wheat accessions were studied out of the National Institute for Agricultural Research (INRA) bread wheat Core Collection of 372 accessions (Balfourier et al., 2007) in order to favor diversity in geographic origin and registration date, while limiting plant susceptibility to lodging. The material was evaluated for yield related traits, during two years (2007 and 2008), in three French locations each year (Estrées-Mons (49°53'N, 3°00'E) (EM), Le Moulon (48°40'N, 2°10'E) (LM) and Joze (45°86'N, 3°30'E) (JO)). In each trial, genotypes were grown under two managements (high and moderate inputs), which differed in terms of sowing density, amount of N fertilizer, disease treatments and growth regulator (Supp. Table 1) resulting in total in 12 year x location x management combinations.

For each year x location x management system, the 194 accessions were sown in 6 incomplete blocks, according to their height in order to minimize plant competition. Six check varieties were repeated in all 6 blocks. Phenotypic data were corrected for block effects, which were estimated on the 6 check varieties because other varieties were not assigned to blocks randomly.

Among the yield-related phenotypic measurements available, we focused on grain number. Since the differences in N applications occurred mainly before anthesis, these N deficiencies probably affected grain number, depending on both intensity and duration of the N stress.

More information on the trials and genetic materials can be found in Bordes et al. (2013).

Genotypic data

Varieties were genotyped with the 420K Axiom array, containing 105 000 Insertion site based polymorphism SNPs (Paux et al., 2010), 140 000 intergenic SNPs, 150 000 genic SNPs, including 10 000 candidate gene-derived polymorphism, and roughly 30 000 additional SNPs from the literature or from the BreedWheat project (<http://www.breedwheat.fr/>). 53 615 SNP markers were detected polymorphic and with a minor allele frequency greater than 0.025. Missing data were imputed by the observed allele frequency of the corresponding SNP over all accessions (Rutkoski et al., 2013).

Calibration and testing of crop-model APSIM-Wheat

APSIM-Wheat version 7.6 (Wang et al., 2002; Holtzworth et al., 2014) was used to simulate the wheat growth of a French winter cultivar, Soissons. First, this variety was parameterized with two sets of experiments: (1) a set of INRA experiments conducted during 5 years (1992-1995-1996-1998-1999) at one location (Grignon (48.9N, 1.9E, France) with 6, 10, 6, 5 and 5 N management systems for

each, respectively (Jeuffroy and Bouchard, 1999); (2) a set of trials conducted by ARVALIS-Institut du Végétal at Villiers-le-Bâcle (48.7N, 2.1E, France) in 2012 and 2013, with 2 N managements each year. In all trials, soil information (depth, texture, organic and inorganic N content), N management and anthesis date were recorded. Leaf area index, crop biomass and crop N at different stages, from emergence to harvest, and grain dry matter at harvest were measured. The N nutrition index (NNI) (Justes et al., 1994; Lemaire and Meynard, 1997) was calculated as the ratio of the actual N content ($CropN$ expressed in $kg\ N\ .ha^{-1}$) to the critical N content ($CropN_c$ in $kg\ N\ .ha^{-1}$), which is defined as the minimum crop N content required to achieve maximum crop growth rate. NNI can be approximated as a function of the crop dry matter ($CropDM$ in $t.ha^{-1}$) (Justes et al., 1994):

$$NNI = \frac{CropN}{CropN_c} = \frac{CropN}{5.35 \times CropDM^{-0.442}} \quad (\text{Equation 1})$$

The parameterization of Soissons was based on existing APSIM parameterization of winter wheat cultivar Marombi. Phenology cultivar-specific parameters (vernalization sensitivity (4.5), photoperiod sensitivity (4.7) and thermal time to floral initiation ($400^{\circ}Cd$)) and a biomass allocation parameter (number of grains per gram of stem set at 35 grains per gram of stem) were adjusted for Soissons, using the parameters obtained from full factorial design-testing all triplets of phenology parameters within their range of possibilities- and which best fitted flowering date and leaf area index for the 36 trials (i.e. best coefficient of determination, error coefficient of variation and root mean square error). Overall, a small difference of 1 to 3 days was achieved between simulated and observed flowering dates across all trials. The model was also tested for biomass, N in biomass, grain yield and NNI (see results).

APSIM-wheat simulations for the variety Soissons were tested in an independent set of experiments, for the MET of the Core Collection.

Environmental Characterization

In the core-collection multi-environment trials, N stress was characterized using either (i) the level of input as a factor (high input (HN) vs low input (LN)), (ii) the N supply (in $g\ m^{-2}$) as a covariate, (iii) the sum of N inorganic in the soil at emergence and N supply (N supply + N soil) as a covariate, or (iv) crop-model based indicators of the plant N satisfaction as covariates. For the latter, we used the APSIM-Wheat crop model (7.6) which was calibrated and tested for French environments as described above. Each of the multi-environment trials was simulated with APSIM for the reference French cultivar Soissons, using local soil and climatic data, and the management practices of the trials. To ensure best estimation of N indices for specific stages, the simulations were adjusted according to the observed booting date. Note that each trial was characterized for a unique variety

(Soissons) and that the potential differences in stress patterns experienced by the different varieties, *e.g.* due to differences in phenology, were not considered in this study. Three crop-model based indicators for N stress were defined. The first one, the integrated NNI between meiosis (Zadoks stage: ZC38; (Zadoks et al., 1974)) to anthesis (ZC65) (NNI_{meiosis-anthesis}), measured the intensity of stress during the period of the elaboration of grain number. The integrated NNI (NNI_i) corresponds to the cumulated stress over the considered period, and is defined as the area delimited by the NNI curve below 1 (NNI below 1 being considered as the non-stress threshold). The second crop-model indicator was the NNI at anthesis (NNI_a). The third characterization consisted of clustering the curves of NNI over thermal time between tillering (estimated as 900°Cd before flowering) and anthesis. NNI values were capped between 0 (no N in the plant) and 1 (no stress). A cluster analysis (R package cluster (R Development Core Team, 2003)) was performed on daily data on all the environments from the Core Collection MET, in a similar way as described for drought by Chapman et al. (2000) and Chenu et al. (2013). This environment characterization allowed the identification of three Environment Clusters based on the crop-NNI patterns.

Explanatory power of characterizations of N stress

We compared statistical models where the interactions with N stress were modeled with the different environment characterization as described above. The frame of the models is described in Equation 2. The fixed-effect terms of each model are developed in Table 1.

$$y_{ijk} = \mu + \alpha p_{ij} + g_i + \left\{ \begin{array}{c} \text{Fixed effect of} \\ \text{the N characterization} \end{array} \right\}_{jt} + e_{jt} + \left\{ \begin{array}{c} \text{Fixed effect} \\ \text{of the interaction between} \\ g_i \text{ and N characterization} \end{array} \right\}_{ijt} + \varepsilon_{ijk}$$

(Equation 2)

Where y_{ijk} corresponds to the grain number, adjusted for block effect, of the genotype i (g_i) in the environment jt (e_{jt}) defined by the trial j (*i.e.* year x location) and the treatment t , for the replicate k . μ is the overall mean, αp_{ij} represents the regression to precocity (*i.e.* booting date) p_{ij} , and ε_{ijk} were the residuals, which were independent and identically distributed, following a normal distribution $N(0, \sigma^2)$.

N characterization	Model	Fixed effect of N characterization	Fixed effect of the interaction GxN characterization
None (additive model)	Model 1	-	-
Level of input	Model 2	$C_{c(j,t)}$	$g_i C_{c(j,t)}$
Environment Cluster			
N supply	Model 3	βx_{jt}	$\beta_i x_{jt}$
N supply + N soil			
NNIa			
NNI _{meiosis-anthesis}			

Table 1. Statistical models to test GxN interaction effects. Three statistical models were tested, all following Equation 2, but differing depending on the characterization of N stress

$C_{c(j,t)}$ represents a group of environments jt categorized by level of input or environmental cluster, $g_i C_{c(j,t)}$ corresponds to the interaction between the genotype i and the environment group $c(j,t)$, βx_{jt} is the regression over the N stress covariate x_{jt} (which corresponds to either (i) N supply, (ii) N supply + N inorganic in the soil, (iii) NNI at anthesis, or (iv) NNli between meiosis and anthesis depending on the N characterization), and $\beta_i x_{jt}$ represents the interaction between the genotype i and the N stress covariate x_{jt} as in a factorial regression model (Denis, 1980).

For the models 1 to 3, the sum of square of the residuals, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) were calculated (Akaike, 1973; Schwarz, 1978), and an analysis of variance (ANOVA) was performed to test the effect of the interaction terms.

Crop-model characterization of intra-year GxE interactions effects

Because of very diverse climatic conditions in 2007 and 2008, we aimed to explore the effect of year on the variability of GxE interactions, and focused on intra-year GxE interactions effects. The objective was to evaluate precisely the part of G x N stress among the total GxE interactions per year, and assess how much of the interaction with the N stress indicators could captured each year.

$$y = \mu \mathbf{1} + \left\{ \begin{array}{c} \text{Fixed effect of} \\ \text{the N characterization} \end{array} \right\} + \mathbf{Z}_G \mathbf{u}_G + \mathbf{Z}_E \mathbf{u}_E + \left\{ \begin{array}{c} \text{Random effect} \\ \text{of the interaction between} \\ \mathbf{u}_G \text{ and the N characterization} \\ \text{by year} \end{array} \right\} + \epsilon$$

(Equation 3)

Where y is phenotypic observation of grain number adjusted for block effect, μ the overall mean, \mathbf{u}_E the environmental (year x location x management system) effects, and \mathbf{u}_G is the [(number of genotypes x number of year) x 1] vector of genotypic effects per year, which follows a normal distribution with a different variance each year (σ_{G2007}^2 in 2007 and σ_{G2008}^2 in 2008) so that

$$\mathbf{u}_G \sim N\left(0, \begin{pmatrix} \sigma_{G2007}^2 \mathbf{I}_G & \mathbf{0} \\ \mathbf{0} & \sigma_{G2008}^2 \mathbf{I}_G \end{pmatrix}\right).$$

$\mathbf{Z}_G, \mathbf{Z}_E$ are the incidence matrices for the genotypic effects per year and the environmental effects respectively, and ε the residuals, which were independent and identically distributed.

According to this structure, Models 4, 5 and 6 are developed in Table 2.

CHAPITRE 2

N characterization	Model	Fixed effect of N characterization	GxN characterization genomic interaction effect per year	Variance Covariance structure
Environment (year x location x level of input combination)	Model 4	-	$\mathbf{Z}_{G \times E \times Y} \mathbf{u}_{G \times E \times Y}$	$\left\{ \begin{array}{l} \mathbf{u}_G \sim N \left(0, \begin{pmatrix} \sigma_{G2007}^2 \mathbf{I}_G & \mathbf{0} \\ \mathbf{0} & \sigma_{G2008}^2 \mathbf{I}_G \end{pmatrix} \right) \\ \mathbf{u}_E \sim N(0, \sigma_E^2 \mathbf{I}_E) \\ \mathbf{u}_{G \times E \times Y} \sim N \left(0, \begin{pmatrix} \sigma_{G \times E 2007}^2 \mathbf{I}_G & \mathbf{0} \\ \mathbf{0} & \sigma_{G \times E 2008}^2 \mathbf{I}_G \end{pmatrix} \right) \\ \boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I}) \end{array} \right.$
Level of input	Model 5	$\mathbf{Z}_T \boldsymbol{\tau}$	$\mathbf{Z}_{G \times T \times Y} \mathbf{u}_{G \times T \times Y}$	$\left\{ \begin{array}{l} \mathbf{u}_G \sim N \left(0, \begin{pmatrix} \sigma_{G2007}^2 \mathbf{I}_G & \mathbf{0} \\ \mathbf{0} & \sigma_{G2008}^2 \mathbf{I}_G \end{pmatrix} \right) \\ \mathbf{u}_E \sim N(0, \sigma_E^2 \mathbf{I}_E) \\ \mathbf{u}_{G \times T \times Y} \sim N \left(0, \begin{pmatrix} \sigma_{G \times T \times Y 2007}^2 \mathbf{I}_G & \mathbf{0} \\ \mathbf{0} & \sigma_{G \times T \times Y 2008}^2 \mathbf{I}_G \end{pmatrix} \right) \\ \boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I}) \end{array} \right.$
Environment Cluster				
N supply	Model 6	$\beta \mathbf{Z}_E \mathbf{x}_E$	$(\mathbf{Z}_{GY} \mathbf{b}_{GY}) \circ (\mathbf{Z}_E \mathbf{x}_E)$	$\left\{ \begin{array}{l} \mathbf{u}_G \sim N \left(0, \begin{pmatrix} \sigma_{G2007}^2 \mathbf{I}_G & \mathbf{0} \\ \mathbf{0} & \sigma_{G2008}^2 \mathbf{I}_G \end{pmatrix} \right) \\ \mathbf{u}_E \sim N(0, \sigma_E^2 \mathbf{I}_E) \\ \mathbf{b}_{GY} \sim N \left(0, \begin{pmatrix} \sigma_{B2007}^2 \mathbf{I}_G & \mathbf{0} \\ \mathbf{0} & \sigma_{B2008}^2 \mathbf{I}_G \end{pmatrix} \right) \\ \boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I}) \end{array} \right.$
N supply + N soil				
NNIa				
NNI _{meiosis-anthesis}				

Table 2. Statistical models to estimate the interaction variances per year (eq. 3). Fixed and Random effects and their variance covariance structure are to be considered in Equation 3 and depend on the characterization of N stress. Recall that in general, \mathbf{u}_E the environmental (year x location x management

CHAPITRE 2

system) effects, σ_E^2 the environmental variance, \mathbf{u}_G are the genotypic effects per year, with σ_{G2007}^2 and σ_{G2008}^2 the genotypic variance each year. ϵ represents the residuals independent, identically distributed, following the normal distribution $\epsilon \sim N(0, \sigma^2 \mathbf{I})$ centered and with a variance σ^2 .

Model 4 : $\mathbf{u}_{G \times E \times Y}$ is the GxE interactions effects per year, each year modeled with a different variance to be estimated, $\sigma_{G \times E 2007}^2$ in 2007, $\sigma_{G \times E 2008}^2$ in 2008.

Model 5: τ is a vector of the N stress indicators fixed effects (either 'Levels of input' or 'Environment Cluster'), $\mathbf{u}_{G \times T \times Y}$ is the genotype x group of environments interactions per year, with a different variance each year, $\sigma_{G \times T \times Y 2007}^2$ or $\sigma_{G \times T \times Y 2008}^2$.

Model 6: \mathbf{x}_E is the vector of the environmental covariates, with one unique value of the covariate per environment, and which was centered to 0 and scaled to a variance of 1, β is the slope coefficient (fixed effect) of the regression to the environmental covariate \mathbf{x}_E , and \mathbf{b}_{GY} is the genotypic regression effect per year to the environmental covariate \mathbf{x}_E (which contains the number of years times the number of genotypes levels), with a different regression variance per year (σ_{B2007}^2 or σ_{B2008}^2). The symbol \circ denotes the Hadamard product.

$\mathbf{Z}_G, \mathbf{Z}_E, \mathbf{Z}_T, \mathbf{Z}_{G \times E \times Y}, \mathbf{Z}_{G \times T \times Y}$ are the incidence matrices for the genotypic effects per year, the environmental effects, the group effect, the the genotype-by-year effect and genotype-by-group-by-year respectively. $\mathbf{I}_G, \mathbf{I}_E, \mathbf{I}$ are identity matrices of dimension number of genotypes, number of environments or number of observations, respectively.

Genomic predictions for the different characterizations of N stress

Three genomic models were developed to assess the predictive ability of the N stress characterizations (eq. 4, Table 3). Genomic predictions of the additive values were modeled using the genomic BLUP model (gBLUP, (Henderson, 1975; Habier et al., 2013)) (Model 7). For the qualitative N stress characterization (Level of input, Environment Clusters), the interactions between genotypes and the groups of environments is modeled as in Model 8. Genomic predictions of the additive values and the response to a given quantitative characterization of N stress were performed using the FR-gBLUP model (Ly et al., in press) (Model 9). All these models (7-9) have the following model structure and are described in Table3.

$$\mathbf{y} = \mu \mathbf{1} + \left\{ \begin{array}{c} \text{Fixed effect of} \\ \text{the N characterization} \end{array} \right\} + \mathbf{Z}_G \mathbf{u}_G + \left\{ \begin{array}{c} \text{G} \times \text{N characterization} \\ \text{Genomic interaction} \\ \text{effect} \end{array} \right\} + \mathbf{Z}_E \mathbf{u}_E + \boldsymbol{\varepsilon}$$

(Equation 4)

N characterization	Model	Fixed effect of N characterization	GxN characterization genomic interaction effect	Variance Covariance structure
Additive Model	Model 7	-	-	$\begin{pmatrix} \mathbf{u}_G \\ \mathbf{u}_E \\ \boldsymbol{\varepsilon} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma_G^2 \mathbf{K} & 0 & 0 \\ 0 & \sigma_E^2 \mathbf{I}_E & 0 \\ 0 & 0 & \sigma^2 \mathbf{I} \end{pmatrix} \right)$
Level of input Environment Cluster	Model 8	$\mathbf{Z}_T \boldsymbol{\tau}$	$\mathbf{Z}_{GT} \mathbf{u}_{GT}$	$\begin{pmatrix} \mathbf{u}_G \\ \mathbf{u}_{GT} \\ \mathbf{u}_E \\ \boldsymbol{\varepsilon} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma_G^2 \mathbf{K} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_{GT}^2 \mathbf{I}_T \otimes \mathbf{K} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_E^2 \mathbf{I}_E & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma^2 \mathbf{I} \end{pmatrix} \right)$
N supply N supply + N soil NNIa NNI _{meiosis-anthesis}	Model 9	$\beta \mathbf{Z}_E \mathbf{x}_E$	$(\mathbf{Z}_G \mathbf{b}_G) \circ (\mathbf{Z}_E \mathbf{x}_E)$	$\begin{pmatrix} \mathbf{u}_G \\ \mathbf{b}_G \\ \mathbf{u}_E \\ \boldsymbol{\varepsilon} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma_G^2 \mathbf{K} & \sigma_{gb} \mathbf{K} & \mathbf{0} & \mathbf{0} \\ \sigma_{gb} \mathbf{K} & \sigma_B^2 \mathbf{K} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_E^2 \mathbf{I}_E & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma^2 \mathbf{I} \end{pmatrix} \right)$

Table 3. Statistical models to predict the genomic response to N stress characterization (eq. 4). Fixed and Random effects and their variance covariance structure to be considered in Equation 4 depending on the characterization of N stress.

Model 7 is an additive genomic BLUP (Meuwissen et al., 2001; VanRaden, 2008; Habier et al., 2013). \mathbf{u}_G are the genomic additive effects which follow a normal distribution $\mathbf{u}_G \sim N(0, \sigma_G^2 \mathbf{K})$, where σ_G^2 is the genetic variance and where the genomic relationship matrix \mathbf{K} was calculated using VanRaden's (2008) formula based on the centered genotypic data (\mathbf{M}) and an allelic frequencies constant $k = 2 \sum_m p_m (1 - p_m)$, so that $\mathbf{K} = \mathbf{M} \mathbf{M}' / k$.

CHAPITRE 2

Model 8: \mathbf{u}_{GT} is a vector of the interactions between the genotypes and the N group ('Level of input' or 'Environment Cluster') effects. $\mathbf{Z}_T, \mathbf{Z}_{GT}$ are the incidence matrix of the group effects and the genotype by group effects, respectively. \mathbf{I}_T is the identity matrix which has the dimension of the number of groups. The symbol \otimes denoted the Kronecker product. For the genomic prediction of response to an Environment Cluster or level of input, we made the assumption that the GxE interactions variance was the same across types of environments, because of the lack of degrees of freedom due to the few repetitions in the experimental design.

Model 9: \mathbf{b}_G is a vector of the effects of the genomic response to the environmental covariate \mathbf{x}_E and σ_{gb} is the covariance between the random effects of the intercepts (u_G) and the slopes (b_G).

The variances were estimated using REML with ASReml in R (Gilmour et al., 2006).

Assessment of predictive ability of the different genomic prediction models

The prediction quality of the different models was assessed by cross-validations (Kohavi, 1995). We used 3 different 4-fold cross-validation schemes as in Ly et al. (in press), using 75% of the data in the training set and 25% in the validation set, and sampling different combinations of tested or untested genotypes in tested or untested environments: (i) CVrandom where about ¼ of the random genotype and environment (location x year x treatment) combinations were excluded from the training of the model but used for the model testing, (ii) CVnewG, where about ¼ of the genotypes were randomly excluded from the training set to be used in the validation set, and (iii) CVnewE, where 3 environments (location-year-treatment) out of the 12 were excluded from the model training but used for its testing.

The cross-validations were repeated 10 times. We calculated within each environment (location-year-treatment combination) the Pearson correlation between the observation \mathbf{y} and the predictions $\hat{\mathbf{y}}$: $\text{cor}(\mathbf{y}, \hat{\mathbf{y}})$ where $\hat{\mathbf{y}} = \hat{\mathbf{u}}_G$ for models without interaction (Model 7), or $\hat{\mathbf{y}} = \hat{\mathbf{u}}_G + \hat{\mathbf{u}}_{GT}$ for models considering interaction with a qualitative co-variable as in Model (8), or $\hat{\mathbf{y}} = \hat{\mathbf{u}}_G + \hat{\mathbf{b}}_G \mathbf{x}_E$ considering interaction with a quantitative c-variable as in Model (9).

Results

Environmental characterization of N stress using a crop-model

Yield, grain number and booting varied greatly across genotypes of the Core Collections, and were also highly impacted by year, location and N treatment (Fig. 1). As expected, low input treatment had indeed lower yield and grain number than high input. The season in 2007 had more radiation, higher temperatures and more rainfall than 2008, except in Joze (Table 4).

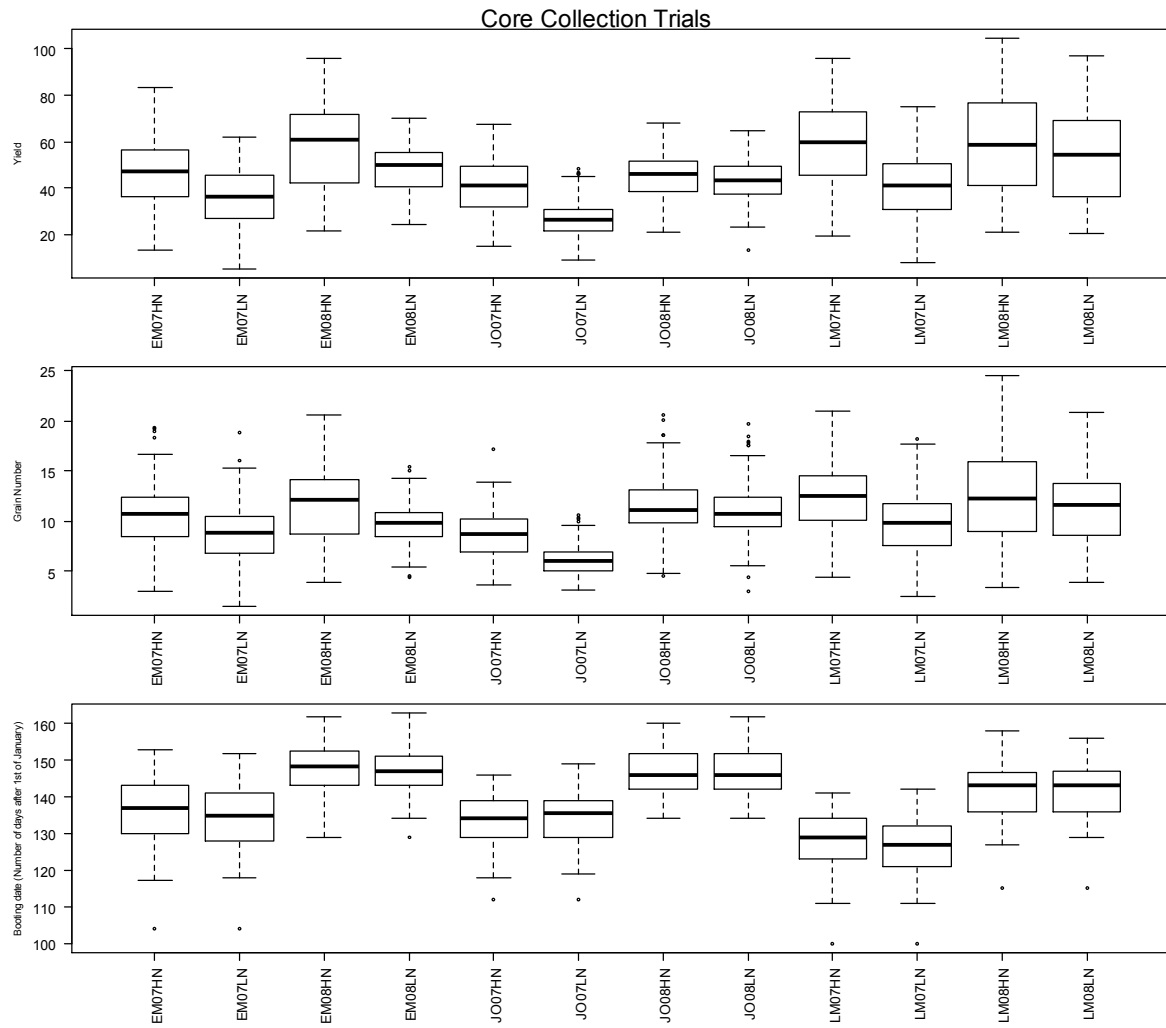


Figure 1. Distributions of yield (100 kg ha^{-1}), grain number (1000 grains per m^2) and booting date (Number of days after 1st of January) in the Core Collection trials for each of tested environments.

The first two characters represent the location: EM stands for Estrées-Mons, JO for Joze, and LM for Le Moulon. The two following character represent the year of the trial (07 for 2007 and 08 for 2008). The last two characters represent the management system (LN for low input and HN for high input).

CHAPITRE 2

Environment	Year	Location	Yield (t /ha)	Booting date (Number of days after 1st january)	Grain Number (1000 grains m ⁻²)	Thousand Kernel Weight (g)	Cumulated Rain (mm)	Average Temperature (°C)	Maximum Temperature (°C)	Minimum Temperature (°C)	Radiation (MJ m ⁻²)
EM07HN	2007	Estrées-Mons	4.68+/-1.34	136.1+/-8.5	10.6+/-3.0	44.5+/-6.1	629.7	11.5	31.6	-5.1	6.29
EM07LN		49°53'N, 3°00'E	3.65+/-1.18	134.5+/-8	8.8+/-2.5	41.2+/-7.4					
JO07HN		Joze	4.05+/-1.07	133.2+/-6.7	8.7+/-2.3	46.7+/-5.8	473.2	12.2	33.4	-14.5	4.73
JO07LN		45°86'N, 3°30'E	2.69+/-0.72	134.3+/-6.9	6.2+/-1.5	43.8+/-6.6					
LM07HN		Le Moulon	5.99+/-1.6	128.5+/-7.9	12.5+/-3.2	47.9+/-5.6	651.8	11.9	31.7	-5.0	6.51
LM07LN		48°40'N, 2°10'E	4.06+/-1.32	127.1+/-7.3	9.8+/-3.0	41.5+/-6.5					
EM08HN	2008	Estrées-Mons	5.76+/-1.77	148+/-6.2	11.6+/-3.6	49.9+/-6.2	601.4	10.5	32.9	-5.2	6.01
EM08LN		49°53'N, 3°00'E	4.89+/-0.96	147.5+/-6	9.8+/-1.9	50.2+/-5.9					
JO08HN		Joze	4.51+/-0.96	146.9+/-6.3	11.3+/-2.8	40.4+/-5.1	482.8	10.5	33.0	-10.2	4.82
JO08LN		45°86'N, 3°30'E	4.35+/-0.89	147.1+/-6.5	11+/-2.6	40.3+/-5.3					
LM08HN		Le Moulon	5.99+/-2.11	142.2+/-6.8	12.7+/-4.5	47.4+/-6.2	568.3	11.0	32.9	-5.7	5.68
LM08LN		48°40'N, 2°10'E	5.41+/-1.88	142+/-6.7	11.4+/-3.6	47.2+/-5.9					

Table 4. Phenotypic and climate characteristics of the Core Collection trials. Maximum temperatures correspond to the maximum daily temperature during the crop cycle, minimum temperatures correspond to the minimum daily temperature during the cycle.

The APSIM-Wheat crop model was used to estimate crop N indices, which were not measured in the trial. To ensure the quality of the simulations, the model was first calibrated and tested for the reference variety Soissons for a set of 37 trials (2 locations, 7 years, various N fertilization levels). Simulations slightly under-estimated crop biomass and grain yield (Fig. 2 a –b) but the model allowed accurate estimations of flowering date (data not shown, $y = -6.36 + 1.09x$; $r^2 = 0.96$), N biomass ($y = 0.87x + 0.59$; $r^2 = 0.86$; Fig 2c) and NNI ($y = 0.71x + 0.18$; $r^2 = 0.59$; Fig. 2d).

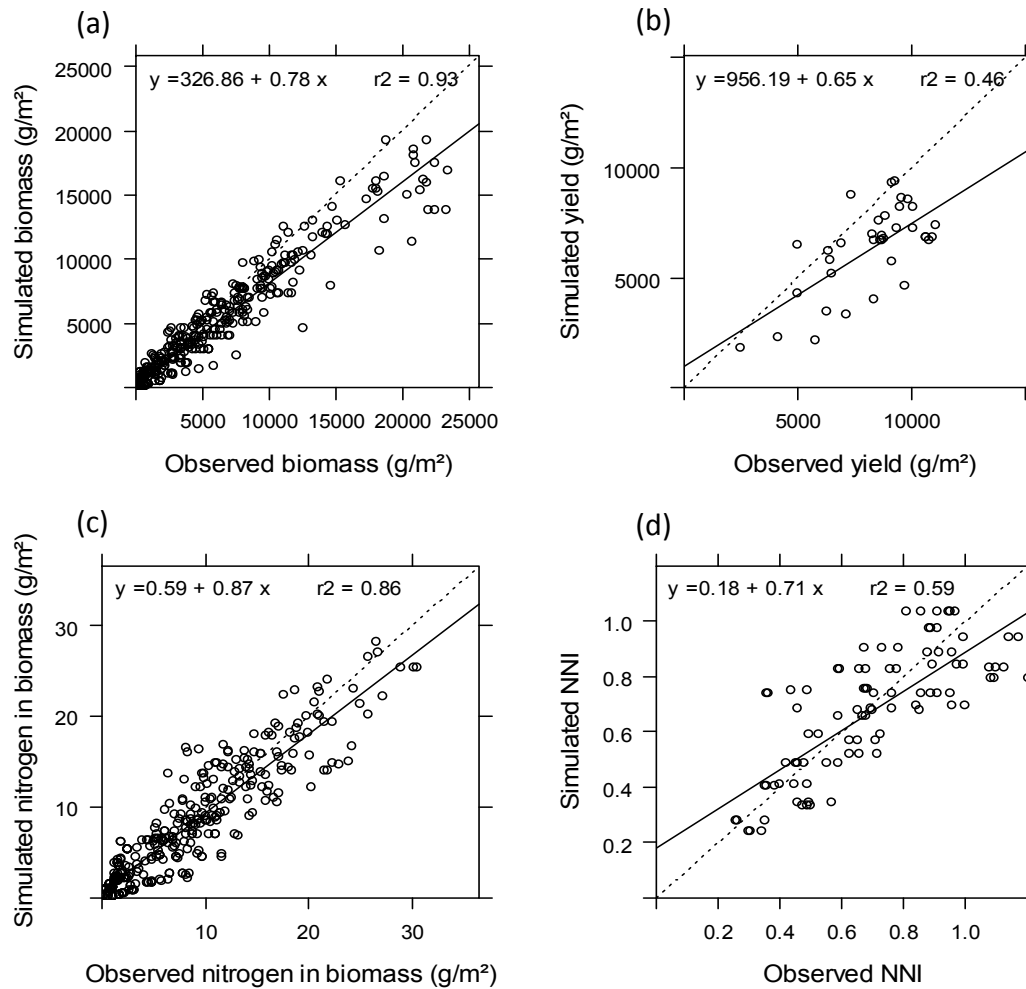


Figure 2: Comparison of APSIM simulations and observations in the calibration set for biomass (a), yield (b), nitrogen in biomass (c) and NNI (d).

Using the genotypic parameters of the Soissons that were calibrated with the 36 trials, simulations of yield for Soissons in the independent set of trials issued from the Core Collection experiment slightly over-estimated the lowest yields and under-estimated the highest values ($y = 2141.1 + 0.67x$, $r^2 = 0.8$) (Fig. 3).

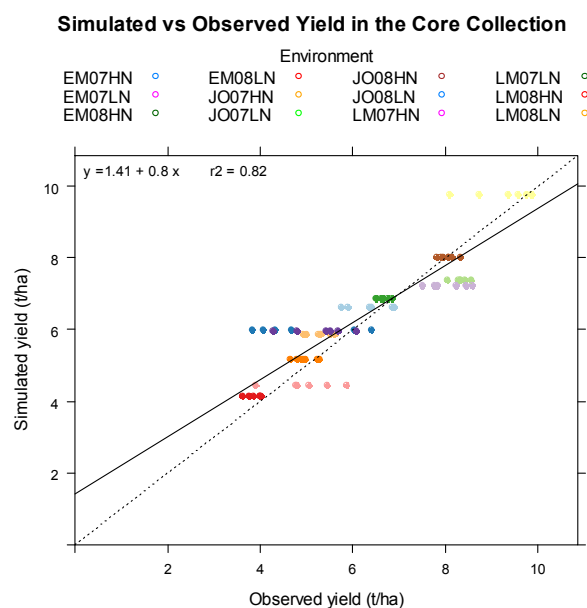


Figure 3. Comparison of APSIM simulations and observations of yield in the Core Collection trials

N indices for 12 trials of the Core Collection (Table 5) reflected the diversity in the N nutrition status that occurred in these trials. NNI at flowering varied from 0.57 (moderately severe N stress) to 0.93 (virtually no stress), with, as expected, higher values for high- than for low-input conditions within each trial (Table 1). However, some environments with low input conditions displayed lower stress than other environments with high input conditions, highlighting the effects that the year, location, soil properties can have on the crop N status.

Environment	Treatment	N _{supply} (kg ha ⁻¹)	N _{supply} + N _{soil} (kg ha ⁻¹)	Environment Cluster	NNI _a	NNI _i _{meiosis-anthesis}
EM08HN	HN	160	218	1	0.93	0.02
LM08LN	LN	40	109	1	0.83	0.08
LM08HN	HN	160	229	1	0.83	0.08
LM07HN	HN	153	148	1	0.75	0.23
EM08LN	LN	40	98	1	0.65	0.29
EM07HN	HN	153	196	2	0.78	0.18
JO07HN	HN	125	225	3	0.66	0.28
JO08HN	HN	130	200	3	0.66	0.3
EM07LN	LN	80	123	3	0.63	0.34
LM07LN	LN	80	148	3	0.61	0.37
JO07LN	LN	90	190	3	0.57	0.39
JO08LN	LN	90	160	3	0.57	0.39

Table 5. N Stress indicators from Crop-model

The timing of the N stress varied across trials and the clustering of NNI curves over thermal time allowed the identification of 3 Environment Clusters representing the pre-flowering temporal variations (Fig. 4). Environment Cluster EC1 corresponded to the highest level of NNI and was characterized by a non-stressed level of NNI (NNI = 1) up to 400 °Cd before anthesis, which declined slowly afterwards. EC1 encompassed environments with no or only a light pre-flowering N stress. Environment Cluster 2 (EC2) related to a trial (EM07LN) where N stress occurred early and was released by supplemental fertilization at jointing. Finally, Environment Cluster 3 (EC3) was characterized by early stress that increased over time up to flowering.

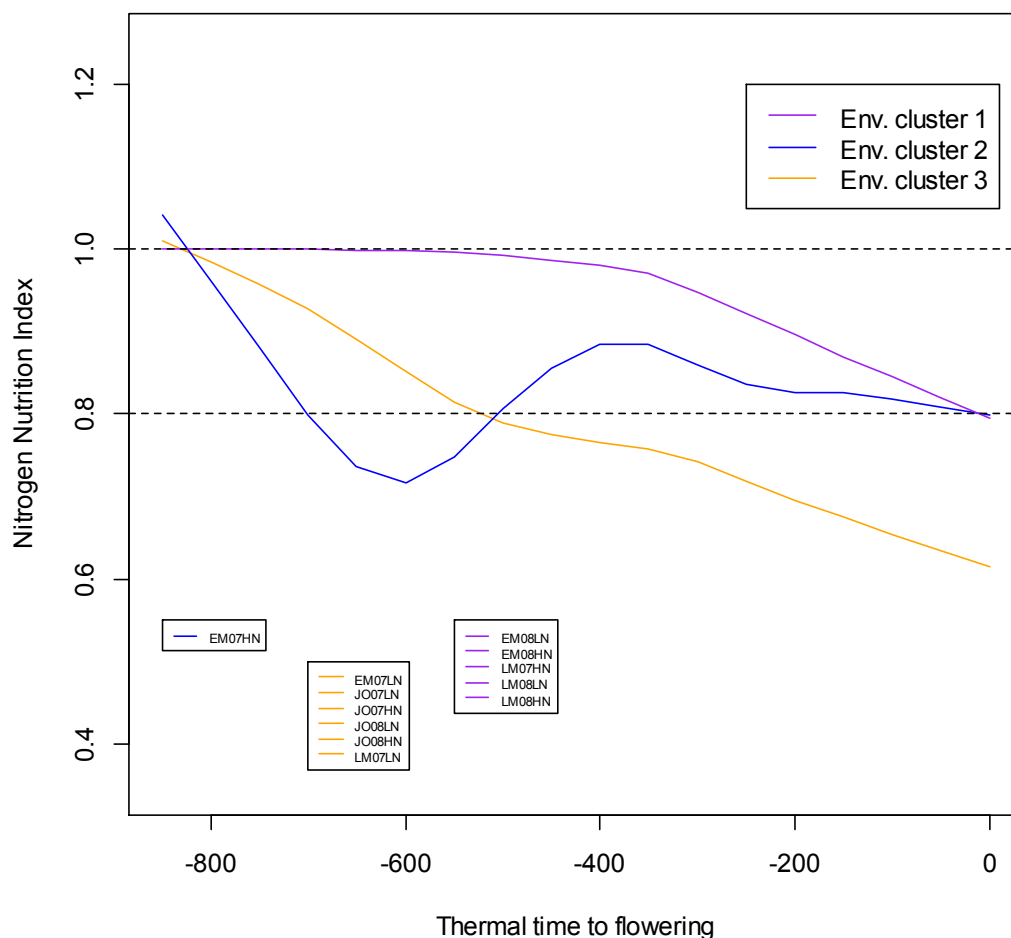


Figure 4. Environmental clusters (1-3) of the Core Collection MET according to their NNI throughout the pre-anthesis development of the crop, represented as a function of thermal time relative to flowering. A NNI of 1 indicates no nitrogen stress while a ratio of 0 indicates full stress. Data of all the environments were centered at flowering and averaged every 50°Cd and clustered in three Environment Clusters. The curves presented here correspond to the average NNI across the different environments of a cluster smoothed using the R function lowess.

Model selection criteria favor crop-model based indicators to characterize G x N stress interactions

N characterization	Model	p-value of interaction effect	Ratio Sum Squares Interaction / Sum Squares Residuals	Residuals sum of squares	AIC	BIC
None (additive model)	(1)	-	-	4607	8907	10095
Level of Input	(2)	0.9532	0.08	4272	9114	11429
Nsupply	(3)	0.949	0.08	4271	9112	11428
Nsupply + Nsoil	(3)	0.9997	0.06	4330	9144	11459
Environmental Cluster	(2)	1.01E-08	0.31	3510	9004	12449
NNIa	(3)	<2.2E-16	0.23	3578	8677	10991
NNI _{meiosis-anthesis}	(3)	<2.2E-16	0.25	3698	8758	11073

Table 6. Model selection criteria for the statistical models of grain number considering different N characterization. AIC is the Akaike Information Criterion, and BIC the Bayesian Information Criterion.

Modeling the interaction between genotypes and N stress indicators reduced the residuals compared to those of the additive model. Crop-model based indicators (NNIa, NNI_{meiosis-anthesis}, Environment Cluster) better captured G x N stress interaction than the other N indicators such as the level of input or the amount of N available in the soil (N supply, N supply + N inorganic in the soil). Indeed, only interactions with crop-model indicators were significant ($P = 0.05$). Besides, both sum of squares of the residuals and AIC (which account for the number of parameters estimated) criteria indicated that the models with crop-model based N characterization were better to capture the phenotypic variability of the MET. BIC (which penalizes the number of parameters estimated more strongly than AIC) favored the additive model, since it was the simplest model (*i.e.* less parameters), and preferred then the interaction models using the crop-model indicators NNIa and NNI_{meiosis-anthesis}. BIC strongly penalized the Environment Cluster characterization, as it required more parameters to estimate (1 for each genotype in each cluster of environments while only 1 for each genotype in the factorial regression).

CHAPITRE 2

N characterization	Model	$\hat{\sigma}_{G2007}^2$	Standard deviation of $\hat{\sigma}_{G2007}^2$	$\hat{\sigma}_{G2008}^2$	Standard deviation of $\hat{\sigma}_{G2008}^2$	$\hat{\sigma}_{\text{interaction} \times Y 2007}^2$	Standard deviation of $\hat{\sigma}_{\text{interaction} \times Y 2007}^2$	$\hat{\sigma}_{\text{interaction} \times Y 2008}^2$	Standard deviation of $\hat{\sigma}_{\text{interaction} \times Y 2008}^2$	$\hat{\sigma}_{\text{interaction} \times Y 2007}^2 / \hat{\sigma}_{G2007}^2$	$\hat{\sigma}_{\text{interaction} \times Y 2008}^2 / \hat{\sigma}_{G2008}^2$
Environment (location x level of input)	(4)	3.58	0.39	4.29	0.48	0.03	0.08	0.93	0.13	0.01	0.22
Nsupply	(6)	3.50	0.39	4.36	0.48	1.98E-07	6.19E-09	1.98E-07	6.19E-09	5.66E-08	4.54E-08
Nsupply + Nsoil	(6)	3.50	0.39	4.36	0.48	1.98E-07	6.19E-09	1.98E-07	6.19E-09	5.66E-08	4.54E-08
Level of input	(5)	3.50	0.39	4.36	0.48	1.98E-07	6.19E-09	1.98E-07	6.19E-09	5.66E-08	4.54E-08
Environmental Cluster	(5)	3.55	0.39	3.72	0.48	1.64E-07	5.38E-09	1.19	0.18	4.62E-08	0.32
NNIa	(6)	3.54	0.39	4.23	0.46	1.70E-07	5.57E-09	0.32	0.05	4.80E-08	0.08
NNI _{meiosis-anthesis}	(6)	3.55	0.39	4.14	0.45	1.63E-07	5.35E-09	0.42	0.06	4.60E-08	0.10

Table 7. Variance components of interaction for model on grain number tested for each year on data of the Core Collection MET. Models and abbreviations are described in the materials and methods (Table 2). $\hat{\sigma}_{G2007}^2$ is the genotypic variance in 2007 ($\hat{\sigma}_{G2008}^2$ in 2008), and $\hat{\sigma}_{\text{interaction} \times Y 2007}^2$ the variance of the genotype-by-N stress characterization interaction effect in 2007 ($\hat{\sigma}_{\text{interaction} \times Y 2008}^2$ in 2008). In model (4), the interaction variance is $\sigma_{G \times E 2007}^2$ in 2007 and $\sigma_{G \times E 2008}^2$ in 2008. In model (5), the interaction variance corresponds $\sigma_{G \times T \times Y 2007}^2$ in 2007 ($\sigma_{G \times T \times Y 2008}^2$ for 2008). In model (6), the interaction variance is $\sigma_{B 2007}^2$ in 2007 ($\sigma_{B 2008}^2$ in 2008).

Ability of crop-model indicators to capture the differences in GxE interactions variances per year

The GxE interactions variance is much higher in 2008 than in 2007 (Table 7). Even though similar N fertilization regimes were applied in 2007 and 2008 (Supp. Table1), cooler temperature in 2008 slowed crop development thus allowing greater biomass and yield potential (Table 3, Fig 1) but also generating greater GxE interactions in response to low inputs (Table 7). The variances per year of the genotypic interaction with the classical characterization of N stress (Level of input, Nsupply, and Nsupply+Nsoil) showed that they did not capture the GxE interactions signal, neither in 2007 nor 2008. By contrast, modeling the interaction between genotype and crop-model indicators for N stress allowed capturing the high GxE signal in 2008, and even the lower one in 2007 also.

Genomic prediction gains by modeling the interaction with N stress using crop-model indicators

Genomic predictions of the response to crop-model based environment characterizations (NNIa, NNli_{meiosis-anthesis}, and Environment Cluster) were compared to basic characterizations (Level of inputs, N supply or N supply + N soil). The crop-model based indicators captured up to almost 10 times more interaction signal (from 3.9 to 6.9% of the additive genetic variance) compared to the other indicators which captured only 0.4 to 0.9% of the genetic additive variance. They allowed decreasing the error variance and the environmental variance with virtually no change in the genetic variance but a greater GxE interaction variance. Overall, the environmental effect was better captured with crop-model based indicators (Table 8).

The estimation of prediction accuracies using a testing set which contained either similar individuals (CVnewG), similar environment (CVnewE) or both (CVrandom) to the training set, showed that all models (additive and GxN models) performed accurately. Accuracies of the additive model ranged from 0.715 to 0.882 in CVrandom, from 0.417 to 0.568 in CVnewG, and from 0.728 to 0.89 in CVnewE (Table 9), which concords with the fact that phenotypic variance was mainly explained by the genetic variance, and a to a lesser extent by the GxE variance (Table 7). The G x Level of input had similar or slightly lower prediction accuracies than the additive model in all cross-validations. Considering the covariates N supply or N supply + N soil in the FR-gBLUP model globally provided similar prediction accuracies than the additive model in CVrandom, CVnewG and CVnewE. With crop-model based covariates (NNIa, NNli_{meiosis-anthesis} and Environment Cluster) to describe N stress in the FR-gBLUP model (models 8 and 9), the prediction accuracies were higher to those obtained with the additive model in CVrandom, and quite similar in CVnewG and CVnewE, on average across all environments. In 2008 particularly (*i.e.* when GxE were relatively high; Table7), their prediction accuracy gains were

higher than the additive or G x basic N stress characterizations models, with *e.g.* + 0.78% of accuracy gains compared to the additive model in CVrandom (instead of -0.28% in 2007).

	Model	σ_G^2	σ_{gb}	$\sigma_{\text{interaction}}^2$ (σ_{GT}^2 or σ_B^2)	σ_E^2	σ^2
Additive	(7)	2.25+/- 0.25	-	-	7.69+/- 3.29	2.06+/- 0.06
Level of Input	(8)	2.24+/- 0.25	-	0.013+/- 0.01	6.74+/- 3.02	2.05+/- 0.06
Nsupply	(9)	2.26+/- 0.25	0.22+/- 0.04	0.02+/-0.01	6.75+/- 3.02	2.02+/- 0.06
Nsupply + Nsoil	(9)	2.26+/- 0.25	0.13+/- 0.04	0.008+/- 0.006	6.75+/- 3.03	2.04+/- 0.06
Environment Cluster	(8)	2.22+/- 0.25	-	0.15+/-0.04	4.62+/- 2.18	1.90+/- 0.06
NNIa	(9)	2.28+/- 0.25	0.37+/- 0.06	0.11+/-0.02	3.65+/- 1.64	1.85+/- 0.06
NNI _{meiosis-anthesis}	(9)	2.29+/- 0.25	-0.37+/- 0.06	0.13+/-0.03	3.37+/- 1.51	1.83+/- 0.06

Table 8. Variance components of the genomic models considering grain-number response to N stress. Models and abbreviations are described in the materials and methods (Table 3). σ_G^2 is the variance of the genomic additive effect \mathbf{u}_G , σ_E^2 is the environmental variance and σ^2 the residual variance. $\sigma_{\text{interaction}}^2$ represents the G x N stress interaction variance, which corresponds to the G x Group of environment variance σ_{GT}^2 in Model 8 or the variance σ_B^2 of the genomic reaction norms \mathbf{b}_G to the environmental covariate \mathbf{x}_E in Model 9. For model 9 in particular, there is also a covariance σ_{gb} representing the relation between the additive genomic additive effects \mathbf{u}_G (intercepts) and the reaction norms \mathbf{b}_G (slopes).

CHAPITRE 2

CVrandom	Model	EM07LN	EM07HN	LM07LN	LM07HN	JO07LN	JO07HN	Average in 2007	EM08LN	EM08HN	LM08LN	LM08HN	JO08LN	JO08HN	Average in 2008	Average
Additif	(7)	0.86+/- 0.003	0.808+/- 0.003	0.811+/- 0.005	0.882+/- 0.003	0.807+/- 0.005	0.858+/- 0.003	0.838	0.822+/- 0.005	0.827+/- 0.002	0.82+/- 0.004	0.797+/- 0.005	0.715+/- 0.004	0.727+/- 0.004	0.785	0.811
Level of input	(8)	0.859+/- 0.003	0.807+/- 0.003	0.811+/- 0.005	0.88+/- 0.003	0.805+/- 0.005	0.858+/- 0.004	0.837	0.823+/- 0.005	0.825+/- 0.002	0.816+/- 0.004	0.796+/- 0.005	0.713+/- 0.004	0.723+/- 0.004	0.783	0.81
Nsupply	(9)	0.857+/- 0.002	0.8+/- 0.003	0.809+/- 0.004	0.886+/- 0.003	0.817+/- 0.003	0.858+/- 0.002	0.845	0.818+/- 0.004	0.82+/- 0.003	0.823+/- 0.003	0.79+/- 0.003	0.733+/- 0.003	0.729+/- 0.005	0.786	0.812
Nsupply + Nsoil	(9)	0.86+/- 0.002	0.806+/- 0.003	0.811+/- 0.003	0.881+/- 0.003	0.807+/- 0.006	0.86+/- 0.004	0.838	0.816+/- 0.005	0.828+/- 0.003	0.821+/- 0.005	0.792+/- 0.005	0.719+/- 0.003	0.73+/- 0.004	0.784	0.811
NNI _{meiosis-anthesis}	(9)	0.855+/- 0.003	0.802+/- 0.004	0.797+/- 0.005	0.881+/- 0.003	0.804+/- 0.005	0.858+/- 0.003	0.833	0.825+/- 0.005	0.803+/- 0.003	0.843+/- 0.004	0.822+/- 0.005	0.74+/- 0.004	0.733+/- 0.004	0.794	0.814
NNI _a	(9)	0.855+/- 0.003	0.803+/- 0.004	0.802+/- 0.005	0.879+/- 0.003	0.805+/- 0.006	0.859+/- 0.003	0.834	0.826+/- 0.005	0.805+/- 0.003	0.838+/- 0.004	0.816+/- 0.005	0.74+/- 0.004	0.733+/- 0.004	0.793	0.813
Environment Cluster	(8)	0.837+/- 0.003	0.809+/- 0.005	0.796+/- 0.005	0.876+/- 0.003	0.816+/- 0.004	0.85+/- 0.003	0.831	0.804+/- 0.002	0.821+/- 0.005	0.833+/- 0.004	0.808+/- 0.005	0.751+/- 0.009	0.745+/- 0.005	0.794	0.812

CVnewG		EM07LN	EM07HN	LM07LN	LM07HN	JO07LN	JO07HN	Average in 2007	EM08LN	EM08HN	LM08LN	LM08HN	JO08LN	JO08HN	Average in 2008	Average
Additif	(7)	0.568+/- 0.013	0.563+/- 0.011	0.549+/- 0.013	0.553+/- 0.012	0.563+/- 0.012	0.553+/- 0.012	0.558	0.489+/- 0.009	0.516+/- 0.011	0.499+/- 0.011	0.489+/- 0.01	0.435+/- 0.013	0.417+/- 0.011	0.474	0.516
Level of input	(8)	0.568+/- 0.013	0.563+/- 0.011	0.549+/- 0.013	0.553+/- 0.012	0.563+/- 0.012	0.554+/- 0.012	0.558	0.489+/- 0.01	0.516+/- 0.011	0.499+/- 0.011	0.488+/- 0.01	0.435+/- 0.013	0.417+/- 0.011	0.474	0.516
Nsupply	(9)	0.576+/- 0.008	0.571+/- 0.006	0.56+/- 0.007	0.544+/- 0.008	0.553+/- 0.007	0.548+/- 0.008	0.559	0.504+/- 0.007	0.53+/- 0.007	0.492+/- 0.011	0.485+/- 0.011	0.431+/- 0.014	0.413+/- 0.013	0.476	0.517
Nsupply + Nsoil	(9)	0.563+/- 0.012	0.559+/- 0.01	0.552+/- 0.012	0.535+/- 0.011	0.537+/- 0.013	0.533+/- 0.01	0.547	0.485+/- 0.014	0.497+/- 0.01	0.477+/- 0.012	0.468+/- 0.015	0.426+/- 0.009	0.406+/- 0.012	0.460	0.503
NNI _{meiosis-anthesis}	(9)	0.564+/- 0.013	0.564+/- 0.011	0.54+/- 0.013	0.552+/- 0.012	0.558+/- 0.013	0.554+/- 0.012	0.555	0.491+/- 0.01	0.499+/- 0.011	0.495+/- 0.01	0.484+/- 0.01	0.449+/- 0.013	0.42+/- 0.011	0.473	0.514
NNI _a	(9)	0.563+/- 0.013	0.566+/- 0.01	0.54+/- 0.013	0.552+/- 0.012	0.558+/- 0.013	0.553+/- 0.012	0.555	0.491+/- 0.01	0.504+/- 0.011	0.5+/-0.01	0.488+/- 0.01	0.449+/- 0.014	0.42+/- 0.012	0.470	0.515
Environment Cluster	(8)	0.549+/- 0.01	0.583+/- 0.008	0.531+/- 0.01	0.535+/- 0.007	0.536+/- 0.012	0.541+/- 0.009	0.546	0.474+/- 0.011	0.497+/- 0.007	0.478+/- 0.007	0.472+/- 0.009	0.427+/- 0.009	0.409+/- 0.008	0.460	0.503

CHAPITRE 2

CVnewE		EM07LN	EM07HN	LM07LN	LM07HN	JO07LN	JO07HN	Average in 2007	EM08LN	EM08HN	LM08LN	LM08HN	JO08LN	JO08HN	Average in 2008	Average
Additif	(7)	0.867+/- 0.003	0.815+/- 0.002	0.822+/- 0.004	0.89+/- 0.001	0.824+/- 0.001	0.865+/- 0.001	0.847	0.832+/- 0.002	0.834+/- 0.001	0.826+/- 0.007	0.802+/- 0.005	0.728+/- 0.008	0.734+/- 0.006	0.793	0.82
Level of input	(8)	0.863+/- 0.004	0.815+/- 0.001	0.819+/- 0.005	0.888+/- 0.002	0.822+/- 0.001	0.862+/- 0.001	0.845	0.833+/- 0.002	0.833+/- 0.001	0.819+/- 0.006	0.802+/- 0.005	0.723+/- 0.006	0.723+/- 0.004	0.789	0.817
Nsupply	(9)	0.869+/- 0.003	0.808+/- 0.004	0.824+/- 0.003	0.882+/- 0.003	0.821+/- 0.002	0.86+/- 0.003	0.844	0.821+/- 0.003	0.828+/- 0.003	0.825+/- 0.005	0.798+/- 0.009	0.729+/- 0.008	0.741+/- 0.007	0.790	0.817
Nsupply + Nsoil	(9)	0.864+/- 0.004	0.81+/- 0.003	0.818+/- 0.004	0.89+/- 0.002	0.824+/- 0.002	0.865+/- 0.002	0.845	0.825+/- 0.003	0.831+/- 0.002	0.839+/- 0.005	0.802+/- 0.004	0.733+/- 0.005	0.743+/- 0.005	0.796	0.82
NNI _{meiosis-anthesis}	(9)	0.857+/- 0.005	0.811+/- 0.002	0.806+/- 0.007	0.888+/- 0.001	0.82+/- 0.002	0.865+/- 0.001	0.841	0.834+/- 0.002	0.808+/- 0.004	0.839+/- 0.01	0.821+/- 0.009	0.734+/- 0.009	0.736+/- 0.007	0.795	0.818
NNI _a	(9)	0.858+/- 0.005	0.812+/- 0.002	0.811+/- 0.006	0.885+/- 0.001	0.82+/- 0.002	0.865+/- 0.001	0.842	0.835+/- 0.002	0.803+/- 0.004	0.839+/- 0.009	0.82+/- 0.008	0.728+/- 0.009	0.734+/- 0.007	0.793	0.8175
Environment Cluster	(8)	0.847+/- 0.008	NaN+/-NA	0.791+/- 0.009	0.87+/- 0.003	0.817+/- 0.003	0.847+/- 0.004	0.834	0.803+/- 0.006	0.822+/- 0.004	0.852+/- 0.005	0.829+/- 0.005	0.744+/- 0.011	0.75+/- 0.01	0.800	0.816

Table 9. Prediction accuracies for grain number in the 4 cross-validation schemes for all N stress characterizations. The 4 cross-validations split the training and testing sets so that testing set contained either similar individuals (CVnewG), similar environment (CVnewE) or both (CVrandom) to the training set. Prediction accuracies were measured as the correlation between observed and predicted values.

Discussion

Crop modelling provides N stress indicators well suited for modelling GxE interactions

In a context of reasoning fertilization for reducing nitrate leaching and glasshouse gas emission, the characterization of N stress in the field is a prerequisite for breeding wheat varieties more adapted to limited N supply (Justes et al., 1994; Laperche et al., 2007). Despite the availability of crop models to simulate N status in wheat crops (Azodyn, APSIM, SIRIUS), they have rarely been used yet for environmental characterization of N stress (Lacaze and Roumet, 2004). This study aimed at evaluating the ability of crop-model simulations to (i) assess the N stress experienced by crops in field trials, taking account for the soil conditions, the climatic conditions, and N applications and to (ii) allow unraveling of genotype x environment (GxE) interactions in multi-environment trials. Here, crop modeling successfully allowed estimation of N indicators, when measured (Fig. 2). The estimation of such indicators in trials where they were not measured successfully allowed the characterization of genotype x N stress interactions (Table 6) and to extract N-specific interaction from the complete GxE interaction (Tables 7-8). Crop model based indicators of N stress were also good predictors of the environmental effect (Table 7). Overall, the indicators based on crop-model simulations for the cultivar Soissons better characterized N stress than basic information such as the treatment ('Level of input') or the amount of N available ('Nsupply + Nsoil'; Table 6). Some environments with high inputs were actually more stressful than others with limited input levels (Table 5), as environmental factors (e.g. temperature, radiation, rainfall) can influence greatly how the level of input is impacting crops.

Across all cross-validations, NNI at anthesis (NNI_a) and between meiosis and anthesis ($NNI_{meiosis-anthesis}$) were the crop-model N environmental covariates which performed the best for unraveling GxE interactions on grain number. Both reflect N stress at crucial stages for grain number determination. The environmental characterization of N stress using NNI, as proposed by (Lecomte, 2008), remains relatively rarely used as it requires time consuming, destructive, and expensive measurements. In this context, simulating NNI with crop models represents an advantageous and original approach to assess the N nutrition status of a trial. In addition, simulating trials with crop models can be used to characterize different stresses and their interactions, and could be used for multiple purposes by breeding programs (Cooper et al., 2014; Chenu, 2015),

How much gain in prediction accuracy to expect from GxE modelling?

In terms of genomic predictions, the accuracies of the statistical model were expected to be very high with the additive model (from 0.715 to 0.882 in CVrandom; from 0.417 to 0.568 in CVnewG and from 0.728 to 0.89 in CVnewE) due to the high genetic variance observed for this dataset. When

considering in addition the interaction with N stress, the model got slightly improved but not in all trials. The differences in prediction accuracies between models (additive model vs GxE models) were small, since the heritability (Cullis et al., 2006) calculated for grain number was very high (0.96) and the variance of the GxN was relatively low (less than 10% of the genetic additive variance in all the tested models). Despite significant G x Management system effects observed within the same trials, their variance represented 6.3% of the genetic additive effect for grain number (Bordes et al., 2013). Even though the experiments were designed to have a wide range of N stress, the tested population came from a Core Collection (Balfourier et al., 2007) and had a very high genetic variability in terms of yield and grain number, and an unexpectedly low GxN variance compared to the genetic additive variance (Table 8).

Despite the low GxN interactions observed, the average prediction accuracy across all trials was increased when using models including crop-model N indicators (NNIa, NNI_{meiosis-anthesis}, and Environment Cluster) for CV random, *i.e.* when random genotype and environment combinations were excluded from the training set of the model (Table 9). In addition for 2008, *i.e.* when most GxE and GxN interactions occurred, greater gains were made (+ 0.78% in CVrandom). In studies with greater GxE variance, higher prediction gains have been observed. For instance, an average prediction gain of 11.1% in accuracies was found in wheat trials where the GxE variance represented 63% of the genetic additive variance (compared to 0.4%-9% in our study), and where the prediction accuracy of the additive model were lower than 0.5 (whereas almost all above this value in our study) (Heslot et al., 2014). Overall, the dataset used in this study was characterized by a very high genetic variation and an atypically low GxE interaction.

Genomic predictions of the response to crop-model stress indicators for N stress per year

However, by looking more closely at the GxE interactions variance components, the comparison of the 2 years of experiments revealed higher GxE interactions in 2008 (Table 7). The lower grain number values and variability in 2007 compared to 2008 (Fig.1) are probably due to a stronger environmental factor effect, possibly heat (Table 1); which did not allow as much GxE interactions. Oppositely, radiation deficit around meiosis and excess of water during booting which happened in 2008 (Communication Arvalis) are stresses known to generate GxE (Fischer, 1975; Reynolds et al., 2002; Hossain and Uddin, 2011).

Therefore, while any trend was hardly noticeable in 2007, more differences appeared the second year. In 2008 crop-model N stress indicators differed from the basic characterizations: they better captured the GxE interactions signal (Table 7) and allowed better prediction accuracies of grain number (Table 9) in CVrandom and CVnewE. The prediction accuracies of the genomic models for the

genotype by crop-model N stress indicators (NNI_a, Environmental Cluster, NNI_{meiosis-anthesis}) were modest but were consistently higher relatively to the additive model, the G x N supply, G x (N supply + N soil) or G x Level of input models in 2008. Therefore, it might be that crop-model based N stress indicators actually accounted for other tresses encountered during plant growth such as anoxia of radiation deficiency. For CVnewG, the losses of prediction accuracies in 2008 may be due to the low genetic variability of N response despite the use of a Core Collection: genotypes may have been not related enough to predict the little variation in their reaction norms. The definition of N stress indicators based on the simulation of the biomass, such as (Chapman, 2008b) for drought, thus allowed prediction of the response to N stress of tested individuals, and should be able to predict for new individuals if they shared more relationship with the training individuals.

Potential of genomics and crop modeling for predicting GxE interactions

Handling GxE interactions has always been a challenge for plant breeding. Methodologies have been developed to map QTLs that control the variation of traits with differential effects in different environments, or QTLs that control the variation of crop model parameters (Yin et al., 2003; Bogard et al., 2014). Furthermore, integrating crop growth models with whole genome prediction of eco-physiologic parameters seems to be a promising approach (Technow et al., 2015). However, the cost of phenotyping, when using precision traits to train the statistical model, and also the computational power required for the analysis, which can restrict the number of markers to consider, are some limitations to this method.

Another approach is to model GxE interactions in the genomic prediction model itself. As we include the variance covariance of individuals in the genomic selection model using a kinship matrix, some authors have used a resemblance matrix for the environments calculated from a large number of environmental covariates (Burgueño et al., 2012a; Jarquín et al., 2014b). Because of the difficulty to predict the multiplicity of environmental covariates that will be relevant in the future, our study focuses on one single stress, nitrogen satisfaction. We showed the possibility to use a crop model to characterize environments based on the simulated biomass for nitrogen stress, such as (Chapman et al. (2000a) and Chenu et al. (2011) for drought. As long as GxE interactions, and in particular in response to the stress of interest, are large enough, we can obtain genomic predictions of the reaction norms to a stress characterized by a crop model. Using diversity panels, crop model characterization of nitrogen stress actually allowed genomic predictions of the reaction norms to N stress when GxE signal was high, in our case in 2008. In the literature, using the elite material or breeding advanced material where GxE interactions represented a large proportion of the total variance (about 20 to 40 %), factorial regression models with appropriate environmental covariate(s) explained large variations of the GxE interactions (van Eeuwijk et al., 1995; Voltas et al., 2005; Lopes

et al., 2012; Ciancaleoni et al., 2016, Ly et al., in press). We expect gains on prediction accuracy of GxE interactions when using crop models to characterize environments.

Conclusion

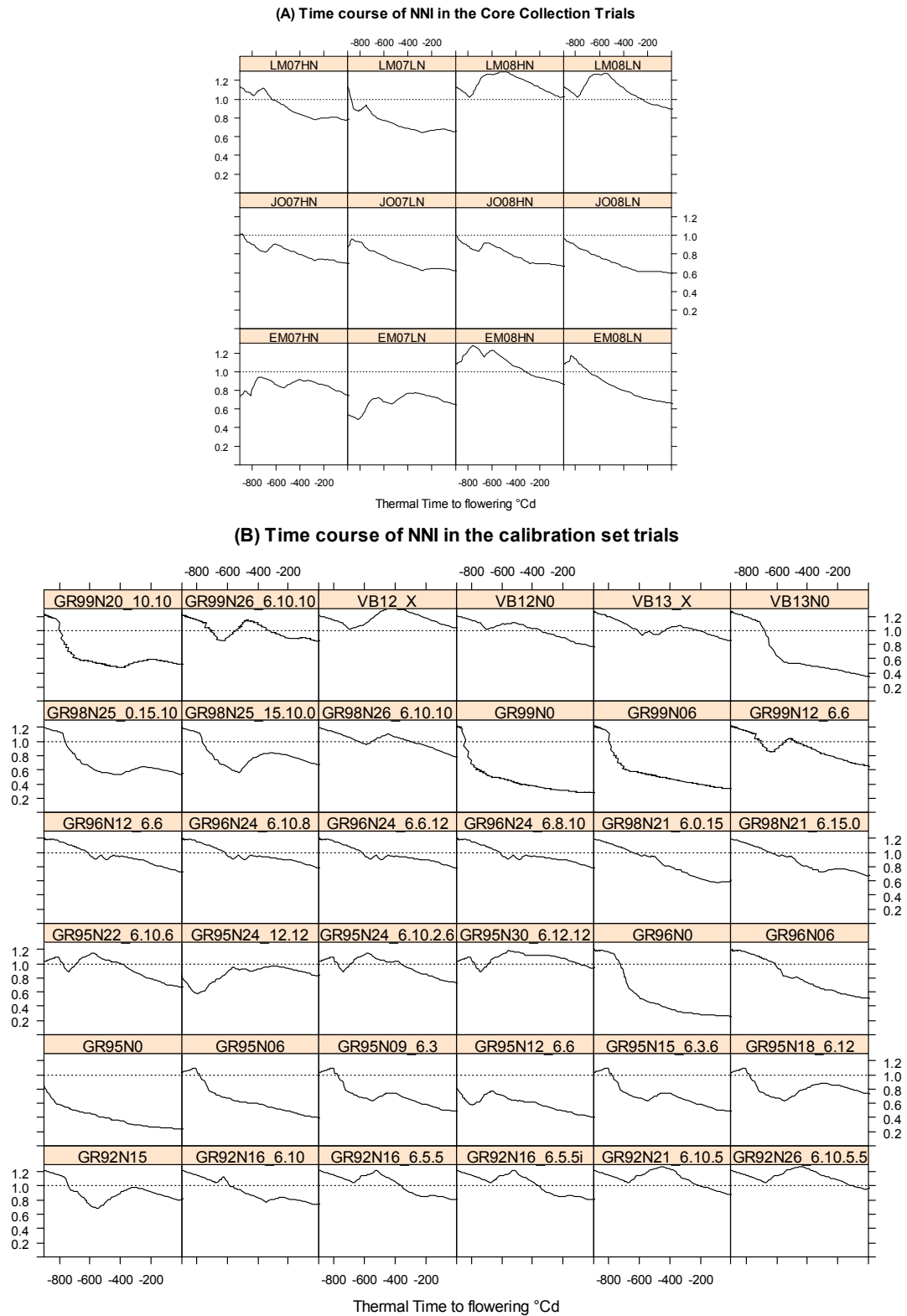
The use of crop-model based stress indicators clearly proved powerful in explaining a significant part of GxE in multi-environment wheat trials. Integrating these indicators as covariates in our recently developed FR-gBLUP model improved the genomic prediction accuracy in MET where GxE was observed (*i.e.* 2008 in this study). While this study focused on a dataset with great genetic variability and atypically poor GxE interaction, the developed method appears as promising to enhance breeding programs with genomic selection strategy integrating GxE interactions.

Supplementary data

Environment Code	Estrées-Mons (49°53'N, 3°00'E)				Le Moulon (45°86'N, 3°30'E)				Joze (45°86'N, 3°30'E)			
	2007		2008		2007		2008		2007		2008	
	HN	LN	HN	LN	HN	LN	HN	LN	HN	LN	HN	LN
	EM07HN	EM07LN	EM08HN	EM08LN	LM07HN	LM07LN	LM08HN	LM08LN	JO07HN	JO07LN	JO08HN	JO08LN
Sowing date	17 oct		17 oct		26 oct		23 oct		29 oct		25 oct	
N applying	33		40		33		40		35		50	
1 Tillering												
2	60	40	80	40	60	40	80	40		40		50
Jointing												
3	60	40	40		60	40	40		40		40	
Booting												
4									50	50	40	40
Flowering												
Seeding densely/m ²	220	220	220	220	220	220	220	220	280	170	280	170
Diseases treatment	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes
Growth regulator	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes

Supplementary Table 1 From (Bordes et al., 2013) Description of experiments conducted at Mons, Le Moulon and Joze, France, for the 2006-2007 (2007) and 2007-2008 (2008) growing seasons on a collection of bread wheats.

CHAPITRE 2



Supplementary Figure 1: Change in NNI over thermal time in the experiments used for the model calibration (a) and in the Core Collection multi-environment trials (b). NNI were simulated with the APSIM-Wheat crop model and presented from 900°Cd before anthesis to anthesis.

References

- Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. p. 267–281. *In* Petrov, B.N., Csaki, F. (eds.), Second International Symposium on Information Theory. Akademiai Kiado, Budapest.
- Balfourier, F., and G. Charmet. 1991. Relationships between agronomic characters and ecogeographical factors in a collection of French perennial ryegrass populations. *Agronomie* 11: 645–657 Available at <http://www.agronomy-journal.org/10.1051/agro:19910802>.
- Balfourier, F., V. Roussel, P. Strelchenko, F. Exbrayat-Vinson, P. Sourdille, G. Boutet, J. Koenig, C. Ravel, O. Mitrofanova, M. Beckert, and G. Charmet. 2007. A worldwide bread wheat core collection arrayed in a 384-well plate. *Theor. Appl. Genet.* 114: 1265–1275.
- Berger, J.D., and N.C. Turner. 2007. The Ecology of Chickpea: evolution, distribution, stresses and adaptation from an agro-climatic perspective. p. 47–71. *In* Yadav, S., Redden, R., Chen, W., Sharma, B. (eds.), Chickpea Breeding and Management. CAB International: Wallingford.
- Bogard, M., C. Ravel, E. Paux, J. Bordes, F. Balfourier, S.C. Chapman, J. Le Gouis, and V. Allard. 2014. Predictions of heading date in bread wheat (*Triticum aestivum* L.) using QTL-based parameters of an ecophysiological model. *J. Exp. Bot.* 65(20): 5849–5865 Available at <http://jxb.oxfordjournals.org/lookup/doi/10.1093/jxb/eru328>.
- Bordes, J., C. Ravel, J.P. Jaubertie, B. Duperrier, O. Gardet, E. Heumez, a L. Pissavy, G. Charmet, J. Le Gouis, and F. Balfourier. 2013. Genomic regions associated with the nitrogen limitation response revealed in a global wheat core collection. *Theor. Appl. Genet.* 126(3): 805–22 Available at <http://www.ncbi.nlm.nih.gov/pubmed/23192671>.
- Brisson, N., P. Gate, D. Gouache, G. Charmet, F.-X. Oury, and F. Huard. 2010. Why are wheat yields stagnating in Europe? A comprehensive data analysis for France. *F. Crop. Res.* 119(1): 201–212.
- Burgueño, J., G. de los Campos, K. Weigel, and J. Crossa. 2012. Genomic Prediction of Breeding Values when Modeling Genotype × Environment Interaction using Pedigree and Dense Molecular Markers. *Crop Sci.* 52(2): 707 Available at <https://www.crops.org/publications/cs/abstracts/52/2/707>.
- Chapman, S.C. 2008. Use of crop models to understand genotype by environment interactions for drought in real-world and simulated plant breeding trials. *Euphytica* 161(1-2): 195–208 Available at <http://www.springerlink.com/index/10.1007/s10681-007-9623-z>.
- Chapman, S.C., M. Cooper, D.G. Butler, and R.G. Henzell. 2000a. Genotype by environment interactions affecting grain sorghum. I. Characteristics that confound interpretation of hybrid yield. *Aust. J. Agric. Res.* 51(2): 197.
- Chapman, S.C., G.L. Hammer, D.G. Butler, and M. Cooper. 2000b. Genotype by environment interactions affecting grain sorghum. Temporal sequences and spatial patterns in the target population of environments. *Aust. J. Agric. Res.* 51(2): 223.
- Chauhan, Y.S., and R.C.N. Rachaputi. 2014. Defining agro-ecological regions for field crops in variable target production environments: A case study on mungbean in the northern grains region of

- Australia. Agric. For. Meteorol. 194: 207–217 Available at <http://dx.doi.org/10.1016/j.agrformet.2014.04.007>.
- Chauhan, Y.S., K.F. Solomon, and D. Rodriguez. 2013. Characterization of north-eastern Australian environments using APSIM for increasing rainfed maize production. F. Crop. Res. 144: 245–255 Available at <http://dx.doi.org/10.1016/j.fcr.2013.01.018>.
- Chenu, K. 2015. Characterising the crop environments - Nature, significance and applications. p. 321–348. In Sadras, V., Calderini, D. (eds.), Crop Physiology. Applications for Genetic Improvement and Agronomy. Academic Press.
- Chenu, K., M. Cooper, G.L. Hammer, K.L. Mathews, M.F. Dreccer, and S.C. Chapman. 2011. Environment characterization as an aid to wheat improvement: interpreting genotype-environment interactions by modelling water-deficit patterns in North-Eastern Australia. J. Exp. Bot. 62(6): 1743–1755 Available at <http://www.ncbi.nlm.nih.gov/pubmed/21421705>.
- Chenu, K., R. Dehifard, and S.C. Chapman. 2013. Large-scale characterization of drought pattern: a continent-wide modelling approach applied to the Australian wheatbelt--spatial and temporal trends. New Phytol. 198(3): 801–20 Available at <http://www.ncbi.nlm.nih.gov/pubmed/23425331>.
- Ciancaleoni, S., A. Onofri, R. Torricelli, and V. Negri. 2016. Broccoli yield response to environmental factors in sustainable agriculture. Eur. J. Agron. 72: 1–9 Available at <http://www.sciencedirect.com/science/article/pii/S1161030115300307>.
- Cormier, F., S. Faure, P. Dubreuil, E. Heumez, K. Beauchêne, S. Lafarge, S. Praud, and J. Le Gouis. 2013. A multi-environmental study of recent breeding progress on nitrogen use efficiency in wheat (*Triticum aestivum* L.). Theor. Appl. Genet. 126(12): 3035–48 Available at <http://www.ncbi.nlm.nih.gov/pubmed/24057081>.
- Cullis, B.R., a. B. Smith, and N.E. Coombes. 2006. On the design of early generation variety trials with correlated data. J. Agric. Biol. Environ. Stat. 11(4): 381–393.
- Denis, J.B. 1980. Analyse de la régression factorielle. Biom Praxim 20: 1–34.
- Van Eeuwijk, F. a., L.C.P. Keizer, and J.J. Bakker. 1995. Linear and bilinear models for the analysis of multi-environment trials: II. An application to data from the Dutch Maize Variety Trials. Euphytica 84(1): 9–22.
- Fischer, R.A. 1975. Yield Potential in a Dwarf Spring Wheat and the Effect of Shading. Crop Sci. 15(5): 607.
- Gilmour, A.R., B.J. Gogel, B.R. Cullis, and R. Thompson. 2006. ASReml User Guide Release 2.0 (VSN International Ltd, Ed.). Hemel Hempstead, HP1 1ES, UK.
- Habash, D.Z., S. Bernard, J. Schondelmaier, J. Weyen, and S. A. Quarrie. 2007. The genetics of nitrogen use in hexaploid wheat: N utilisation, development and yield. Theor. Appl. Genet. 114(3): 403–19 Available at <http://www.ncbi.nlm.nih.gov/pubmed/17180378>.

- Habier, D., R.L. Fernando, and D.J. Garrick. 2013. Genomic BLUP Decoded: A Look into the Black Box of Genomic Prediction. *Genetics* 194(3): 597–607 Available at <http://www.ncbi.nlm.nih.gov/pubmed/23640517>.
- Hammer, G.L., and R.C. Muchow. 1994. Assessing climatic risk to sorghum production in water-limited subtropical environments I. Development and testing of a simulation model. *F. Crop. Res.* 36(3): 221–234.
- Harrison, M.T., F. Tardieu, Z. Dong, C.D. Messina, and G.L. Hammer. 2014. Characterizing drought stress and trait influence on maize yield under current and future conditions. *Glob. Chang. Biol.* 20(3): 867–878.
- Henderson, C.R. 1975. Best Linear Unbiased Estimation and Prediction under a Selection Model. *Biometrics* 31(2): 423 Available at <http://www.ncbi.nlm.nih.gov/pubmed/1174616>.
- Heslot, N., D. Akdemir, M.E. Sorrells, and J.-L. Jannink. 2014. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theor. Appl. Genet.* 127(2): 463–480 Available at <http://link.springer.com/10.1007/s00122-013-2231-5>.
- Hirel, B., J. Le Gouis, B. Ney, and A. Gallais. 2007. The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches. *J. Exp. Bot.* 58(9): 2369–87 Available at <http://www.ncbi.nlm.nih.gov/pubmed/17556767>.
- Hossain, M. a, and S.N. Uddin. 2011. Mechanisms of waterlogging tolerance in wheat: Morphological and metabolic adaptations under hypoxia or anoxia. *Aust. J. Crop Sci.* 5(9): 1094–1101.
- Jarquín, D., J. Crossa, X. Lacaze, P. Du Cheyron, J. Daucourt, J. Lorgeou, F. Piraux, L. Guerreiro, P. Pérez, M. Calus, J. Burgueño, and G. de los Campos. 2014. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* 127(3): 595–607.
- Jeuffroy, M.-H., and C. Bouchard. 1999. Intensity and Duration of Nitrogen Deficiency on Wheat Grain Number. *Crop Sci.* 39: 1385–1393.
- Justes, E., B. Mary, J.-M. Meynard, J.-M. Machet, and L. Thelier-Huche. 1994. Determination of a critical nitrogen dilution curve for winter wheat crops. *Ann. Bot.* 74: 397–4007.
- Kohavi, R. 1995. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *In* International Joint Conference on Artificial Intelligence.
- Lacaze, X., and P. Roumet. 2004. Environment characterisation for the interpretation of environmental effect and genotype x environment interaction. *Theor. Appl. Genet.* 109(8): 1632–40.
- Laperche, A., J. Le Gouis, E. Hanocq, and M. Brancourt-Hulmel. 2007. Modelling nitrogen stress with probe genotypes to assess genetic parameters and genetic determinism of winter wheat tolerance to nitrogen constraint. *Euphytica* 161: 259–271.

- Lecomte, C. 2008. L'évaluation expérimentale des innovations variétales Proposition d'outils d'analyse de l'interaction génotype-milieu adaptés à la diversité des besoins et des contraintes des acteurs de la filière semences. Available at <http://prodinra.inra.fr/record/244938>.
- Lemaire, G., and J.-M. Meynard. 1997. Use of the Nitrogen Nutrition Index for the Analysis of Agronomical Data. p. 44–55. *In* Lemaire, G. (ed.), *Diagnosis of the Nitrogen Status in Crops*. Springer Berlin Heidelberg.
- Lobell, D.B., G.L. Hammer, K. Chenu, B. Zheng, G. McLean, and S.C. Chapman. 2015. The shifting influence of drought and heat stress for crops in northeast Australia. *Glob. Chang. Biol.* 21: 4115–4127 Available at <http://doi.wiley.com/10.1111/gcb.13022>.
- Löffler, C.M., J. Wei, T. Fast, J. Gogerty, S. Langton, M. Bergman, B. Merrill, and M. Cooper. 2005. Classification of Maize Environments Using Crop Simulation and Geographic Information Systems. *Crop Sci.* 45(5): 1708 Available at <https://www.crops.org/publications/cs/abstracts/45/5/1708>.
- Lopes, M.S., M.P. Reynolds, M.R. Jalal-Kamali, M. Moussa, Y. Feltaous, I.S. a Tahir, N. Barma, M. Vargas, Y. Mannes, and M. Baum. 2012. The yield correlations of selectable physiological traits in a population of advanced spring wheat lines grown in warm and drought environments. *F. Crop. Res.* 128: 129–136 Available at <http://dx.doi.org/10.1016/j.fcr.2011.12.017>.
- Meuwissen, T.H., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157(4): 1819–1829.
- Ortiz, R., K.D. Sayre, B. Govaerts, R. Gupta, G.V. Subbarao, T. Ban, D. Hodson, J.M. Dixon, J. Iván Ortiz-Monasterio, and M. Reynolds. 2008. Climate change: Can wheat beat the heat? *Agric. Ecosyst. Environ.* 126(1-2): 46–58 Available at <http://linkinghub.elsevier.com/retrieve/pii/S0167880908000194>.
- Paux, E., S. Faure, F. Choulet, D. Roger, V. Gauthier, J.P. Martinant, P. Sourdille, F. Balfourier, M.C. Le Paslier, A. Chauveau, M. Cakir, B. Gandon, and C. Feuillet. 2010. Insertion site-based polymorphism markers open new perspectives for genome saturation and marker-assisted selection in wheat. *Plant Biotechnol. J.* 8(2): 196–210.
- Pollak, L., and J.D. Corbett. 1993. Using GIS Datasets to Classify Maize-Growing Regions in Mexico and Central America. *Agron. J.* 85: 1133–1139 Available at <https://dl.sciencesocieties.org/publications/aj/pdfs/85/6/AJ0850061133>.
- R Development Core Team. 2003. R: a language and environment for statistical computing. Available at <http://www.r-project.org>.
- Reynolds, M.P., R. Trethowan, J. Crossa, M. Vargas, and K.D. Sayre. 2002. Physiological factors associated with genotype by environment interaction in wheat. *F. Crop. Res.* 75(2-3): 139–160.
- Rutkoski, J.E., J. Poland, J.-L. Jannink, and M.E. Sorrells. 2013. Imputation of unordered markers and the impact on genomic selection accuracy. *G3 (Bethesda)*. 3(3): 427–39.
- Schwarz, G. 1978. Estimating the dimension of a model. *Ann. Stat.* 6: 461–464.

- Technow, F., C.D. Messina, L.R. Totir, and M. Cooper. 2015. Integrating Crop Growth Models with Whole Genome Prediction through Approximate Bayesian Computation (I De Smet, Ed.). PLoS One 10(6): e0130855 Available at <http://dx.doi.org/10.1371/journal.pone.0130855>.
- VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91(11): 4414–23 Available at <http://www.ncbi.nlm.nih.gov/pubmed/18946147>.
- Voltas, J., H. López-Córcoles, and G. Borrás. 2005. Use of biplot analysis and factorial regression for the investigation of superior genotypes in multi-environment trials. Eur. J. Agron. 22(3): 309–324.
- Yin, X., P. Stam, M.J. Kropff, and A.H.C.M. Schapendonk. 2003. Crop modeling, QTL mapping, and their complementary role in plant breeding. Agron. J. 95(1): 90–98.
- Zadoks, J.C., T.T. Chang, and C.F. Konzak. 1974. A Decimal Code for the Growth Stages of Cereals. Weed Res. 14(6): 415–421.
- Zheng, B., S.C. Chapman, J.T. Christopher, T.M. Frederiks, and K. Chenu. 2015. Frost trends and their estimated impact on yield in the Australian wheatbelt. J. Exp. Bot. 66(12): 3611–3623 Available at <http://jxb.oxfordjournals.org/lookup/doi/10.1093/jxb/erv163>.
- Zheng, B., K. Chenu, M. Fernanda Dreccer, and S.C. Chapman. 2012. Breeding for the future: what are the potential impacts of future frost and heat events on sowing and flowering time requirements for Australian bread wheat (*Triticum aestivum*) varieties? Glob. Chang. Biol. 18(9): 2899–2914 Available at <http://doi.wiley.com/10.1111/j.1365-2486.2012.02724.x>.

Conclusion du Chapitre 2

Dans cette étude, nous avons cherché à évaluer la capacité d'un modèle de culture, APSIM, à (i) caractériser les stress azotés subis par un couvert de blé, en prenant en compte les conditions du sol, du climat et des pratiques culturales, et à (ii) décrire et expliquer l'interaction GxE dans des essais multi-environnementaux, et à (iii) fournir des indicateurs pertinents pour les prédictions génomiques des réponses au stress azoté.

Nous avons montré que des indicateurs simulés par le modèle de culture APSIM-Wheat version 7.6 (Wang et al., 2002; Holtzworth et al., 2014) basés sur l'index de nutrition azoté (INN) (Environment cluster, NNla, NNli_{meiosis-anthesis}) permettent de mieux capturer et expliquer non seulement les effets principaux environnementaux, mais aussi les effets d'interactions GxE, par rapport à des indicateurs de stress azotés que nous avons appelé « classiques » (tels que la quantité d'azote apporté, ou apporté + disponible dans le sol en sortie hiver, ou encore le mode de conduite azotée). Les modèles statistiques utilisant les indicateurs simulés présentent des termes d'interactions significatifs, et des critères de sélection de modèles AIC et BIC généralement plus faibles.

En termes de prédictions génomiques, nous avons mesuré des précisions de prédiction élevées, quel que soit le modèle, additif ou considérant une interaction entre les génotypes et le stress azoté. En raison de la forte héritabilité du caractère étudié (0.96), et le niveau très faible de la part d'interaction entre les génotypes et le stress azoté (de 0.4 à 9% de la variance génétique additive, selon l'indicateur de stress considéré), les gains des modèles avec interaction par rapport au modèle additif, qui montre déjà des niveaux de précisions de prédictions très importants, sont relativement modestes.

Néanmoins, nous avons remarqué une différence de gains entre les environnements 2007 et 2008. A des niveaux de gains toujours modestes, il apparaît une tendance selon laquelle il y a un avantage à modéliser la réponse à une covariable de stress azoté en 2008, plutôt qu'en 2007 (+0.78 en 2008 vs -0.28 en 2007 en CVrandom). En effet, en 2008, des températures plus fraîches auraient ralenti le développement des plantes, générant des biomasses et des potentiels de rendement plus importants, mais aussi plus d'interactions GxE. La variance de l'interaction GxE en 2008 est effectivement 22 fois plus élevée qu'en 2007. Que ce soit en 2007 ou en 2008, les indicateurs « classiques » ne parvenaient pas à expliquer les interactions GxE. A l'opposé, les caractérisations basées sur les modèles de culture ont permis de capturer les interactions GxE lorsque le signal était plus fort, en 2008. Ainsi les indicateurs simulés par APSIM semblent mieux intégrer les autres stress qui ont eu lieu en 2008 et ainsi caractériser les environnements ayant subis des stress multiples.

Références complémentaires

- Chapman, S.C. 2008. Use of crop models to understand genotype by environment interactions for drought in real-world and simulated plant breeding trials. *Euphytica* 161: 195–208.
- Chapman, S.C., G.L. Hammer, D.G. Butler, and M. Cooper. 2000. Genotype by environment interactions affecting grain sorghum. Temporal sequences and spatial patterns in the target population of environments. *Aust. J. Agric. Res.*
- Chenu, K., M. Cooper, G.L. Hammer, K.L. Mathews, M.F. Dreccer, and S.C. Chapman. 2011. Environment characterization as an aid to wheat improvement: interpreting genotype-environment interactions by modelling water-deficit patterns in North-Eastern Australia. *J. Exp. Bot.* 62(6): 1743–55 Available at <http://www.ncbi.nlm.nih.gov/pubmed/21421705>.
- Löffler, C.M., J. Wei, T. Fast, J. Gogerty, S. Langton, M. Bergman, B. Merrill, and M. Cooper. 2005. Classification of Maize Environments Using Crop Simulation and Geographic Information Systems. *Crop Sci.* 45(5): 1708 Available at <https://www.crops.org/publications/cs/abstracts/45/5/1708>.
- Muchow, R., M. Cooper, and G. Hammer. 1996. Characterizing environmental challenges using models. p. 349–364. *In* Cooper, M., Hammer, G. (eds.), *Plant adaptation and crop improvement*. CAB International: Wallingford, UK.
- Podlich, D.W., and M. Cooper. 1998. QU-GENE: a simulation platform for quantitative analysis of genetic models. *Bioinformatics* 14(7): 632–53 Available at <http://www.ncbi.nlm.nih.gov/pubmed/9730929>.

DISCUSSION GENERALE

TABLE DES MATIERES : DISCUSSION

1- La prédiction des normes de réaction aux stress environnementaux par une approche éco-physiologique et génétique.....	136
1.1 Le FR-gBLUP : Un modèle de prédiction génomique des réponses à un stress environnemental.....	136
1.2- La caractérisation environnementale des stress à l'aide de modèle de culture éco-physiologique	138
1.3- Intérêts de la prédiction génomique des normes de réactions aux stress environnementaux par rapport aux autres approches éco-physiologiques et génétiques.....	140
2- Les conditions d'application de notre approche éco-physiologique et génétique.....	143
3- De nouvelles possibilités pour sélectionner des variétés adaptées ?	145
4- Une recommandation variétale par environnement ?	147
5- Vers quelles perspectives est-ce que les avancées génomiques futures et le développement du phénotypage haut-débit conduiront la prédiction des interactions GxE ?	148
5.1- Vers une prédiction basée sur l'identification des polymorphismes causaux	148
5.2- ... et une amélioration des capacités de prédiction grâce au phénotypage à haut débit....	149
Références	150

1- La prédiction des normes de réaction aux stress environnementaux par une approche éco-physiologique et génétique

1.1 Le FR-gBLUP : Un modèle de prédiction génomique des réponses à un stress environnemental

Face au changement climatique et à l'évolution des pratiques culturales, les environnements dans lesquels le blé est cultivé sont en bouleversement. De ce fait, l'accélération de la sélection visée par la sélection génomique doit aujourd'hui intégrer la composante des interactions entre les génotypes et les environnements. Des modèles proposés dans la littérature (Burgueño et al., 2012a; Jarquín et al., 2014; Heslot et al., 2014) permettent de prédire les performances d'individus génotypés, dans des environnements caractérisés par un jeu de covariables environnementales. Cependant, la caractérisation précise de ces covariables dans les environnements futurs est difficile à prédire. La prédiction génomique des interactions GxE est plutôt à raisonner sur des tendances climatiques futures, prédites notamment par le GIEC (GIEC, 2014). Notre travail s'inscrit dans cette optique, de prédire les normes de réactions des génotypes à un stress environnemental particulier, à partir de leur information génomique.

Dans un premier temps, nous avons développé une série de critères permettant d'identifier l'intérêt d'une covariable environnementale pour étudier l'interaction des génotypes avec cette dernière. Nous nous sommes intéressés particulièrement (i) aux résultats de régression de la covariable considérée sur le nombre de grains, pour des variétés témoins de référence, Soissons et Camp-Rémy, (ii) à la connaissance agronomique de l'impact de ce stress sur le stade de développement auquel il intervient (Lecomte, 2008) (Figure 3- Introduction générale), (iii) à l'importance de l'interaction entre les génotypes et la covariable environnementale dans le panel étudié et (iv) à la validité du sens des normes de réactions. A partir de ces critères, nous avons identifié le stress hydrique durant la montaison comme facteur climatique d'intérêt pour notre étude des normes de réactions à un stress. Nous avons également choisi de nous intéresser au stress azoté qui était l'objet de l'expérimentation choisie. Alors que l'interaction avec le stress azoté, caractérisé par l'indice de nutrition azoté de Soissons, ne représente que 4.2% des interactions GxE, le stress hydrique pendant la montaison a permis de capturer 66% des interactions GxE.

Les résultats de simulations montrent que la part de variance des régressions génétiques à un stress environnemental joue un rôle majeur dans la capacité du modèle FR-gBLUP à prédire les normes de réactions au stress considéré. Effectivement, alors que la modélisation des normes de réaction au stress azoté n'apporte que très peu par rapport au modèle additif, les gains en précision de prédictions obtenus en modélisant une réponse au stress hydrique (+3.5% à +15.4%, toutes validations croisées confondues) sont comparables à ceux obtenus par (Jarquín et al., 2014; Heslot et

al., 2014), qui modélisent l'interaction GxE totale (+17% (Burgueño et al., 2012), +11.1% (Heslot et al., 2014), et +17 à +34% (Jarquín et al., 2014)). Cependant, ces études présentent des différences de populations, d'environnements, d'architectures génétiques des caractères, de choix de prédiction (des interactions complètes ou d'une part liée à un stress environnemental particulier) et également pour les précisions de prédiction du modèle additif utilisé comme référence. Les comparaisons entre les études pour les gains de précisions apportés par les différents modèles restent à nuancer.

Des améliorations du FR-gBLUP sont envisageables pour prendre en compte plusieurs stress qui auraient un effet important sur les interactions GxE. Toutefois, le nombre de covariables environnementales à considérer étant limité par souci d'estimation des paramètres, le reste des interactions non expliqué par les interactions avec la/les covariable(s) environnementale(s) d'intérêt pourrait être modélisé globalement comme dans le modèle proposé par Jarquín et al. (2014).

Ainsi, la démarche adoptée dans ce travail combine agronomie et génétique, dans le but de pouvoir interpréter plus aisément les prédictions de la part des interactions GxE sur laquelle nous nous sommes focalisée, ici l'interaction avec le stress hydrique durant la montaison.

1.2- La caractérisation environnementale des stress à l'aide de modèle de culture éco-physiologique

Dans l'objectif de prédire les réponses génomiques à un stress environnemental particulier, la caractérisation du stress d'intérêt est un enjeu essentiel. L'intérêt de la caractérisation des stress hydriques à l'aide de modèle de culture pour l'étude des interactions GxE a été démontré chez plusieurs espèces (sorgho, maïs, blé) dans les conditions climatiques australiennes mais aussi américaines (Chapman et al., 2000a; Löffler et al., 2005; Chenu et al., 2013). Notre étude propose de caractériser dans un réseau d'essai français, à partir d'un modèle de culture, le stress azoté pour lequel les mesures directes sont actuellement coûteuses et destructives (même si de nouvelles méthodes d'imagerie sont également en cours de développement (Vigneau et al., 2011)). Nous avons cherché à évaluer la qualité de la caractérisation des stress azotés par le modèle de culture, en comparaison avec des indicateurs « classiques » (niveau d'intrants, azote apporté ou azote apporté + azote disponible dans le sol en sortie hiver) pour expliquer les interactions GxE, et également pour prédire les normes de réactions des génotypes au stress azoté.

Nous avons montré que des indicateurs simulés par le modèle de culture APSIM-Wheat version 7.6 (Wang et al., 2002; Holtzworth et al., 2014) basé sur l'index de nutrition azoté (INN) (Environment cluster, NNla, NNli_{meiosis-anthesis}) permettent de mieux capturer et expliquer, non seulement les effets principaux environnementaux, mais aussi les effets d'interactions GxE, par rapport à des indicateurs de stress azotés « classiques ».

En termes de prédictions génomiques, la forte héritabilité du caractère étudié (0.96), et le niveau très faible de la part d'interaction entre les génotypes et le stress azoté (de 0.4 à 9% selon l'indicateur de stress considéré) permettent d'expliquer les gains de précision modestes mesurés par rapport au modèle additif. Une différence de gains est tout de même perceptible entre les environnements 2007 et 2008. En 2008, la variance d'interactions GxE représente plus de 22 fois celle de 2007. Les gains de précisions sont effectivement généralement plus importants dans les environnements de 2008 que dans ceux de 2007, et pour des indicateurs de stress azoté issus du modèle de culture plutôt que les indicateurs « classiques ». Du fait de la présence de nombreux stress générant de l'interaction GxE en 2008 (déficit de rayonnement et anoxie), il semble que les indicateurs simulés par APSIM pour le stress azoté ont pu intégrer les différents stress qui ont eu lieu en 2008 et ainsi caractériser les environnements ayant subis des stress multiples. En effet, l'INN est calculé à partir de la biomasse de la plante, qui peut être elle-même affectée par les autres stress qui ont lieu au cours du développement. Dans le modèle APSIM, le stress de rayonnement, en particulier, a un impact sur l'accumulation de la biomasse, et la manière dont cette accumulation est réduite peut également être affectée par le stress azoté.

DISCUSSION GENERALE

Dans notre étude, les indicateurs les plus pertinents sont l'INN à floraison et l'INN intégré entre la méiose et la floraison. La floraison représente effectivement un stade crucial pour la détermination du nombre de grains et le statut azoté à ce stade est donc bien représentatif de l'essai. Quant au fait de considérer un INN intégré entre 2 stades de développement, la caractérisation est théoriquement plus fine car elle intègre le moment du stress et son intensité dans le temps. Dans d'autres dispositifs expérimentaux, d'autres indicateurs de stress azotés seraient éventuellement à définir, en fonction des stress qui y ont lieu.

Notre étude conforte la proposition de caractériser le stress azoté par l'INN (Lecomte, 2008). Néanmoins, cette mesure nécessitant beaucoup de ressources, les modèles de cultures sont une bonne alternative pour estimer le statut azoté d'un essai sans mesures directes.

1.3- Intérêts de la prédiction génomique des normes de réactions aux stress environnementaux par rapport aux autres approches éco-physiologiques et génétiques

Devant la complexité des interactions GxE, de plus en plus d'auteurs utilisent les modèles de culture pour mieux appréhender l'environnement au sein duquel la plante évolue et comprendre comment la physiologie de la plante en est impactée. Plusieurs approches combinant des modèles de culture avec des modèles génétiques ont été développées.

L'approche que nous avons adoptée est la plus adaptée à l'exploitation de données existantes, venant par exemple de programmes de sélection. Cette approche statistique des réponses écophysiologiques consiste à caractériser l'environnement par la manière dont il impacte le développement de la culture, et à utiliser cette caractérisation dans les prédictions génomiques des réponses aux stress environnementaux.

Jarquín et al. (2014 et Heslot et al. (2014) définissent des covariables agro-climatiques en prenant en compte la précocité des variétés, pour pouvoir déterminer le stress impactant un stade de développement particulier. Heslot et al. 2014 utilise un modèle de culture (SIRIUS) pour déterminer la phénologie de différents groupes de variétés en fonction des conditions climatiques. Des travaux de caractérisation environnementale à partir de modèles de culture (Löffler et al., 2005; Chapman, 2008b; Chenu et al., 2011) ont démontré la possibilité d'expliquer les interactions GxE en définissant des environnements-type à partir de simulations journalières d'un indice de stress hydrique calculé en fonction de la biomasse de la plante et de paramètres environnementaux. Dans notre travail, nous avons cherché à développer des covariables de stress environnementaux basées sur la biomasse simulée par les modèles de culture. Ainsi, cette caractérisation est basée sur le développement journalier de la plante, face aux stress multiples possibles et à leurs potentielles interactions.

L'approche recourant à la définition de covariables environnementales nécessite alors un modèle génomique adapté pour prendre en compte l'interaction entre les génotypes et les covariables définies. (Jarquín et al., 2014b; Heslot et al., 2014; Tiezzi et al., 2015) visent, quant à eux, à prédire l'ensemble des interactions GxE, en considérant toutes les covariables environnementales disponibles, et en faisant l'hypothèse que les réponses à différentes covariables sont indépendantes et qu'elles ont le même poids quel que soit le stress considéré. Dans notre travail de développement du FR-gBLUP, nous nous sommes focalisés sur la prédiction des normes de réactions à un stress environnemental identifié préalablement, pour mieux comprendre cette part d'interaction GxE. Des extensions de ce modèle sont envisageables (Modèles 2 et 3 - Chapitre 1, Partie 1.3) pour considérer un plus grand nombre de stress environnementaux.

Une autre approche combinant des modèles de culture avec des modèles génétiques est en essor : il s'agit du gene-based modeling ou le QTL-based modeling. Dans le gene/QTL based modeling, le modèle écophysologique rend compte des spécificités de chaque variété grâce à des paramètres génétiques, non soumis aux interactions GxE, mais impliqués dans des relations non linéaires avec les stress environnementaux. Les interactions GxE sont alors une propriété émergente du modèle écophysologique. Une fois les paramètres génétiques estimés pour un panel de diversité, on peut ensuite s'intéresser à leur déterminisme génétique (supposé indépendant de l'environnement). On peut ensuite prédire la valeur des paramètres génétiques d'autres variétés grâce à la connaissance des QTL ou à des modèles de prédictions génomiques. Le modèle éco-physiologique peut alors être utilisé pour simuler le comportement de ces variétés dans différents environnements. Le gene/QTL based modeling repose sur l'idée que les modèles écophysologiques ont été construits de manière à ce que les paramètres génétiques soient indépendants de l'environnement.

Yin et al., (2003) ont cherché à explorer les possibilités de collaborations entre les modélisateurs et les généticiens, notamment dans la définition d'idéotypes pour des cultures à haut rendement. Ils proposent une stratégie pour établir un idéotype dans un environnement donné. Cela consiste à détecter les QTLs de paramètres génétiques d'un modèle éco-physiologique, et de cumuler les allèles favorables dans un environnement donné. Bogard et al., (2014), quant à eux, utilisent le gene/QTL based modeling pour prédire la date d'épiaison à partir d'un modèle éco-physiologique. Pour cela, il identifie des QTLs des paramètres de précocité, plus précisément de la vernalisation et de la photopériode, du modèle éco-physiologique. Avec 11 et 12 marqueurs génétiques, réciproquement, il explique 71 et 68% de la variance de ces paramètres. Les QTLs de vernalisation et photopériode permettent alors de prédire la date d'épiaison dans un jeu d'environnements indépendant, expliquant de 43 à 63% de sa variation. Technow et al. (2015) proposent une nouvelle méthode (Crop Growth Model Whole Genome Prediction) qui combine modèle de culture et sélection génomique, en utilisant tous les marqueurs du génome et en incorporant les modèles de culture directement dans l'estimation des effets des marqueurs. En termes de prédiction, la démarche proposée surpasse le gBLUP dans des simulations, que ce soit pour prédire un environnement déjà observé ou un nouvel environnement. Remarquons néanmoins que dans cette approche la faisabilité des calculs sollicite des ressources informatiques importantes et conduit à utiliser un nombre relativement faible de marqueurs (140).

Bien que l'approche gene/QTL based modeling présente l'avantage d'accéder à la compréhension du mécanisme génétique responsable des interactions GxE, elle nécessite des mesures éco-physiologiques, souvent complexes et coûteuses, éventuellement sur plateforme de phénotypage, et ceci pour un nombre de génotypes suffisant pour permettre la détection de QTL ou les prédictions

DISCUSSION GENERALE

genome-wide. Du fait du coût des mesures écophysiologiques, et du nombre important de paramètres génétiques à estimer (si l'on s'intéresse à des caractères plus complexes que la floraison), ces approches sont aujourd'hui encore peu répandues.

2- Les conditions d'application de notre approche éco-physiologique et génétique

La prédiction génomique de la norme de réaction à un stress environnemental est possible sous certaines conditions.

Comme ont pu le montrer nos études par simulations dans le chapitre 1, le modèle FR-GBLUP n'est valorisable que si la variabilité des normes de réactions des génotypes étudiés au stress environnemental considéré est suffisante. Plusieurs facteurs influencent la variabilité des pentes de réponse. Tout d'abord, en fonction de l'héritabilité du caractère étudié, le comportement des individus sera plus ou moins variable en fonction des environnements. Même si décomposer le rendement en différentes composantes permet de mieux appréhender l'impact des stress environnementaux, étudier les interactions GxE sur le nombre de grain s'est avéré assez difficile puisque ce caractère s'est révélé finalement peu interactif. De plus, la diversité génétique du panel influe sur la diversité possible des réponses à l'environnement. Toutefois, le choix de s'orienter sur une core collection ne maximise pas forcément la variabilité des réponses. Bien que cette collection ait été construite pour maximiser la variabilité génétique de marqueurs neutres, il s'est avéré qu'elle élargissait la gamme de variation additive, mais pas celle des interactions ou normes de réponse. De même, augmenter la gamme des stress environnementaux permettrait de mieux mettre en valeur les différences génotypiques des réponses à l'environnement, si elles existent. La prédiction des performances des individus pour de nouveaux environnements n'est valable que si ces environnements se situent dans la gamme des conditions testées. Ainsi, la prédiction pour des conditions climatiques futures ne serait pas forcément possible avec des expérimentations sous les conditions climatiques actuelles. Certains dispositifs expérimentaux permettent de simuler, en plein champ, plus de stress hautes températures, ou sécheresse par exemple et pourraient être utilisés pour accroître la gamme des stress étudiés (Araus and Cairns, 2014; Lopes et al., 2014). Par ailleurs, l'identification des covariables de stress influentes sur les interactions GxE peut s'avérer complexe dans des expérimentations plein-champ, sans contrôle des stress environnementaux. La connaissance agronomique ou éco-physiologique d'un lien entre la covariable de stress et le trait étudié, et de la variabilité génétique existante dans cette réponse, permettent de mieux raisonner le choix de la covariable. Ainsi, pour avoir une variabilité suffisante des normes de réaction à un stress particulier à prédire, il est intéressant d'avoir des caractères pour lesquels l'effet génétique additif n'explique pas toute la variabilité du caractère, pour lesquels l'interaction peut être expliquée physiologiquement, et où les environnements et les génotypes explorent une gamme diversifiée des conditions et des réponses au stress.

D'autres conditions d'application de notre approche sont liées au dispositif expérimental nécessaire pour l'estimation des paramètres dans un modèle de régression aléatoire. Martin et al. (2011) indiquent notamment que pour des jeux de données de petite taille (~100 observations), le meilleur ratio nombre d'individus / nombre d'observations par individus pour détecter significativement l'effet de régression aléatoire était de 0.5, par exemple 10 individus avec 20 observations chacun. Avec un nombre d'observations supérieur à 200, des ratios supérieurs à 0.5, c'est-à-dire plus d'individus avec chacun moins d'observations, procurent des plus grandes puissances de détection des interactions GxE. Par ailleurs, l'estimation des variances de la régression aléatoire peut s'avérer également difficile, du fait de l'hypothèse de linéarité des réponses au stress environnemental qui ne serait pas toujours justifiée. Dans ce cas, il serait possible d'envisager d'autres types de réponse, par exemple une courbe quadratique, ou encore une réponse en pente puis plateau.

L'application directe de notre approche de caractérisation environnementale serait également à réfléchir. Le fait de choisir un génotype représentatif fait l'hypothèse forte que les stress subis par ce génotype sont représentatifs des stress subis par les autres génotypes. La phénologie peut effectivement avoir un rôle majeur sur l'impact des stress, notamment dans le cas du stress hydrique (Gupta et al., 2001). Heslot et al. (2014) choisit de considérer 3 groupes de précocité. Cependant, la définition de ces groupes est difficile, notamment pour des génotypes qui, à un jour près, peuvent se retrouver dans un groupe de précocité ou un autre. L'approche la plus rigoureuse serait de construire des covariables environnementales par précocité, à partir d'une prédiction préalable de la précocité de chaque génotype. Bogard et al. (2014) ont montré les possibilités du gene/QTL based modeling pour la prédiction de la date d'épiaison. Ainsi, l'approche consisterait à utiliser les marqueurs génomiques impliqués dans la précocité pour modéliser la phénologie des différentes variétés. A partir de cette simulation de la phénologie, il serait alors possible de construire des variables de stress par environnement et par précocité. La caractérisation environnementale ainsi réalisée serait spécifique à chaque génotype et à sa précocité, et serait la caractérisation la plus fine de la tolérance intrinsèque au stress, parce qu'elle ne considérerait plus les effets d'échappement au stress lié à la précocité.

3- De nouvelles possibilités pour sélectionner des variétés adaptées ?

Ainsi en prenant en compte les différentes conditions d'applications que nous avons pu identifier dans le paragraphe précédent, il serait possible d'appliquer notre approche éco-physiologique et génétique pour sélectionner des variétés adaptées à un stress particulier, comme le stress hydrique ou le stress azoté.

La sélection de variétés adaptées à un stress rechercherait idéalement des variétés à haute performance et stables, c'est-à-dire performantes aussi bien en conditions stressées qu'en conditions non stressées. Le type de variétés à éviter absolument serait des variétés à faible performance en conditions non stressées et qui s'effondreraient avec un stress. Au-delà de ces évidences, les situations intermédiaires seraient alors plus spécifiques à un type d'environnement. Par exemple, des variétés performantes en conditions non stressées mais fortement impactées par un stress seront plus adaptées à des environnements favorables. Inversement, des variétés performantes en conditions stressées, mais ne valorisant pas la disponibilité de la ressource, seront plus adaptées à des environnements stressants. Les choix variétaux seraient alors dictés par des prévisions à long terme des environnements les plus probables de la culture, basés par exemple sur des analyses fréquentielles de climats passés et intégrant des tendances du changement climatique.

Dans le cas de la réponse au stress azoté, avec l'expansion de l'utilisation des fertilisants, la sélection variétale a sans doute développé des variétés qui répondent bien à l'azote en augmentant le nombre de grains. Aujourd'hui, les variétés présentent des normes de réaction au stress azoté assez similaires entre elles. Parmi ces variétés actuelles, à fort nombre de grains en conditions azotées non stressantes, il semble alors difficile d'entreprendre une démarche inverse, pour identifier des variétés peu sensibles à une baisse de N. Les travaux de Cormier et al. (2013) détectent, dans le même panel que celui que nous avons utilisé, plus d'interaction entre les géotypes et l'azote pour le rendement et la NUE (Nitrogen Use Efficiency) que pour le nombre de grains. Ceci suggère que pour une adaptation au stress azoté, les progrès génétiques à faire, ou qui ont été réalisés au cours des 20 dernières années de sélection, se situent non pas au niveau du nombre de grains, mais plus au niveau de la remobilisation de l'azote après floraison et du remplissage du grain.

L'intégration de la recherche de variétés adaptées dans les schémas de sélection variétale actuels serait possible, à partir du moment où les individus sont testés dans des essais multi-environnementaux. Toutefois, le coût de la sélection génomique par rapport au gain apporté ne serait peut-être pas avantageux. Dans l'objectif de la sélection génomique pour l'adaptation, les schémas devraient sans doute être repensés. Par exemple, si on souhaitait faire de la sélection pour des variétés avec une remobilisation de l'azote adaptées à différentes conditions azotées possibles,

DISCUSSION GENERALE

on chercherait à avoir une large population d'individus testés dans de nombreuses conditions, ou peut-être dans les conditions les plus extrêmes, sans forcément nécessiter de nombreuses répétitions. Les environnements seraient contrastés pour le niveau d'azote apporté après floraison, pour mieux évaluer l'impact du stress sur l'absorption et la remobilisation. Le choix du nombre d'environnements et du niveau de stress est à réfléchir en fonction de la courbe de réponse. En supposant une réponse linéaire au stress azoté, en théorie, deux environnements extrêmes devraient suffire à estimer les pentes de régression. Si la réponse est de type pente puis plateau, il faudrait plus d'environnements pour estimer la pente, le seuil et le niveau du plateau. La génomique permettrait alors d'évaluer les réponses des individus dès un stade assez précoce et d'appliquer une forte pression de sélection, afin de retenir idéalement les génotypes à fort effet additif génétique (ou forte ordonnée à l'origine) et à grande stabilité (faible pente).

4- Une recommandation variétale par environnement ?

En recommandation variétale, l'objectif est de pouvoir conseiller les variétés les plus adaptées à un environnement donné. La question est alors différente de la sélection, dans la mesure où l'environnement doit être caractérisé par tous les stress qui interviennent.

Un des enjeux principaux est d'évaluer le risque climatique et les stress potentiels pour la prochaine saison. La difficulté en Europe est la prise en compte de la multiplicité des stress, contrairement à d'autres pays comme l'Australie où la sécheresse est le stress majeur. Dans le cas du stress azoté, les stress attendus pour la prochaine saison sont liés en partie à l'environnement (reliquats azotés, minéralisation, mais aussi aux pratiques de fertilisation). Le niveau de stress est fonction du mode de culture que l'agriculteur choisit d'appliquer (ex : faibles intrants, AB). Dans le cas du stress azoté, les stress attendus pour la prochaine saison sont des stress que l'agriculteur choisit d'appliquer, mais qui dépendront également d'autres facteurs tels que la pluviométrie (nécessaire à l'absorption des engrais en culture non irriguée). Ce stress azoté peut interagir avec d'autres stress, mais le fait de définir l'environnement cible de sélection réduit les incertitudes liées à la prédiction des conditions environnementales futures. Pour la recommandation variétale, notre approche de prédiction des normes de réaction à des stress environnementaux est limitée à l'étude d'un stress majeur ou d'un stress appliqué. A partir des prédictions génomiques des normes de réactions aux stress environnementaux, et des probabilités de stress, par exemple par analyse fréquentielle de séries climatiques comme le fait Arvalis, il serait alors possible de mieux choisir les variétés à recommander.

Dans les autres cas, la détermination d'environnements types sur des données historiques dans le lieu à prédire permet d'évaluer les fréquences de différents patterns de stress (Chenu et al., 2013). Les modèles de cultures permettraient alors d'utiliser les fréquences attendues de différents scénarios de stress pour prédire le rendement des variétés. De plus, l'utilisation des prédictions climatiques futures peuvent également être utilisées comme indicateurs des conditions environnementales prochaines. En absence de stress majeur, mais d'effets faibles liés à chaque stress, le modèle génomique de régression factorielle FR-gBLUP serait moins adapté que l'extension proposée par (Jarquín et al., 2014). Néanmoins, en présence d'un ou deux stress majeurs, qui génèrent de l'interaction GxE, la modélisation explicite des normes de réactions aux stress choisis, et la modélisation d'un terme de GxE non expliqué par ces régressions pourraient permettre de mieux prédire les interactions GxE pour ces nouveaux environnements.

5- Vers quelles perspectives est-ce que les avancées génomiques futures et le développement du phénotypage haut-débit conduiront la prédiction des interactions GxE ?

5.1- Vers une prédiction basée sur l'identification des polymorphismes causaux ...

Au vue des avancées technologiques, il sera possible d'avoir prochainement accès au séquençage complet du génome du blé. De meilleures prédictions sont-elles alors à espérer ? Dans le cadre du projet 1000 génomes bovins, qui encourage l'accélération du gain génétique de la vache laitière, 1750 bovins ont été séquencés à une « profondeur » moyenne de 8.3 (génome bovin de 3Gb) (www.1000bullgenomes.com). En capturant les variations du déséquilibre de liaison, même les variations les plus fines, l'identification des polymorphismes causaux par génétique d'association est alors possible. La prédiction génomique pourrait alors se faire sur les polymorphismes causaux, et on attendrait des équations plus « universelles » que les prédictions actuelles qui ne sont fiables qu'au sein d'une race donnée.

Aujourd'hui, les polymorphismes identifiés chez le blé sont majoritairement des polymorphismes de substitution d'un nucléotide (SNP) (Paux et al., 2010; Wang et al., 2014), alors qu'il existe d'autres types de variations, telles que les variations du nombre de copies, d'insertions/délétions ou de présence/absence de séquences, qui interviennent dans le phénotype d'une plante et permettent d'expliquer une partie de variabilité du caractère phénotypique non expliquée par les SNPs (Springer et al., 2009; Saxena et al., 2014). A l'heure actuelle, il n'existe pas encore de méthode unique pour capturer complètement les variations structurales du génome. Le re-séquençage et l'imputation à partir de l'information des pedigrees seraient un moyen d'avoir des séquences complètes pour tous les individus génotypés. Chez le génome du blé, 5 à 6 fois plus grand que le génome des bovins (17Gb contre 3Gb), et dont le séquençage de référence est toujours en cours en cours (<http://www.wheatgenome.org/>), la possibilité de re-séquencer efficacement et totalement les individus à tester en sélection n'arrivera que plus tardivement.

Dans une approche de sélection d'individus adaptés à un stress environnemental, l'identification de polymorphismes impliqués dans la réponse à l'environnement semble primordiale. Les résultats de génétique d'association permettent alors d'explorer de nouvelles pistes pour l'étude du déterminisme génétique des caractères de tolérance à un stress. Pour la NUE, les différentes études d'association ont identifié des QTL colocalisant avec des enzymes impliquées dans l'absorption ou assimilation d'N, offrant ainsi une certaine validation de l'étude (Li et al., 2011; Bordes et al., 2013; Cormier et al., 2015). Les nouveaux QTLs qui ont été également identifiés constituent alors des voies à

approfondir. Néanmoins, il est possible que les variations phénotypiques observées soient dues à d'autres variations que celles de la séquence nucléotidique.

Dans les années à venir, la sélection variétale devra probablement prendre en compte le catalogue de petits ARN et l'épigénome (Molinier et al., 2006). Dans l'étude des réponses aux stress, l'adaptation épigénétique peut-être plus rapide que génétique et pourrait permettre une identification plus immédiate des épimutations favorables à certaines conditions environnementales. Cependant, les épimutations sont plus labiles et donc plus compliquées à conserver. Il existe toutefois la possibilité de fixer un caractère à long terme, en transformant une épimutation en mutation. La désamination des cytosines méthylées en thymine par exemple permet de transmettre la mutation lors de la réplication suivante. Toutefois, les investissements faits pour le développement de marqueurs épigénétiques en sélection devront surement être relatifs à l'ampleur des variations phénotypiques causées par les seules différences épigénétiques, qui peut s'avérer très faible.

A partir d'une information complète des polymorphismes génétiques et épi-génétiques, il serait possible d'identifier des polymorphismes causaux impliqués dans la tolérance aux stress. Les connaissances sur les mécanismes génétiques et épigénétiques de tolérance à des stress environnementaux qui vont alors s'accumuler, seraient un moyen de valider les pentes de régression que nous prédisons dans des modèles de régression factorielle, et justifieraient une sélection sur ces pentes. Une autre approche consisterait à pyramider de nombreux loci, à effets faibles, impliqués dans la tolérance à un stress. Toutefois, la tolérance à un stress environnemental est souvent un caractère complexe, hautement polygénique, et pour lesquels les effets des différents loci peuvent dépendre du fond génétique, ce qui rend la détection des polymorphismes causaux difficile.

5.2- ... et une amélioration des capacités de prédiction grâce au phénotypage à haut débit

Aujourd'hui, notamment en amélioration des plantes, face à la quantité d'information génomique grandissante, la sélection assistée par marqueur est limitée par la quantité et la qualité du phénotypage. En effet, des études de simulation (Yu et al., 2008; Clark et al., 2011; Hickey et al., 2014) et des études empiriques (VanRaden et al., 2009; Habier et al., 2010; Clark et al., 2012) ont démontré l'importance d'avoir une population de référence large, avec idéalement plusieurs milliers d'individus pour des céréales autogames par exemple (Hickey et al., 2014).

Au cours de la dernière décennie, l'essor du phénotypage à haut débit a été fulgurant (Araus and Cairns, 2014). Ces nouveaux outils, qu'ils soient au sol (Andrade-Sanchez et al., 2013; Busemeyer et al., 2013) ou aériens (Bendig et al., 2014; Chapman et al., 2014), portent de nombreux capteurs, permettant ainsi de caractériser les variétés (Svensgaard et al., 2014), mais aussi les environnements, de manière précise, rapide et non destructive. Les mesures multispectrales de rayonnements émis

par la culture, obtenues par ces capteurs permettent de calculer des indices utiles à la caractérisation des stress environnementaux sur une parcelle, par exemple pour le stress azoté (Vigneau et al., 2011; Agüera et al., 2012), thermique et hydrique (Zarco-Tejada et al., 2002; Moller et al., 2006; Liu et al., 2011; Cohen et al., 2012).

Le développement du phénotypage à haut-débit constitue non seulement une opportunité d'élargir nos populations de référence pour la sélection génomique, mais également un moyen d'accéder à plus grande échelle à une meilleure compréhension au champ des réponses éco-physiologiques des variétés aux stress environnementaux (Leinonen et al., 2006; Guilioni et al., 2008). L'intégration de ces nouvelles données dans la sélection génomique conduit à de nouveaux projets (Mackay et al., 2015) et pourrait amener à réfléchir à de nouvelles approches pour traiter les interactions GxE.

Références

- Agüera, F., F. Carvajal, and M. Pérez. 2012. MEASURING SUNFLOWER NITROGEN STATUS FROM AN UNMANNED AERIAL VEHICLE-BASED SYSTEM AND AN ON THE GROUND DEVICE. *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* XXXVIII-1/(C22): 33–37 Available at <http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XXXVIII-1-C22/33/2011/>.
- Andrade-Sanchez, P., M.A. Gore, J.T. Heun, K.R. Thorp, A.E. Carmo-Silva, A.N. French, M.E. Salvucci, and J.W. White. 2013. Development and evaluation of a field-based high-throughput phenotyping platform. *Funct. plant Biol.* 41(1): 68–79.
- Araus, J.L., and J.E. Cairns. 2014. Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* 19(1): 52–61 Available at <http://linkinghub.elsevier.com/retrieve/pii/S1360138513001994>.
- Bendig, J., A. Bolten, S. Bennertz, J. Broscheit, S. Eichfuss, and G. Bareth. 2014. Estimating Biomass of Barley Using Crop Surface Models (CSMs) Derived from UAV-Based RGB Imaging. *Remote Sens.* 6(11): 10395–10412 Available at <http://www.mdpi.com/2072-4292/6/11/10395/>.
- Bogard, M., C. Ravel, E. Paux, J. Bordes, F. Balfourier, S.C. Chapman, J. Le Gouis, and V. Allard. 2014. Predictions of heading date in bread wheat (*Triticum aestivum* L.) using QTL-based parameters of an ecophysiological model. *J. Exp. Bot.* 65(20): 5849–5865 Available at <http://jxb.oxfordjournals.org/lookup/doi/10.1093/jxb/eru328>.
- Bordes, J., C. Ravel, J.P. Jaubertie, B. Duperrier, O. Gardet, E. Heumez, a L. Pissavy, G. Charmet, J. Le Gouis, and F. Balfourier. 2013. Genomic regions associated with the nitrogen limitation response revealed in a global wheat core collection. *Theor. Appl. Genet.* 126(3): 805–22 Available at <http://www.ncbi.nlm.nih.gov/pubmed/23192671>.
- Burgueño, J., G. de los Campos, K. Weigel, and J. Crossa. 2012. Genomic Prediction of Breeding Values when Modeling Genotype × Environment Interaction using Pedigree and Dense Molecular Markers. *Crop Sci.* 52(2): 707 Available at <https://www.crops.org/publications/cs/abstracts/52/2/707>.

- Bussemeyer, L., D. Mentrup, K. Möller, E. Wunder, K. Alheit, V. Hahn, H. Maurer, J. Reif, T. Würschum, J. Müller, F. Rahe, and A. Ruckelshausen. 2013. BreedVision — A Multi-Sensor Platform for Non-Destructive Field-Based Phenotyping in Plant Breeding. *Sensors* 13(3): 2830–2847 Available at <http://www.mdpi.com/1424-8220/13/3/2830/>.
- Chapman, S.C. 2008. Use of crop models to understand genotype by environment interactions for drought in real-world and simulated plant breeding trials. *Euphytica* 161(1-2): 195–208 Available at <http://link.springer.com/10.1007/s10681-007-9623-z>.
- Chapman, S.C., G.L. Hammer, D.G. Butler, and M. Cooper. 2000. Genotype by environment interactions affecting grain sorghum. Temporal sequences and spatial patterns in the target population of environments. *Aust. J. Agric. Res.* 51(2): 223.
- Chapman, S., T. Merz, A. Chan, P. Jackway, S. Hrabar, M. Dreccer, E. Holland, B. Zheng, T. Ling, and J. Jimenez-Berni. 2014. Pheno-Copter: A Low-Altitude, Autonomous Remote-Sensing Robotic Helicopter for High-Throughput Field-Based Phenotyping. *Agronomy* 4(2): 279–301.
- Chenu, K., M. Cooper, G.L. Hammer, K.L. Mathews, M.F. Dreccer, and S.C. Chapman. 2011. Environment characterization as an aid to wheat improvement: interpreting genotype-environment interactions by modelling water-deficit patterns in North-Eastern Australia. *J. Exp. Bot.* 62(6): 1743–1755 Available at <http://www.ncbi.nlm.nih.gov/pubmed/21421705>.
- Chenu, K., R. Dehifard, and S.C. Chapman. 2013. Large-scale characterization of drought pattern: a continent-wide modelling approach applied to the Australian wheatbelt--spatial and temporal trends. *New Phytol.* 198(3): 801–20 Available at <http://www.ncbi.nlm.nih.gov/pubmed/23425331>.
- Clark, S. A, J.M. Hickey, H.D. Daetwyler, and J.H. van der Werf. 2012. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. *Genet. Sel. Evol.* 44(1): 4.
- Clark, S. A, J.M. Hickey, and J.H.J. van der Werf. 2011. Different models of genetic variation and their effect on genomic evaluation. *Genet. Sel. Evol.* 43(1): 18 Available at <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3114710&tool=pmcentrez&render type=abstract>.
- Cohen, Y., V. Alchanatis, a Prigojin, a Levi, V. Soroker, and Y. Cohen. 2012. Use of aerial thermal imaging to estimate water status of palm trees. *Precis. Agric.* 13(1): 123–140.
- Cormier, F., S. Faure, P. Dubreuil, E. Heumez, K. Beauchêne, S. Lafarge, S. Praud, and J. Le Gouis. 2013. A multi-environmental study of recent breeding progress on nitrogen use efficiency in wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 126(12): 3035–48 Available at <http://www.ncbi.nlm.nih.gov/pubmed/24057081>.
- Cormier, F., M. Throude, C. Ravel, J. Gouis, M. Leveugle, S. Lafarge, F. Exbrayat, N. Duranton, and S. Praud. 2015. Detection of NAM-A1 Natural Variants in Bread Wheat Reveals Differences in Haplotype Distribution between a Worldwide Core Collection and European Elite Germplasm. *Agronomy* 5(2): 143–151 Available at <http://www.mdpi.com/2073-4395/5/2/143/>.
- GIEC. 2014. Climate Change 2014 Synthesis Report.

- Guilioni, L., H.G. Jones, I. Leinonen, and J.P. Lhomme. 2008. On the relationships between stomatal resistance and leaf temperatures in thermography. *Agric. For. Meteorol.* 148(11): 1908–1912.
- Gupta, N., S. Gupta, and A. Kumar. 2001. Effect of water stress on physiological attributes and their relationship with growth and yield of wheat cultivars at different stages. *J. Agron. Crop Sci.* 186: 55–62 Available at <http://onlinelibrary.wiley.com/doi/10.1046/j.1439-037x.2001.00457.x/full>.
- Habier, D., L.R. Totir, and R.L. Fernando. 2010. A two-stage approximation for analysis of mixture genetic models in large pedigrees. *Genetics* 185(2): 655–70 Available at <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2881145&tool=pmcentrez&render type=abstract>.
- Heslot, N., D. Akdemir, M.E. Sorrells, and J.-L. Jannink. 2014. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theor. Appl. Genet.* 127(2): 463–480 Available at <http://link.springer.com/10.1007/s00122-013-2231-5>.
- Hickey, J.M., S. Dreisigacker, J. Crossa, S. Hearne, R. Babu, B.M. Prasanna, M. Grondona, A. Zambelli, V.S. Windhausen, K. Mathews, and G. Gorjanc. 2014. Evaluation of Genomic Selection Training Population Designs and Genotyping Strategies in Plant Breeding Programs Using Simulation. *Crop Sci.* 54(4): 1476 Available at <https://dl.sciencesocieties.org/publications/cs/abstracts/54/4/1476>.
- Jarquín, D., J. Crossa, X. Lacaze, P. Du Cheyron, J. Daucourt, J. Lorgeou, F. Piraux, L. Guerreiro, P. Pérez, M. Calus, J. Burgueño, and G. de los Campos. 2014. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* 127(3): 595–607.
- Lecomte, C. 2008. L'évaluation expérimentale des innovations variétales Proposition d'outils d'analyse de l'interaction génotype-milieu adaptés à la diversité des besoins et des contraintes des acteurs de la filière semences. Available at <http://prodinra.inra.fr/record/244938>.
- Leinonen, I., O.M. Grant, C.P.P. Tagliavia, M.M. Chaves, and H.G. Jones. 2006. Estimating stomatal conductance with thermal imagery. *Plant, Cell Environ.* 29(8): 1508–1518.
- Li, X.P., X.Q. Zhao, X. He, G.Y. Zhao, B. Li, D.C. Liu, A.M. Zhang, X.Y. Zhang, Y.P. Tong, and Z.S. Li. 2011. Haplotype analysis of the genes encoding glutamine synthetase plastic isoforms and their association with nitrogen-use- and yield-related traits in bread wheat. *New Phytol.* 189(2): 449–458.
- Liu, Y., C. Subhash, J. Yan, C. Song, J. Zhao, and J. Li. 2011. Maize leaf temperature responses to drought: Thermal imaging and quantitative trait loci (QTL) mapping. *Environ. Exp. Bot.* 71(2): 158–165 Available at <http://dx.doi.org/10.1016/j.envexpbot.2010.11.010>.
- Löffler, C.M., J. Wei, T. Fast, J. Gogerty, S. Langton, M. Bergman, B. Merrill, and M. Cooper. 2005. Classification of Maize Environments Using Crop Simulation and Geographic Information Systems. *Crop Sci.* 45(5): 1708 Available at <https://www.crops.org/publications/cs/abstracts/45/5/1708>.

- Lopes, M.S., G.J. Rebetzke, and M. Reynolds. 2014. Integration of phenotyping and genetic platforms for a better understanding of wheat performance under drought. *J. Exp. Bot.* 65(21): 6167–77 Available at <http://www.ncbi.nlm.nih.gov/pubmed/25246446>.
- Mackay, I., E. Ober, and J. Hickey. 2015. GplusE: beyond genomic selection. *Food Energy Secur.* 4(1): 25–35 Available at <http://doi.wiley.com/10.1002/fes3.52>.
- Martin, J.G. a, D.H. Nussey, A.J. Wilson, and D. Réale. 2011. Measuring individual differences in reaction norms in field and experimental studies: A power analysis of random regression models. *Methods Ecol. Evol.* 2(4): 362–374.
- Molinier, J., G. Ries, C. Zipfel, and B. Hohn. 2006. Transgeneration memory of stress in plants. *Nature* 442(7106): 1046–1049.
- Moller, M., V. Alchanatis, Y. Cohen, M. Meron, J. Tsipris, a Naor, V. Ostrovsky, M. Sprintsin, and S. Cohen. 2006. Use of thermal and visible imagery for estimating crop water status of irrigated grapevine. *J. Exp. Bot.* 58(4): 827–838 Available at <http://jxb.oxfordjournals.org/lookup/doi/10.1093/jxb/erl115>.
- Paux, E., S. Faure, F. Choulet, D. Roger, V. Gauthier, J.P. Martinant, P. Sourdille, F. Balfourier, M.C. Le Paslier, A. Chauveau, M. Cakir, B. Gandon, and C. Feuillet. 2010. Insertion site-based polymorphism markers open new perspectives for genome saturation and marker-assisted selection in wheat. *Plant Biotechnol. J.* 8(2): 196–210.
- Saxena, R.K., D. Edwards, and R.K. Varshney. 2014. Structural variations in plant genomes. *Brief. Funct. Genomics* 13(4): 296–307 Available at <http://www.ncbi.nlm.nih.gov/pubmed/24907366>.
- Springer, N.M., K. Ying, Y. Fu, T. Ji, C.T. Yeh, Y. Jia, W. Wu, T. Richmond, J. Kitzman, H. Rosenbaum, a. L. Iniguez, W.B. Barbazuk, J. a. Jeddelloh, D. Nettleton, and P.S. Schnable. 2009. Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genet.* 5(11).
- Svensgaard, J., T. Roitsch, and S. Christensen. 2014. Development of a Mobile Multispectral Imaging Platform for Precise Field Phenotyping. *Agronomy* 4(3): 322–336 Available at <http://www.mdpi.com/2073-4395/4/3/322/htm>.
- Technow, F., C.D. Messina, L.R. Totir, and M. Cooper. 2015. Integrating Crop Growth Models with Whole Genome Prediction through Approximate Bayesian Computation (I De Smet, Ed.). *PLoS One* 10(6): e0130855 Available at <http://dx.doi.org/10.1371/journal.pone.0130855>.
- Tiezzi, F., K.L.P. Gaddis, J.S. Clay, and C. Maltecca. 2015. Accounting for Genotype by Environment Interaction in Genomic Predictions for US Holstein Dairy Cattle. (49).
- VanRaden, P.M., C.P. Van Tassell, G.R. Wiggans, T.S. Sonstegard, R.D. Schnabel, J.F. Taylor, and F.S. Schenkel. 2009. Invited review: reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* 92(1): 16–24 Available at <http://www.ncbi.nlm.nih.gov/pubmed/19109259>.
- Vigneau, N., M. Ecartot, G. Rabatel, and P. Roumet. 2011. Potential of field hyperspectral imaging as a non destructive method to assess leaf nitrogen content in Wheat. *F. Crop. Res.* 122(1): 25–31 Available at <http://linkinghub.elsevier.com/retrieve/pii/S0378429011000451>.

- Wang, S., D. Wong, K. Forrest, A. Allen, S. Chao, B.E. Huang, M. Maccaferri, S. Salvi, S.G. Milner, L. Cattivelli, A.M. Mastrangelo, A. Whan, S. Stephen, G. Barker, R. Wieseke, J. Plieske, M. Lillemo, D. Mather, R. Appels, R. Dolferus, G. Brown-Guedira, A. Korol, A.R. Akhunova, C. Feuillet, J. Salse, M. Morgante, C. Pozniak, M.-C. Luo, J. Dvorak, M. Morell, J. Dubcovsky, M. Ganai, R. Tuberosa, C. Lawley, I. Mikoulitch, C. Cavanagh, K.J. Edwards, M. Hayden, and E. Akhunov. 2014. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnol. J.* 12(6): 787–796 Available at <http://www.ncbi.nlm.nih.gov/pubmed/24646323>.
- Yin, X., P. Stam, M.J. Kropff, and A.H.C.M. Schapendonk. 2003. Crop modeling, QTL mapping, and their complementary role in plant breeding. *Agron. J.* 95(1): 90–98.
- Yu, J., J.B. Holland, M.D. McMullen, and E.S. Buckler. 2008. Genetic design and statistical power of nested association mapping in maize. *Genetics* 178(1): 539–51 Available at http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2206100&tool=pmcentrez&render_type=abstract.
- Zarco-Tejada, P.J., J.R. Miller, G.H. Mohammed, T.L. Noland, and P.H. Sampson. 2002. Vegetation stress detection through chlorophyll a + b estimation and fluorescence effects on hyperspectral imagery. *J. Environ. Qual.* 31(5): 1433–1441.