



HAL
open science

Repenser les méthodes associées au choix des dates de semis du blé tendre

Marion Gauthier

► **To cite this version:**

Marion Gauthier. Repenser les méthodes associées au choix des dates de semis du blé tendre. Sciences agricoles. 2012. dumas-00797924

HAL Id: dumas-00797924

<https://dumas.ccsd.cnrs.fr/dumas-00797924>

Submitted on 7 Mar 2013

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.



AGROCAMPUS OUEST
CFR Rennes
65 rue de Saint-Brieuc CS 84215
35042 RENNES Cedex



ARVALIS - Institut du végétal
IBP – Université Paris Sud
Rue de Noetzlin – Bât. 630
91405 ORSAY Cedex

Mémoire de Fin d'Études

**Diplôme d'Ingénieur de l'Institut Supérieur des Sciences
Agronomiques, Agroalimentaires, Horticoles et du Paysage ***

Année universitaire : 2011-2012

Spécialisation : Statistique appliquée

Repenser les méthodes associées au choix des dates de semis du blé tendre

Marion GAUTHIER

Volet à renseigner par l'enseignant responsable de l'option/spécialisation

Date :/.../... Signature :

Bon pour dépôt (version définitive)

Autorisation de diffusion : Oui Non

Devant le jury :

Soutenu à Rennes le 11 septembre 2012

Sous la présidence de : François HUSSON

Maître de stage : David GOUACHE et Jean-Charles DESWARTE

Enseignant référent : François HUSSON

Autre membre du jury : Jérôme PAGES

*"Les analyses et les conclusions de ce travail d'étudiant n'engagent
que la responsabilité de son auteur et non celle d'AGROCAMPUS OUEST".*

Fiche de diffusion du mémoire

A remplir par l'auteur⁽¹⁾ avec le maître de stage.

Aucune confidentialité ne sera prise en compte si la durée n'en est pas précisée.

Préciser les limites de la confidentialité⁽²⁾ :

➤ Confidentialité absolue : oui non

(ni consultation, ni prêt)

☞ Si oui 1 an 5 ans 10 ans

➤ A l'issue de la période de confidentialité **ou** si le mémoire n'est pas confidentiel,

Merci de renseigner les éléments suivants :

Référence bibliographique diffusable⁽³⁾ : oui non

Résumé diffusable : oui non

Mémoire consultable sur place : oui non

Reproduction autorisée du mémoire : oui non

Prêt autorisé du mémoire : oui non

.....
Diffusion de la version numérique du résumé : oui non

☞ Si oui, l'auteur⁽¹⁾ complète l'autorisation suivante :

Je soussigné(e) _____, propriétaire des droits de reproduction dudit résumé, autorise toutes les sources bibliographiques à le signaler et le publier.

Date :

Signature :

Rennes/Angers, le

Le maître de stage⁽⁴⁾,

L'auteur⁽¹⁾,

L'enseignant référent,

(1) auteur = étudiant qui réalise son mémoire de fin d'études

(2) L'administration, les enseignants et les différents services de documentation d'AGROCAMPUS OUEST s'engagent à respecter cette confidentialité.

(3) La référence bibliographique (= Nom de l'auteur, titre du mémoire, année de soutenance, diplôme, spécialité et spécialisation/Option)) sera signalée dans les bases de données documentaires sans le résumé.

(4) Signature et cachet de l'organisme.

Cette fiche de diffusion doit être située après la page de couverture du mémoire

Remerciements

Je tiens à remercier en premier lieu mes deux maitres de stage pour m'avoir permis de réaliser ce stage à Arvalis et de m'avoir fait confiance dans son avancée. Je les remercie particulièrement de m'avoir aiguillé le long des périples et tournants qu'auront suscités ces dates de semis ainsi que pour leur soutien au cours de la seconde manche.

Je souhaite, de plus, remercier toutes les personnes d'Arvalis qui m'ont guidée. Je remercie particulièrement Guillaume R., Fatma, Lygie et Cécile avec qui nous avons eu de longues discussions parfois fructueuses en termes d'idées mais toujours en termes de convivialité.

Je souhaite remercier, de plus, Guillaume B. pour avoir su jouer les éclaireurs, et Fanch qui, je l'espère, m'aura montré que l'auto persuasion est un très bon outil.

Je remercie infiniment tous les membres de *l'open space* (et assimilés amis ou non des canards) pour leur accueil chaleureux, les moments de rigolades, les pauses culturelles du midi et celles (un peu moins culturelles) de l'après-midi.

Enfin, un grand "*Merci!*" à mes amis pour m'avoir épaulée.

Sommaire

Introduction générale	1
1 Phénologie du blé tendre et enjeux des dates de semis	1
2 Stratégies d'analyse envisageables pour le choix des dates de semis	2
3 Approche historique chez ARVALIS Institut du végétal	4
4 Les limites de l'approche historique	6
4.1 Limites liées à l'utilisation actuelle	6
4.2 Limites liées à l'analyse fréquentielle	6
5 Problématiques	6
6 Objectifs	7
7 Déroulement	7
1 Diagnostic de la méthode d'analyse fréquentielle existante	9
1.1 Introduction	9
1.2 Méthodes d'analyse	9
1.2.1 Réalisation de simulations	9
1.2.2 Critère d'évaluation des résultats	10
1.3 Résultats	11
1.3.1 Biais	11
1.3.2 Erreur moyenne	11
1.3.3 Amélioration de la méthode	11
1.4 Commentaires des résultats	15
1.5 Bilan	15
2 Mise en œuvre de nouvelles méthodes d'estimation des niveaux de risque de critères phénoclimatiques	17
2.1 Introduction	17
2.2 Structure du jeu de données	18
2.3 Nouvelles méthodes et évaluation de leur performance	21
2.3.1 Méthodes d'estimation des déciles tenant compte des tendances	21

2.3.2	Évaluation de ces méthodes	22
2.4	Résultats	23
2.4.1	Analyse des valeurs de précision	24
2.4.2	Étude globale de la précision	24
2.4.3	Étude globale de la stabilité	25
2.5	Commentaires des résultats	28
2.5.1	Compromis précision/stabilité	28
2.5.2	Les limites de la méthode <i>lm</i>	28
2.5.3	Les avantages de la méthode <i>rq</i>	28
2.6	Bilan	28
3	Préconiser des dates de semis à partir des estimations de critères phénoclimatiques d'intérêt	31
3.1	Introduction	31
3.2	Données utilisées	31
3.3	Méthodes de choix des dates de semis à préconiser	32
3.3.1	Définition de fenêtres possibles par station	32
3.3.2	Optimisation du choix grâce à l'ensemble des critères phénoclimatiques	32
3.4	Résultats de préconisations	33
3.4.1	Quelles dates de semis optimales?	33
3.4.2	Quelle optimisation?	33
3.5	Commentaire des résultats	35
3.5.1	Le choix des limites de la plage de semis	35
3.5.2	Lien avec les préconisations historiques	39
3.6	Bilan	39
4	Discussion générale	41
4.1	Discussion agronomique des résultats de préconisations de dates de semis	41
4.1.1	Quelle validité des préconisations antérieures?	41
4.1.2	Le choix des variables phénoclimatiques, de leur seuil, des limites acceptables	42
4.2	Discussion générale sur le matériel utilisé et les méthodes mises en œuvre	42
4.2.1	Validité des variables phénoclimatiques estimées par le modèle de stades d'Arvalis	42
4.2.2	Stratégie d'analyse	43
	Conclusion générale et perspectives	45
	Bibliographie	48
	Annexes	49

A	Description du jeu de données	49
B	Quelques algorithmes de calculs de quantiles d'un échantillon	52
C	Table de la loi normale centrée réduite $\mathcal{N}(0; 1)$	53
D	Principe de la régression quantile linéaire	54
E	Distribution et variance des $D8_{i,n}$	55
F	Test de Student de comparaison des moyennes $D\hat{8}_n$ aux $D8_{th}$	57
G	Analyse du modèle du biais $\Delta D8$	58
H	Dates de semis préconisées historiquement	60

Liste des figures

1	Les principaux stades de développement du blé tendre	2
2	Méthodes envisageables dans la détermination des dates de semis conduisant au rendement optimal	3
3	Schéma résumé des modèles existants, de l'acquisition des données et des objectifs	5
4	Schéma résumé du modèle existant diagnostiqué au chapitre 1.	8
5	Biais $\Delta D8$ entre $D\hat{8}_n$ et $D8_{th}$ suivant la variance de la loi Normale dont est issu l'échantillon et suivant la taille de cet échantillon	11
6	Erreur moyenne réalisée lorsqu'on estime un décile 8 à partir d'un petit échantillon suivant la variance de la loi Normale dont est issu l'échantillon et suivant la taille de cet échantillon	12
7	$\Delta D8$ et RMSEP des $D8_{i,n}$ corrigé du biais	13
8	$\Delta D8$ et RMSEP des $D8_{i,n}$ calculés à l'aide de la méthode de type 8	14
9	RMSEP des $D8_{i,n}$ calculés à l'aide de la méthode de type 8 comparé à la RMSEP des $D8_{i,n}$ corrigés du biais	14
10	Schéma résumé de la méthode que l'on cherche à déterminer, au chapitre 2, afin de proposer une méthode alternative à l'analyse fréquentielle couramment utilisée au sein d'Arvalis.	16
11	Évolution de la somme de l'évapotranspiration réelle cumulée entre le stade épi 1 cm et l'épiaison pour la station 3002	18
12	Schéma du modèle de stades d'Arvalis permettant l'estimation de variables phénoclimatiques	19
13	Exemple des estimations du décile 2 de la variable <code>Epiaison.Epi1cm.sETR</code>	26
14	Précision et stabilité moyenne pour chaque méthode suivant le nombre d'années antérieures utilisées	27
15	Exemple des estimations du décile 8 de la variable <code>Epiaison.Epi1cm.nTminiinf4</code> illustrant les limites de la méthode <i>lm</i>	29
16	Schéma résumé de la méthode de choix des dates de semis que l'on cherche à déterminer au chapitre 3.	30
17	Structure des données permettant de calculer le coût par date de semis pour chaque station.	33

18	Dates de semis permettant d'éviter les risques destructeurs de gel par station à partir des estimation réalisées pour 2012 avec le couple méthode rq x 20 années antérieures . . .	34
19	Coût associé aux dates de semis sans gel destructeurs	34
20	Graphiques de individus et des variables de l'ACP décrivant la structure des dates de semis suivant les critères phénoclimatiques	36
21	Coût associé aux dates de semis sans gel destructeurs	38
22	Évolution de la variable Maturite.Epiaison.sRg suivant les années et les dates de semis pour la station 201.	44
23	Répartition géographique des 14 stations du jeu de données	49
24	Distribution des valeurs prises par les variables phénoclimatiques non binaires.	50
24	Distribution des valeurs prises par les variables phénoclimatiques non binaires (suite).	51
25	Écart-type des $D8_{i,n}$ suivant la taille de l'échantillon et la variance de la loi Normale génératrice.	55
26	Distribution des $D8_{i,n}$ calculés sur des échantillons de différentes tailles et issus de loi Normale de différentes variances.	56
27	Ajustement du modèle du biais $\Delta D8$	58
28	Vérification de l'hypothèse de normalité des résidus du modèle du biais $\Delta D8$	59
29	Vérification de l'hypothèse d'homostédasticité des résidus du modèle du biais $\Delta D8$	59
30	Plages de semis préconisées historiquement d'une variété de céréales de précocité type Soissons.	60

Liste des tableaux

1	Méthodes envisageables dans la détermination des dates de semis conduisant au rendement optimal	3
2	Résumé des simulations	10
3	Valeurs de AIC et de RMSEP pour différents modèles expliquant $\Delta D8(x, y)$ ajustés sur les données issues des simulations.	13
4	Résumé des 12 variables phénoclimatiques étudiées	20
5	Pourcentage de précision obtenu	23
6	Table d'analyse de la déviance	25
7	Valeurs prises par les variables phénoclimatiques binaires.	49
8	Table de la loi normale centrée réduite $\mathcal{N}(0; 1)$	53
9	Test de Student de comparaison des moyennes des déciles 8 calculés sur de petits échantillons aux déciles 8 théoriques des distributions correspondantes	57
10	Estimation des paramètres du modèle du biais	58

Glossaire

adventice

plante qui se développe dans une autre culture.

apex

tissu à l'extrémité d'une plante qui permet sa croissance.

échaudage

mauvais remplissage des grains dû à des températures élevées lors de la constitution des réserves des grains (fin de cycle).

épiaison

l'épi gonflé est visible. Il a dépassé la ligule de la dernière feuille.

maturité

stade où le développement du blé est achevé.

méiose

division cellulaire qui conduit aux gamètes sexués.

niveau de risque

Nous désignons par *niveau de risque* d'une variable phénoclimatique, l'estimation du décile (2, 5 ou 8) de cette variable, pour l'année $N + 1$, par la méthode d'analyse fréquentielle d'Arvalis ou une méthode alternative mise en œuvre au chapitre 2.

phénologie

phénomènes biologiques périodiques de développement de la plante en fonction du climat.

physiologie végétale

fonctionnement d'une plante.

pédologie

conditions du sol telles que sa composition et sa réserve en eau.

stade double-ride

stade de développement des céréales où des stries horizontales apparaissent sur le point végétatif. Ce stade marque l'initiation des futurs épis.

stade épi 1cm

stade de développement des céréales où le sommet de l'épi est à 1 cm du plateau de tallage. Ce stade correspond au début d'une phase de croissance importante des plantes et donc à des besoins particuliers.

talle

tige latérale ou secondaire qui se développe après le maître-brin.

variable phénoclimatique

variable climatique qui a lieu à un stade physiologique où la plante y est sensible.

Introduction générale

ARVALIS - Institut du végétal est un organisme français de recherche appliquée en agronomie financé par les agriculteurs. Son objectif principal est d'améliorer la compétitivité des exploitations agricoles. Il a un rôle de conseil auprès des différents acteurs de la filière. Pour cela, l'institut a mis au point de nombreux outils. Il doit notamment déterminer les dates de semis optimales permettant aux agriculteurs d'obtenir les rendements les plus élevés. Dans notre étude, nous nous centrerons sur les dates de semis du blé tendre. Cette culture représente la première production française en volume et est cultivée par 40% des exploitations agricoles françaises (Source AGRESTE - Recensement agricole 2000).

1 Phénologie du blé tendre et enjeux des dates de semis

Le choix des dates de semis est un enjeu majeur de la conduite des cultures. La croissance et le développement de la plante dépendent de la température, de l'eau disponible, de l'ensoleillement, etc. Or, c'est la date de semis qui détermine les conditions climatiques que la plante va rencontrer. Le développement de la plante (vitesse de croissance, risque de destruction et qualité de grains) est donc déterminé en grande partie par sa date de semis. Tout au long d'une campagne, il existe des interactions dynamiques entre le climat et le développement du blé.

Le blé tendre se développe suivant différents stades clairement identifiés par les physiologistes (Figure 1). Après germination, les premières feuilles apparaissent. Ensuite, pour une même graine, des tiges latérales (talles) émergent. L'apex de chaque talle suffisamment développée évolue, pendant l'hiver, en un futur épi qui sort de la gaine formée par la dernière feuille, après élongation des tiges. L'épiaison est alors suivie de la floraison puis du remplissage des grains en amidon et en protéines. Après dessiccation des grains, le blé peut enfin être récolté. Au cours de chacun de ces stades, la plante a besoin de différents éléments pour sa croissance (eau, lumière, azote. . .). Elle est aussi sensible à des stress ou des accidents climatiques distincts (Figure 1). Les physiologistes parlent alors de *variables phénoclimatiques* qui sont des variables *climatiques* définies à un stade *physiologique* où la plante y est sensible. Ce stade est généralement défini par une somme de températures (éventuellement corrigée par des effets de la photopériode et de la vernalisation) nécessaires pour que la plante l'atteigne (Gate et al., 2003). Ainsi, les dates de semis conduisant aux plus hauts rendements sont celles permettant d'éviter les accidents climatiques et de capter les conditions d'eau, de rayonnement, etc. optimales dont la plante a besoin durant des phases de développement précises (Gate, 1995). Les variables phénoclimatiques en question ne font pas l'objet d'un consensus auprès des agronomes. Suivant les sources, leur choix est très fluctuant (Porter and Gawith, 1999).

De plus, le contexte agricole actuel est marqué par le changement climatique. En effet, un plafonnement des rendements dû essentiellement au climat a été constaté depuis 1995 environ alors que les rendements étaient en augmentation constante depuis le début du *XX^{ème}* siècle (Gate et al., 2009; Brisson et al., 2010). On remarque aussi une modification des dates d'apparition des accidents

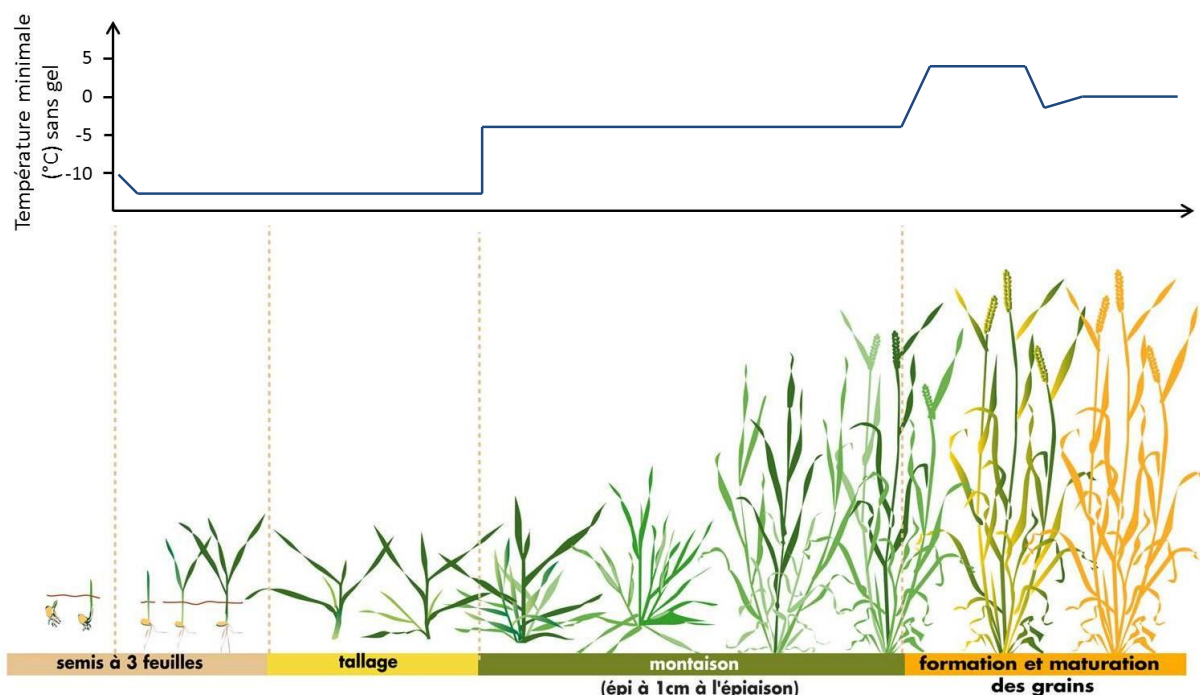


FIGURE 1 – Les principaux stades de développement du blé tendre et UNE courbe de sensibilité au gel suivant les stades (source : ARVALIS Institut du végétal). La courbe présentée ici est proposée par l'institut mais d'autres niveaux de sensibilité sont parfois soumis par d'autres auteurs; ce n'est donc pas une référence absolue.

climatiques, de leur intensité et de leur durée (Olesen et al., 2011). Enfin, on a observé un décalage ou une modification de risques climatiques qui font diminuer le rendement (Gate, 2008). Le choix des dates de semis semble être un levier intéressant pour diminuer les pertes de rendement liées au climat changeant (Porter and Gawith, 1999; Gate, 2007; Gate et al., 2007; Olesen et al., 2011; Gouache et al., 2012).

2 Stratégies d'analyse envisageables pour le choix des dates de semis

Afin de déterminer les dates de semis optimales (qui maximisent le rendement), différentes méthodes sont envisageables pour relier la variable « rendement » au facteur « dates de semis ». Celles-ci sont résumées dans le tableau 1 et à la figure 2.

La méthode la plus naturelle est d'utiliser des données expérimentales pour calculer une valeur de rendement moyen que l'on obtient en semant à différentes dates. Malheureusement, ceci demande un dispositif expérimental très lourd car il faut que plusieurs dates de semis soient testées dans différentes régions pendant au moins 20 ans (nombre d'années nécessaires pour caractériser un climat selon Météo France). L'institut ne possède pas assez de données pour relier les dates de semis et le rendement de cette façon.

Une autre méthode est de prédire le rendement final de la culture, pour plusieurs dates de semis, à partir de données climatiques journalières injectées dans un modèle de culture complet. Il existe, en effet, des modèles visant à simuler l'ensemble des aspects de la croissance du blé tels que l'émission des feuilles, l'élaboration du rendement, l'accumulation en protéines des grains, etc. (Deswarte et al.,

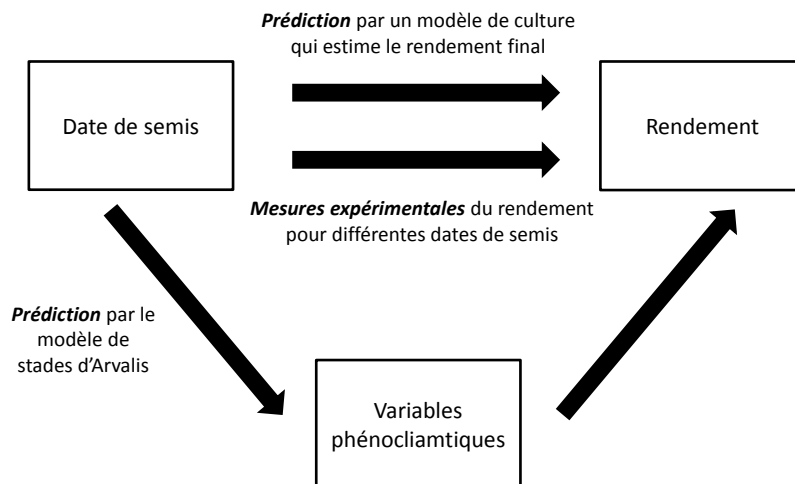


FIGURE 2 – Méthodes envisageables dans la détermination des dates de semis conduisant au rendement optimal

Méthodes	Principes	Avantages	Inconvénients
Expérimentation	Mesurer expérimentalement les rendements réels obtenus pour différents triplets station x année x date de semis	<ul style="list-style-type: none"> – Relation directe entre les dates de semis et le rendement – Pas d'erreur liée à l'utilisation de modèles – Autres facteurs non contrôlables (climat, sol...) 	<ul style="list-style-type: none"> – Dispositif expérimental lourd – Données non disponibles
Prédictions du rendement	A partir des données météo journalières et d'un modèle de culture modélisant son développement, prédire le rendement final pour différentes dates de semis	<ul style="list-style-type: none"> – Relation directe entre les dates de semis et le rendement – Si le modèle est fonctionnel, rapidité d'acquisition des données 	<ul style="list-style-type: none"> – Nécessite de définir toutes les lois de réponse à chaque facteur limitant, et, de pouvoir quantifier correctement ce facteur limitant – Paramétrage très précis du modèle – Prédiction de rendement peu fiable
Recherche indirecte	Méthode historique basée sur l'optimisation de risques ou besoins phénoclimatiques dont on sait qu'ils influencent le rendement	<ul style="list-style-type: none"> – Les variables phénoclimatiques sont issues du modèle de stades d'Arvalis : fiabilité et rapidité d'acquisition des données – Paramétrage simple du modèle de stades d'Arvalis 	<ul style="list-style-type: none"> – Choix des variables phénoclimatiques à prendre en compte – Pas de lien direct ; il faut ensuite coupler les différents variables phénoclimatiques et les hiérarchiser

TABLE 1 – Méthodes envisageables dans la détermination des dates de semis conduisant au rendement optimal

2011). Arvalis a fait le choix de ne pas utiliser ces types de modèles estimant une valeur de rendement final à partir des données climatiques journalières, car ils demandent un paramétrage très précis (Guillaume, 2011). Connaissant le climat quotidien réel d'une année, on ne peut donc pas ici estimer le rendement final puisque nous ne possédons pas les modèles adéquats.

Nous ne sommes donc pas dans la capacité de relier *directement* les dates de semis à un rendement. Arvalis a cependant développé un modèle de stades. Ceux-ci simulent le développement du blé à partir de la connaissance des conditions quotidiennes du climat, pour différentes pédologies et différentes variétés de blé (Partie 1 de la figure 3). Ce modèle donne, en sortie, la date d'apparition des stades, de façon précise. Il permet de positionner dans le temps les principaux stades et, donc, d'estimer des risques phénoclimatiques. C'est pourquoi, Arvalis s'est tourné vers une stratégie *indirecte* pour relier les dates de semis au rendement. Cette stratégie historique est explicitée dans le paragraphe ci-dessous.

3 Approche historique chez ARVALIS Institut du végétal

Historiquement, les dates de semis ont été définies de manière à éviter certains accidents climatiques aux stades de sensibilité correspondants (Masse, 1987; Gate, 1995). Puisque le développement de la plante dépend de façon dynamique du climat, on s'intéresse à des critères phénoclimatiques i.e. des variables climatiques à un stade de développement défini. Par exemple, Rieu et Gate (1997) préconisaient de ne pas semer trop tôt afin d'éviter que des gels n'aient lieu après le stade épi 1cm, ni trop tard pour que le remplissage des grains ne soit pas perturbé par des températures élevées. Le cycle de développement de la plante a donc été calé pour que ces accidents n'aient pas lieu aux stades où la plante y est sensible (Gate, 1995).

Plus précisément, cette démarche historique se décompose en plusieurs étapes. Arvalis possède les données de climat journalier pour de nombreuses stations¹, enregistrées depuis l'année 1956. En intégrant ces données climatiques au modèle de stades d'Arvalis, nous pouvons déterminer les valeurs de variables phénoclimatiques qui déterminent le rendement, tels que le nombre de jours de gels entre deux stades (Partie 1 de la figure 3).

Ces variables phénoclimatiques prennent des valeurs parfois très changeantes d'une année à l'autre. Ainsi, plus que des valeurs de moyenne, il est aussi intéressant de connaître la distribution de ces variables. On souhaite par exemple, estimer le nombre maximum de jours échaudants² que l'on peut rencontrer huit années sur dix, suivant la date de semis. Les écophysiolgistes d'Arvalis calculent alors les déciles 2 ou 8 (second ou huitième quantile au dixième) et médianes des valeurs de la variable phénoclimatique étudiée pour les 20 années précédentes. Ces valeurs sont appelées *niveaux de risque* pour l'année $N + 1$. Les estimations de déciles permettent d'approcher la distribution de la variable. On obtient, par exemple, un nombre maximum de jours échaudants pour 8 années sur 10 (Partie 2 de la figure 3). Ceci permet de savoir quel niveau de stress la culture va devoir supporter a minima 4 années sur 5. De même, l'étude du décile 2 du nombre de jours de gel conduit à évaluer le risque de voir la culture détruite 1 année sur 5. Le recourt aux déciles 2 et 8 permet de nous affranchir des années particulièrement extrêmes du point de vue du climat rencontré. Cette méthode est appelée *analyse fréquentielle* (Partie 2 de la figure 3).

Enfin, la dernière étape consiste à définir les dates de semis à préconiser, à partir des niveaux de risque obtenus (Partie 3 de la figure 3). Les agronomes d'Arvalis ont historiquement défini des seuils d'acceptabilité correspondants aux deux variables phénoclimatiques étudiées, à savoir le gel d'épi (des

1. station expérimentale soit un lieu géographique donné, avec ses conditions climatiques (données de Météo France)

2. mauvais remplissage des grains dû à des températures élevées lors de la constitution des réserves des grains (fin de cycle)

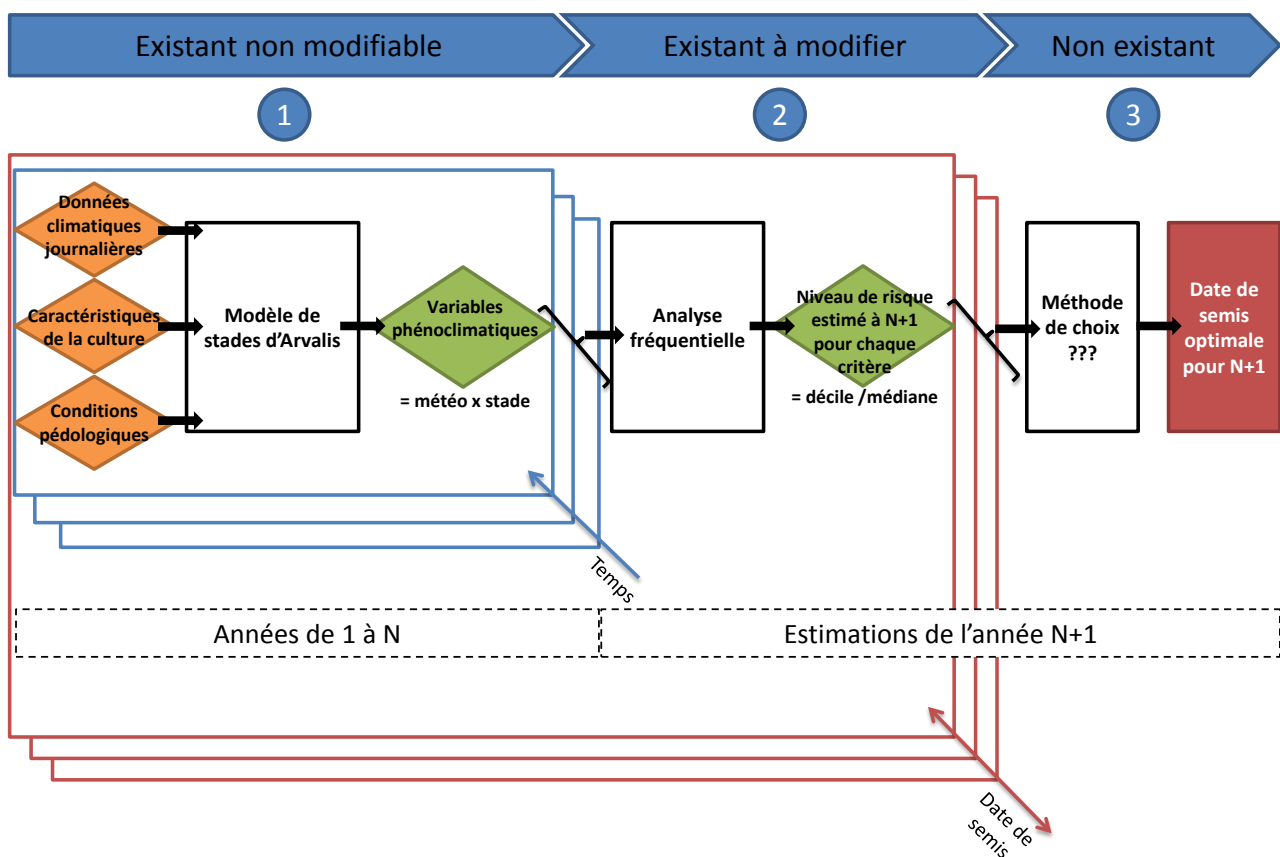


FIGURE 3 – Schéma résumé des modèles existants, de l'acquisition des données et des objectifs. Les boîtes oranges représentent les données d'entrée (mesures ou paramètres) et les vertes les données d'entrée/sortie des modèles qui sont elles-mêmes estimées.

températures inférieures à -4 °C entre le stade épi 1cm et épiaison), et l'échaudage (défini par des températures supérieures à 25 °C entre l'épiaison et la maturité). Les dates de semis optimales, pour une station donnée, étaient celles dont les niveaux de risque pour les deux critères étaient inférieurs aux seuils.

4 Les limites de l'approche historique

4.1 Limites liées à l'utilisation actuelle

Actuellement, les agronomes d'Arvalis souhaitent prendre en compte plus de variables phéno-climatiques. Ils ne connaissent pourtant pas, pour ces nouvelles variables, de seuils d'acceptabilité correspondants. On cherche alors une méthode (Partie 3 de la figure 3) qui optimise chaque critère phéno-climatique à l'année $N + 1$ (décile de la variable phéno-climatique) pour définir les dates de semis à préconiser l'année $N + 1$.

4.2 Limites liées à l'analyse fréquentielle

L'institut étudie les niveaux de risque de variables phéno-climatiques i.e. les déciles ou médianes de variables phéno-climatiques que l'on estime, pour une année future $N + 1$, à partir des valeurs de ces variables pour les années précédentes. Pour cela, Arvalis utilise la méthode d'analyse fréquentielle. Cette méthode est utilisée très généralement au sein d'Arvalis pour traiter diverses problématiques. L'institut travaille notamment sur des variables phéno-climatiques afin d'étudier la faisabilité des chantiers agricoles (semis, récolte), le risque d'apparition de maladies ou la variabilité des plages d'intervention au champ (apports azotés par exemple). L'analyse fréquentielle a ainsi été intégrée dans les outils internes de l'institut. Nous avons cependant identifié plusieurs lacunes relatives à cette méthode.

Premièrement, les niveaux de risque, qui correspondent à des valeurs extrêmes (quantiles extrêmes), sont calculés à partir d'un petit échantillon de 20 valeurs. Nous verrons en quoi les valeurs obtenues sont peu fiables.

Un deuxième problème lié à cette méthode de prédiction des niveaux de risque vient de la nature des données en entrée. Les données, utilisées pour estimer l'année $N + 1$, correspondent à des données antérieures à $N + 1$. Or, nous sommes actuellement dans un contexte de climat changeant. En effet, on observe une évolution de critères agronomiques et climatiques. Il semble donc peu judicieux d'estimer les niveaux de risque de l'année $N + 1$ en attribuant une valeur de décile calculée sur des années passées. Si une variable phéno-climatique possède une tendance, alors celle-ci n'est pas prise en compte par la méthode d'analyse fréquentielle.

5 Problématiques

Avant de s'intéresser au problème du choix des dates de semis, on va regarder les limites liées à l'analyse fréquentielle qui est une méthode générale largement utilisée à Arvalis.

Ensuite, nous travaillerons sur la méthodologie particulière de choix des dates. Il n'existe pas, en effet, actuellement chez Arvalis de méthode permettant de déterminer les dates de semis optimales à partir de critères phéno-climatiques sans seuil d'acceptabilité clairement définis.

L'étude de variables phéno-climatiques étant courante au sein d'Arvalis, ce travail sur le choix des dates de semis est un prototype méthodologique qui a vocation à être décliné dans d'autres approches

de l'institut.

6 Objectifs

Les objectifs sont de :

- Établir un diagnostic de la méthode d'analyse fréquentielle qui est très largement utilisée par l'institut.
- Tester une méthode alternative d'estimation du niveau de risque de variables phénoclimatiques qui prenne en compte des tendances potentielles liées au climat changeant (Partie 2 de la figure 3). Les enjeux de cet objectif sont forts et dépassent le cadre des dates de semis puisque la méthode actuelle est mise en œuvre par Arvalis pour des études variées.
- Mettre en place une nouvelle méthodologie qui permette de définir les dates de semis optimales correspondantes à des estimations de niveaux de risque des variables phénoclimatiques (Partie 3 de la figure 3)
- Plus généralement, reformaliser des règles de décisions quant au choix des dates de semis.

Ce mémoire a donc, avant tout, une vocation méthodologique qui ne s'inscrit pas uniquement dans le choix des dates de semis. La finalité générale de ce mémoire est de "Repenser les méthodes associées au choix des dates de semis".

7 Déroulement

Dans un premier temps, nous étudierons la méthode de calculs de quantiles extrêmes utilisée de façon générale par Arvalis qui est dite d'*analyse fréquentielle*. Nous suivrons ensuite les étapes 2 et 3 de la figure 3. Nous mettrons ainsi en œuvre d'autres méthodes d'estimation des niveaux de risque et des critères permettant d'évaluer leur performance. Enfin, nous définirons une méthodologie permettant de redéfinir les dates de semis optimales correspondant aux nouvelles estimations. Ces trois parties seront traitées de façon indépendante.

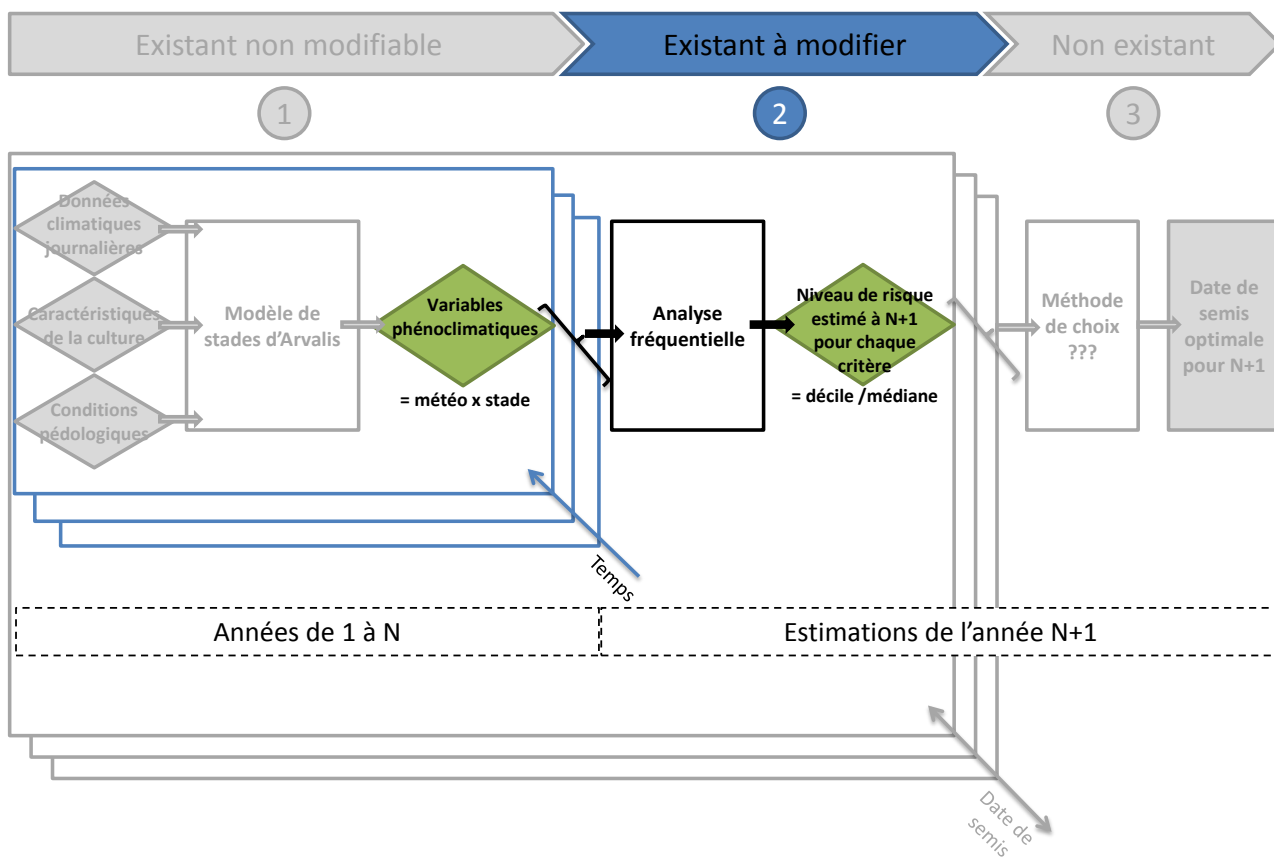


FIGURE 4 – Schéma résumé du modèle existant diagnostiqué au chapitre 1. Les boîtes vertes représentent les données d'entrée et de sortie de ce modèle. Les données d'entrée sont des variables phénoclimatiques estimées par le modèle de culture d'Arvalis pour plusieurs années passées.

Chapitre 1

Diagnostic de la méthode d'analyse fréquentielle existante

1.1 Introduction

Les préconisations de dates de semis passent par l'optimisation¹ des critères phénoclimatiques retenus comme importants dans la composition du rendement final. Plus précisément, on s'intéresse aux *niveaux de risque* de ces variables pour l'année $N + 1$. On définit le niveau de risque d'une variable phénoclimatique comme la valeur maximale de la variable qui est observée un certain nombre de fois. Par exemple, le niveau de risque à l'année $N + 1$ du nombre de jours échaudants correspond au nombre de jours échaudants maximum que l'on rencontre 8 années sur 10. Ceci revient donc à estimer la distribution de la variable. Afin d'approcher cette distribution, Arvalis calcule les médiane et déciles² (2 et 8) à partir des données de cette variable pour les 20 années antérieures.

Outre le choix des dates de semis, cette méthode d'analyse fréquentielle est utilisée de façon générale par l'institut pour répondre à de nombreuses questions telles que connaître la date d'apparition d'un stade pour pouvoir effectuer des mesures, le risque d'apparition d'une maladie dans le but d'anticiper un traitement... Cette méthode est même intégrée à des outils internes. Cette méthode d'analyse fréquentielle a donc un enjeu particulièrement important pour l'institut. Nous cherchons à évaluer la pertinence de cette méthode au vu de la taille des échantillons sur lesquels sont réalisés les calculs (20 observations passées).

Dans un premier temps, nous décrirons les simulations de données réalisées dans le but d'évaluer de façon théorique la méthode d'analyse fréquentielle. Ensuite, nous présenterons les différents critères d'évaluation exploités ainsi que nos tentatives d'amélioration. Enfin, nous commenterons nos résultats dans le cas des variables phénoclimatiques intervenant dans le choix des dates de semis.

1.2 Méthodes d'analyse

1.2.1 Réalisation de simulations

Il est ainsi question de savoir comment varie l'estimation de quantiles d'une variable phénoclimatique si on change d'échantillon. Pour répondre à cette question, on a recours à des simulations d'échantillons qui représentent la variable phénologique (par exemple, date d'apparition d'un stade)

1. éviter les accidents destructeurs, minimiser les risques et maximiser les besoins de la culture
2. quantile au dixième, la distribution de la variable est donc divisée en 10 intervalles de même effectif

ou phénoclimatique (par exemple, nombre de jours où la température est faible entre deux stades) obtenue grâce au modèle de stades d'Arvalis (Figure 4). Ainsi, en prenant le cas d'un décile 8 noté $D8$:

1. Tirage d'un échantillon de n observations
2. Calcul des valeurs de $D8$ pour cet échantillon. L'algorithme utilisé est celui noté 7 dans l'aide de la fonction `quantile()` du logiciel R (Team, 2012) ainsi que par Hyndman et Fan (Hyndman and Fan, 1996), puisque c'est cet algorithme qui est utilisé par Arvalis (via ses outils internes). Il est décrit à l'annexe B. Pour la 1ère simulation, on a ainsi $D8_{1,n}$
3. Itération de l'opération 2000 fois
4. Calcul de la moyenne et de l'écart-type des valeurs obtenues. On obtient ainsi $D8_n$ et σ_n . Puisque le nombre de simulations est très grand, on peut dire que ces valeurs $\hat{D8}_n$ et $\hat{\sigma}_n$ sont les $D8$ et σ théoriques *pour un échantillon de n individus*.

Il est nécessaire de savoir selon quelle loi tirer l'échantillon. On rencontre généralement des variables qui suivent théoriquement :

- des lois Normales. *Ex* : date d'apparition d'un stade, température moyenne à une date.
- des lois de Poisson. *Ex* : nombres de jours où la température est inférieure à un certain seuil entre deux stades.

De plus, la précision des déciles varie peut-être avec la variance de distribution. On effectue donc les simulations précédentes en tirant nos échantillons selon des lois Normales ou de Poisson de différentes variances. Les simulations, résumées dans le tableau 2, sont réalisées à l'aide du logiciel R (Team, 2012). Nous ne détaillerons que les analyses réalisées pour les déciles 8 calculés sur des échantillons issus de lois Normales.

Variable	Notation	Valeurs prises
Taille de l'échantillon	<i>tailleEch</i>	5,10,20,25,30,50,150
Ecart-type de la loi Normale centrée génératrice	<i>sdLoi</i>	1 à 15 avec un pas de 1

TABLE 2 – *Résumé des simulations. On réalise 2000 simulations pour chaque couple taille x écart-type.*

1.2.2 Critère d'évaluation des résultats

Deux critères nous permettent de tester la pertinence de nos estimations :

- le biais noté $\Delta D8$ défini comme la différence entre $\hat{D8}_n$ et $D8_{th}$, décile 8 de la loi considérée,
- la *RMSEP* (Root Mean Square Error of Prediction) soit la racine carrée de la moyenne des carrés des écarts entre le $D8_{i,n}$ qui est calculé sur un échantillon i de taille n , et le $D8_{th}$. Elle correspond à l'erreur moyenne réalisée lorsqu'on estime un décile 8 sur un petit échantillon.

1.3 Résultats

Pour chaque couple taille de l'échantillon x écart-type de la loi Normale génératrice, les 2000 valeurs de décile 8 obtenues suivent une loi Normale dont l'écart-type augmente avec l'écart-type de la loi dont sont issus les échantillons (Annexe E). On note *tailleEch*, la taille de l'échantillon, et *sdLoi*, l'écart-type de la loi selon laquelle sont tirés les échantillons.

1.3.1 Biais

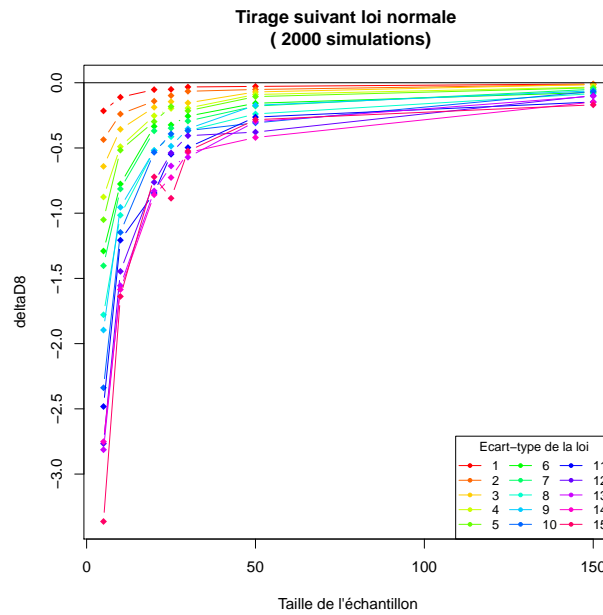


FIGURE 5 – Biais $\Delta D8$ entre $\hat{D8}_n$ et $D8_{th}$ suivant la variance de la loi Normale dont est issu l'échantillon et suivant la taille de cet échantillon

Le biais $\Delta D8$ calculé ici est la différence entre la moyenne des déciles 8 calculés sur des petits échantillons tirés selon des lois Normales de variances différentes notée $\hat{D8}_n$, et le décile 8 de la loi correspondante $D8_{th}$. Des tests de Student de comparaison de moyennes montrent que le décile 8 d'un petit échantillon est inférieur au décile 8 de la distribution dont il est issu (Annexe F). *En moyenne*, il est sous estimé, d'autant plus que la variance de la loi est élevée et que la taille de l'échantillon est faible (Figure 5).

1.3.2 Erreur moyenne

L'erreur moyenne réalisée est d'autant plus élevée que la variance de la loi est élevée et que la taille de l'échantillon est faible (Figure 6). Ceci s'explique par l'augmentation du biais et de la variance des estimations selon ces deux variables (variance de la loi et taille de l'échantillon).

1.3.3 Amélioration de la méthode

Correction du biais

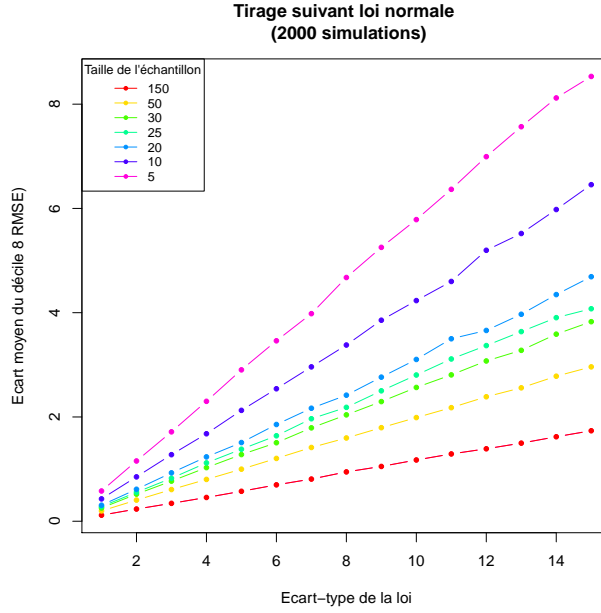


FIGURE 6 – Erreur moyenne réalisée lorsqu'on estime un décile 8 à partir d'un petit échantillon suivant la variance de la loi Normale dont est issu l'échantillon et suivant la taille de cet échantillon

Afin de diminuer la RMSEP, on peut essayer d'enlever la part d'erreurs due au biais. Pour cela, on cherche à modéliser le biais $\Delta D8$ à partir des variables³ quantitatives *tailleEch* et *sdLoi*. On teste classiquement les effets des variables $(tailleEch)^2, (sdLoi)^2, tailleEch, sdLoi$ ainsi que l'interaction *tailleEch* : *sdLoi*. Au vu de la forme des courbes en figure 5 (hyperboles), on introduit de plus le terme $\frac{1}{tailleEch}$ ainsi que la variable *div* qui vaut $\frac{sdLoi}{tailleEch}$. Le choix du modèle se fait selon une sélection pas à pas "avant/arrière" en minimisant à chaque étape le critère AIC⁴. Le modèle retenu, qui ajuste le mieux les données, est :

$$\Delta D8(x, y) = \alpha_0 + \alpha_1 \cdot \frac{x}{y} + \alpha_2 \cdot x + \alpha_3 \cdot x^2 + \alpha_4 \cdot y + \alpha_5 \cdot x \cdot y + \varepsilon$$

avec x = écart-type de la loi Normale génératrice et y = taille de l'échantillon.

D'après l'étude des résidus du modèle, l'hypothèse de leur normalité semble respectée (Annexe G). La variance des résidus en fonction des valeurs prédites indique que le modèle n'est pas optimal. Cependant, il a une très bonne qualité prédictive puisque l'on obtient une RMSEP de 0.00308 après validation croisée (leave-one-out). Le tableau 3, regroupant les valeurs d'AIC et de RMSEP de plusieurs modèles, montre la pertinence de l'ajout du terme $\frac{x}{y}$.

Puisque $\Delta D8 = D\hat{8}_n - D8_{th}$, on retire les valeurs estimées du biais par le modèle aux moyennes des $D8_{n,i}$.

Le premier graphe figure 7 indique que le biais est presque corrigé. La moyenne des $D8_{i,n}$ est donc presque identique au $D8_{th}$. On remarque cependant que la RMSEP n'est que faiblement diminuée par cette correction. Dans le cas d'un échantillon de 20 observations, on ne gagne que très peu de précision. La grande variabilité des $D8_{i,n}$ est donc à l'origine de leur erreur moyenne.

3. pour les notations des variables, voir tableau 2

4. Critère d'Akaike ou Akaike Information Criterion = $-2 \cdot \log(Likhd) + 2 \cdot n$ avec *Likhd* la vraisemblance du modèle et n son nombre de paramètres

Modèle expliquant $\Delta D8(x, y)$	AIC	RMSEP
$\alpha_0 + \alpha_1 \cdot x^2 + \alpha_2 \cdot x + \alpha_3 \cdot y + \alpha_4 \cdot x \cdot y$	167.7706	0.1923976
$\alpha_0 + \alpha_1 \cdot \frac{x}{y} + \alpha_2 \cdot x + \alpha_3 \cdot y + \alpha_4 \cdot x \cdot y$	-400.9312	0.003366179
$\alpha_0 + \alpha_1 \cdot \frac{x}{y} + \alpha_2 \cdot x^2 + \alpha_3 \cdot x + \alpha_4 \cdot y + \alpha_5 \cdot x \cdot y$	-414.1138	0.003080692

TABLE 3 – Valeurs de AIC et de RMSEP pour différents modèles expliquant $\Delta D8(x, y)$ ajustés sur les données issues des simulations. La RMSEP est obtenue après validation croisée (leave-one-out).

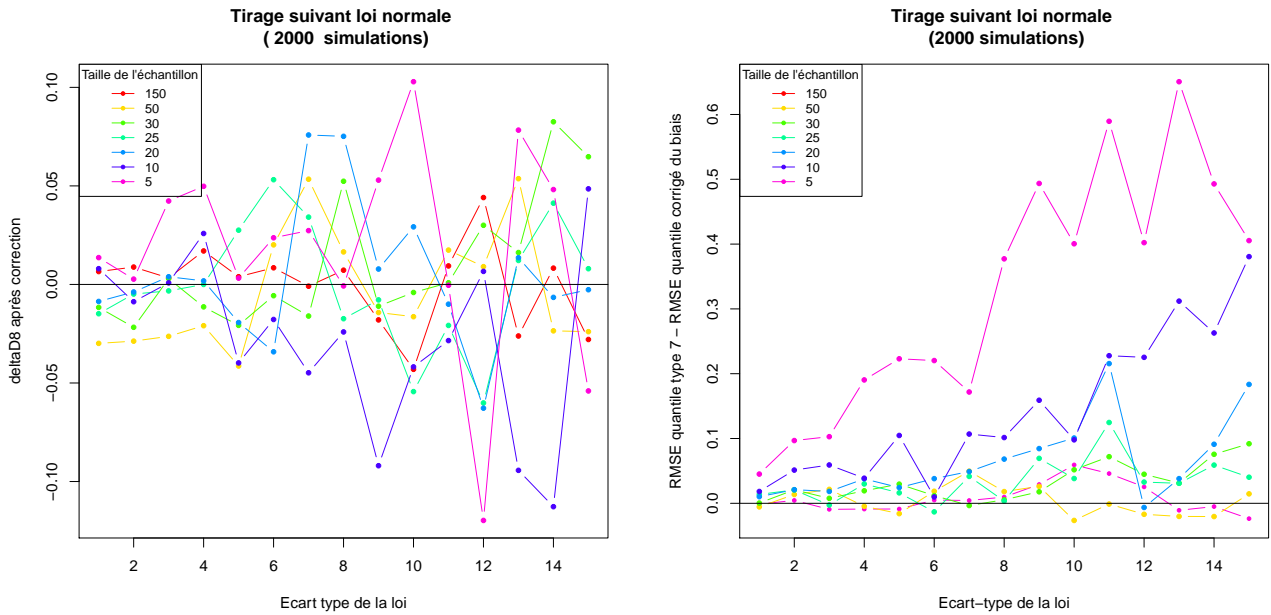


FIGURE 7 – $\Delta D8$ et RMSEP des $D8_{i,n}$ corrigé du biais

Autre méthode de calcul du quantile d'un échantillon

Il existe d'autres méthodes de calcul du quantile d'un échantillon. Notamment, Hyndman et Fan indiquent que la méthode de type 8 est à « médiane non biaisée » quelle que soit la distribution théorique dont est issu l'échantillon (Hyndman and Fan, 1996). On réalise alors les mêmes simulations en estimant les déciles à l'aide de ce mode de calcul. On tire les mêmes conclusions qu'après avoir corrigé le biais précédemment (Figure 8). L'erreur est légèrement plus faible avec la méthode de correction du biais en particulier pour les échantillons de 5 et 10 observations (Figure 9).

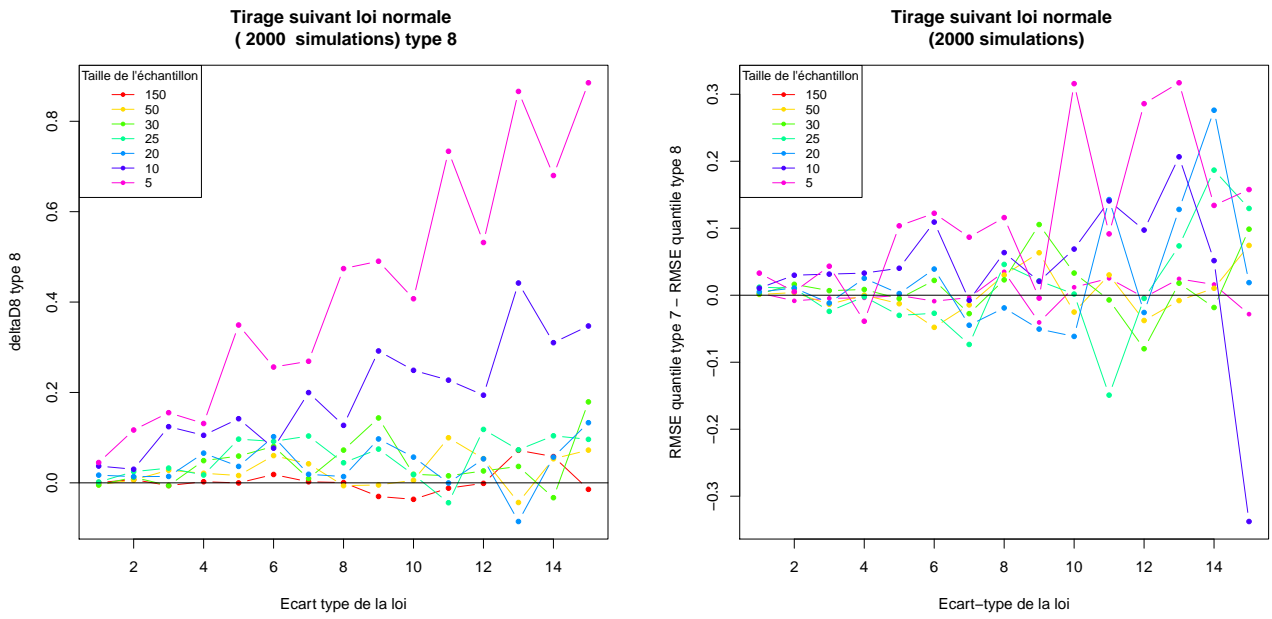


FIGURE 8 – $\Delta D8$ et $RMSEP$ des $D8_{i,n}$ calculés à l'aide de la méthode de type 8

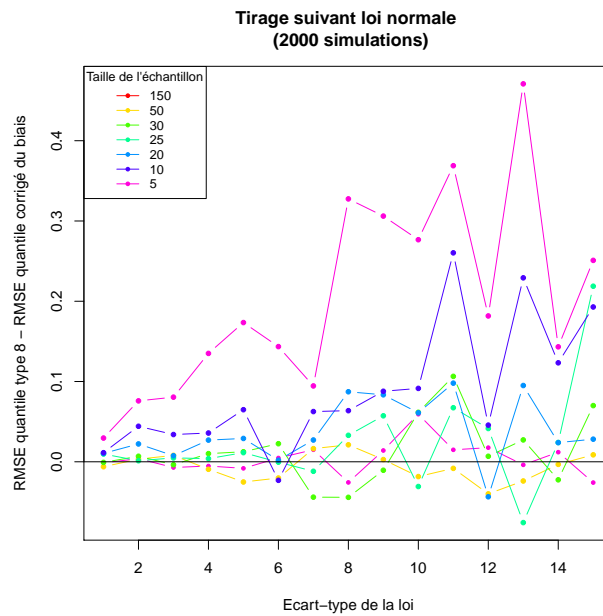


FIGURE 9 – $RMSEP$ des $D8_{i,n}$ calculés à l'aide de la méthode de type 8 comparé à la $RMSEP$ des $D8_{i,n}$ corrigés du biais

1.4 Commentaires des résultats

Actuellement, Arvalis calcule des déciles 8 de variables phéno-climatiques ou d'autres variables telles que la date d'apparition du stade épi 1cm. Les valeurs prises par cette variable suivent une loi Normale. L'écart-type moyen sur 14 stations des données de 1954 à 2011 est de 12,4 jours. Ainsi, avec un échantillon de 20 années, en moyenne, on estime le décile 8 avec 3,7 jours de moins que la théorie. Les mêmes simulations et analyses ont été réalisées pour des échantillons tirés selon des lois de Poisson. L'institut étudie ce type de variable comme par exemple le nombre de jours où la température minimum est inférieure à -4°C entre épi 1cm et épiaison. L'écart-type moyen sur 14 stations des données de 1954 à 2011 est de 0,7 jour. On a alors montré que, avec un échantillon de 20 années, en moyenne, on estime le décile 8 de cette variable phéno-climatique avec un écart de 1 jour à la théorie.

1.5 Bilan

Nous avons mis en évidence que l'estimation de déciles extrêmes à partir de petits échantillons induit un biais par rapport à la valeur théorique attendue. De plus, les valeurs obtenues sont très variables ce qui est à l'origine d'une erreur moyenne d'estimation élevée malgré différentes tentatives de correction du biais. La méthode actuelle consiste à estimer les niveaux de risque de critères phéno-climatiques à l'année $N + 1$ en calculant des déciles sur les 20 années antérieures. Ainsi, si un faible nombre d'années permet mieux de tenir compte d'une tendance éventuelle, il ne permet pas d'estimer correctement les valeurs extrêmes de la distribution des variables phéno-climatiques.

Puisque cette méthode d'analyse fréquentielle est à très fort enjeu pour Arvalis, nous allons donc essayer de mettre en place une méthode alternative.

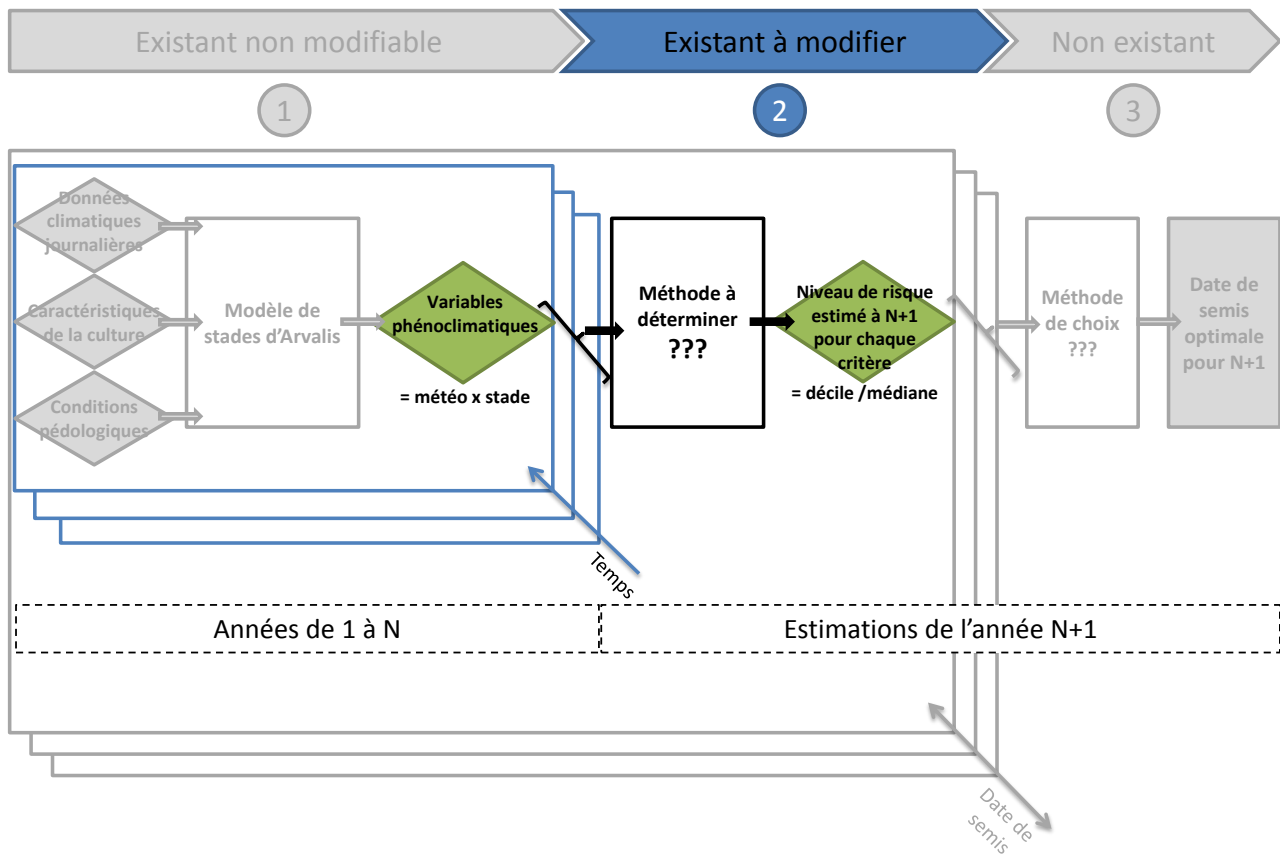


FIGURE 10 – Schéma résumé de la méthode que l'on cherche à déterminer, au chapitre 2, afin de proposer une méthode alternative à l'analyse fréquentielle couramment utilisée au sein d'Arvalis. Les boîtes vertes représentent les données d'entrée et de sortie de ce modèle. Les données d'entrée sont des variables phénoclimatiques estimées par le modèle de culture d'Arvalis pour plusieurs années passées.

Chapitre 2

Mise en œuvre de nouvelles méthodes d'estimation des niveaux de risque de critères phénoclimatiques

2.1 Introduction

Actuellement, Arvalis utilise la méthode d'analyse fréquentielle pour répondre à des questions variées dont *un exemple* est le choix des dates de semis. Cette méthode générale consiste à attribuer comme niveau de risque d'une variable phénoclimatique, pour une année $N + 1$, le décile calculé sur les valeurs de la variable phénoclimatique prises lors des 20 années antérieures à l'année $N + 1$. Cette méthode soulève deux interrogations. D'une part, d'un point de vue technique, nous n'estimons pas correctement des valeurs extrêmes de distribution à partir de petits échantillons (vu au chapitre 1). D'autre part, réalisant des calculs de déciles sur des données passées, nous ne tenons pas compte de potentielles *tendances* des variables pouvant être induites par le changement climatique (Figure 11). Nos objectifs sont donc de :

- Mettre en place une méthode alternative d'estimation du niveau de risque de variables phénoclimatiques qui prend en compte des tendances potentielles liées au climat changeant (Figure 10) et qui soit utilisable par Arvalis
- Estimer sa performance par rapport à la méthode actuelle dans le cas de la problématique du choix des dates de semis

Dans un premier temps, nous décrirons les données qui sont utilisées en entrée des différentes méthodes testées ainsi que la façon dont ces données ont été obtenues. Ensuite, nous présenterons les différentes méthodes testées et les critères mis en œuvre afin d'évaluer leur performance par rapport à la méthode actuelle. Cette performance ne sera expertisée que pour le cas du choix des dates de semis puisque les données d'entrée utilisées ne sont relatives qu'à cette problématique. En effet, l'objectif général que nous nous sommes fixé pour ce mémoire est bien de "Repenser les méthodes associées au choix des dates de semis". Enfin, nous concluons ce chapitre par nos préconisations soit de mise en place d'une méthode alternative à l'analyse fréquentielle, soit de sa conservation.

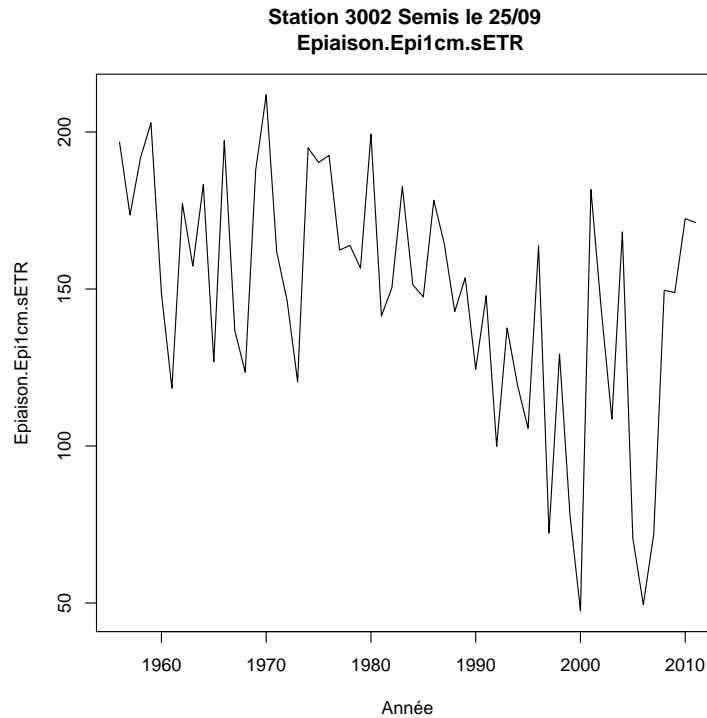


FIGURE 11 – Évolution de la somme de l'évapotranspiration réelle cumulée entre le stade épi 1 cm et l'épiaison pour la station 3002

2.2 Structure du jeu de données

La méthode actuelle d'analyse fréquentielle consiste à attribuer comme niveau de risque d'une variable phénoclimatique, pour une année $N + 1$, le décile calculé sur les valeurs de la variable phénoclimatique prises lors des 20 années antérieures à l'année $N + 1$. Les valeurs prises par une variable phénoclimatique au cours des années antérieures correspondent donc à l'entrée de cette méthode. Nous cherchons une méthode alternative à cette méthode d'analyse fréquentielle, largement utilisée par l'institut. Bien que cette méthode soit générale, nous allons en évaluer la performance dans le cadre de la problématique particulière du choix des dates de semis.

Comme vu précédemment, nous allons mettre en place une méthodologie de choix des dates de semis qui étudie indirectement la relation entre la date de semis et le rendement de la culture. Dans le but de répondre à cette problématique, nous passons par l'estimation des niveaux de risque de critères phénoclimatiques, pour l'année $N + 1$. Les valeurs prises par ces critères phénoclimatiques sont obtenues grâce au modèle de stades d'Arvalis qui permet d'estimer les dates d'apparition des stades de développement du blé mais pas le rendement final de la culture. Nous disposons des données climatiques journalières relatives à 14 stations météorologiques depuis 1956. Ces données sont intégrées au modèle de stades de l'institut. Ce modèle est paramétré pour une culture de blé tendre de variété Soissons qui est la variété utilisée comme référence dans le modèle d'Arvalis, semée à différentes dates. De plus, les conditions pédologiques sont définies identiquement dans toutes les stations. Le modèle permet de simuler les dates d'apparition des stades de développement. En les recoupant aux données climatiques réelles (utilisées en entrée du modèle), on peut donc connaître des variables phénoclimatiques qui témoignent d'un climat à un stade donné (Figure 12). On choisit de simuler, à l'aide de ce modèle, des variables phénoclimatiques connues pour leur influence sur le rendement. Le choix de ces critères s'est fait via une étude préliminaire (cf.infra).

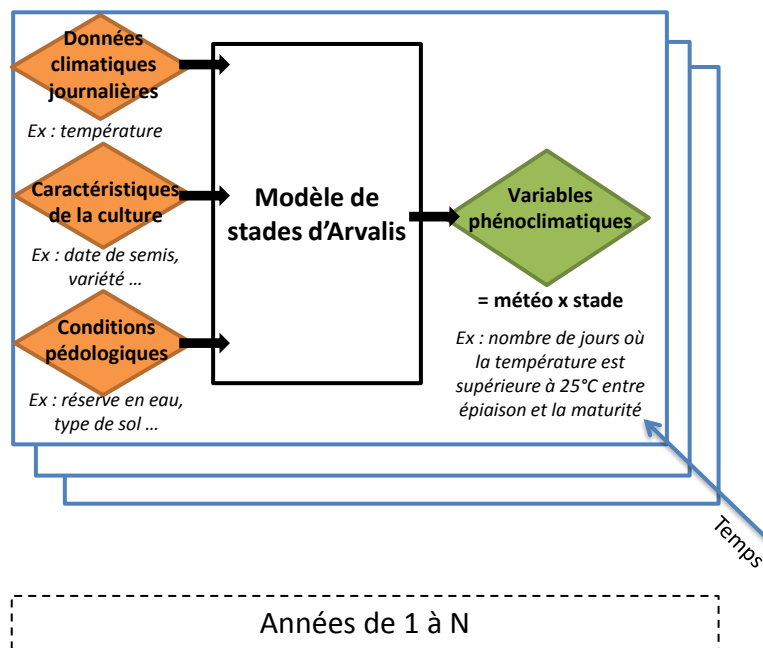


FIGURE 12 – Schéma du modèle de stades d'Arvalis permettant l'estimation de variables phénoclimatiques

Ces simulations sont réalisées en utilisant les données climatiques relatives à tous les couples station x année, et en paramétrant le modèle de stades d'Arvalis avec différentes dates de semis. Les variables *station*, *année* et *date de semis* prennent les modalités :

Station : 14 stations étudiées (Annexe A)

Année : de 1956 à 2011 (données météo quotidiennes disponibles)

Date de semis : 17 modalités du 10 septembre au 30 novembre espacées de 5 jours

En sortie du modèle, on obtient toutes les **variables phénoclimatiques** pour toutes les combinaisons station x année x date de semis. Les variables phénoclimatiques étudiées sont résumées dans le tableau 4 ainsi qu'à l'annexe A.

L'analyse se déroule en deux étapes. Dans un premier temps (correspondant à la partie 2 de la figure 3 page 5), on étudie l'évolution avec les années de chaque variable phénoclimatique afin de déterminer le niveau de risque à l'année $N + 1$. Le niveau de risque d'une variable phénoclimatique est une valeur de décile estimée pour une année. On distingue :

- les besoins à maximiser, dont les niveaux de risque étudiés sont des médianes et des déciles 2 afin de cerner les tendances globales ainsi que les cas minimaux extrêmes
- facteurs de stress « non destructeurs », que l'on cherche à minimiser en étudiant leurs médiane et décile 8 (tendance globale et maxima)
- les risques « de destruction » par le gel, que l'on transforme en présence/absence de gel (codage 1/0) avant de cerner avec un décile 8

Choix des critères phénoclimatiques

Variable phénoclimatique	Notation	Type de critère
Gel hivernal (0/1) : la température est inférieure à $-12/ - 15^{\circ}\text{C}$ entre le semis et le stade épi 1cm	Epi1cm.Semis.nTminiinf12 Epi1cm.Semis.nTminiinf15	Risque destructeur à éviter
Gel d'épi (0/1) : la température est inférieure à $-4/ - 5^{\circ}\text{C}$ entre le stade épi 1cm et l'épiaison	Epiaison.Epi1cm.nTMiniinf4 Epiaison.Epi1cm.nTMiniinf5	Risque destructeur à éviter
Nombre de jours où la température est inférieure à 4°C à $+/- 5$ jours de la méiose	Méiose5j.Méiose5j.nTminiinf4	Stress à minimiser
Nombre de jours où le rayonnement global journalier est inférieur à 200 calories par m^2 à $+/- 5$ jours de la méiose	Méiose5j.Méiose5j.nRginf200	Stress à minimiser
Nombre de jours où la température ne passe pas sous les 15°C entre épiaison et maturité	Epiaison750C.Epiaison.nTminisup15	Stress à minimiser
Nombre de jours où la température est supérieure à 25°C entre l'épiaison et la maturité	Epiaison750C.Epiaison.nTmaxisup25	Stress à minimiser
Quantité d'eau réellement perdue par transpiration entre le stade épi 1cm et épiaison et, entre épiaison et maturité	Epiaison.Epi1cm.sETR Epiaison750C.Epiaison.sETR	Besoin à maximiser
Somme de rayonnement global entre l'épiaison et la maturité	Maturite.Epiaison.sRg	Besoin à maximiser
Somme de rayonnement intercepté au cours du cycle de développement	saPAR	Besoin à maximiser

TABLE 4 – Résumé des variables phénoclimatiques étudiées. Les notations posées ici seront utilisées dans la suite de ce rapport. Les niveaux de risque étudiés pour chaque critère phénoclimatique dépendent du type de critère. Pour les risques destructeurs à éviter, on étudie leur décile 8, dans le cas des stress à minimiser, on s'intéresse à leurs médiane (ou décile 5) et décile 8; enfin, on regarde la médiane et le décile 2 des besoins à maximiser.

Dans un premier temps, nous avons établi la liste des variables phénoclimatiques intervenant dans les préconisations de date de semis d'après la littérature (Masse, 1987; Gate, 1995; Rieu and Gate, 1997). Suivant les sources, les critères énoncés ne sont pas identiques. Il n'existe, en effet, pas de consensus sur les critères à prendre en compte dans le choix des dates de semis. Par exemple, plusieurs seuils de températures létales sont proposés pour un même stade de développement du blé (Porter and Gawith, 1999).

Dans un second temps, nous avons sélectionné les critères énoncés historiquement par l'institut dans des documents internes. Ayant à notre disposition les préconisations correspondantes, nous avons essayé de re-valider ces préconisations historiques en utilisant la méthodologie énoncée (analyse fréquentielle) et les données de l'époque. Contre toute attente, nous avons obtenu des dates possibles ne correspondant pas aux préconisations historiques (Annexe H).

Nous avons donc cherché de nouveaux critères phénoclimatiques intéressants dans le choix des dates de semis puisque d'une part, il n'existe pas de consensus sur ces critères et, d'autres part, les critères énoncés historiquement par l'institut ne semblent pas exacts. Pour cela, un travail préliminaire était la définition même d'une plage de semis optimale. Il est nécessaire de choisir, parmi la littérature, les risques à éviter et les besoins à optimiser au cours du cycle de développement. En concertation avec les experts écophysiologistes, nous avons choisi de retenir certains critères phénoclimatiques à prendre en compte ; il ne s'agit donc pas de critères absolus. Les critères obtenus ont été définis sans limite d'acceptabilité.

2.3 Nouvelles méthodes et évaluation de leur performance

Dans cette partie, nous présentons les méthodes testées permettant d'estimer un niveau de risque d'une variable phénoclimatique soit, un décile estimé pour une année $N + 1$, à partir des années antérieures. L'objectif est de proposer à Arvalis une alternative à la méthode actuelle (analyse fréquentielle) qui prenne en compte des potentielles tendances liées au changement climatique. Le méthode de remplacement doit être à la fois :

- fiable pour obtenir des estimations qui décrivent bien la distribution des variables phénoclimatiques
- stable car il ne serait pas très judicieux pour l'institut de préconiser des dates de semis complètement différentes d'une année à l'autre.

Par ailleurs, nous verrons quels ont été les critères mis en œuvre pour évaluer ces deux points.

Dans la suite de ce rapport, nous noterons AF la méthode d'analyse fréquentielle utilisée actuellement.

2.3.1 Méthodes d'estimation des déciles tenant compte des tendances

Variabilité autour de la régression linéaire

Afin de prendre en compte une possible tendance des variables phénoclimatiques, la manière la plus simple est de réaliser une régression linéaire. Il est alors possible de prolonger la droite obtenue pour estimer la moyenne à l'année $N + 1$. De plus, les résidus du modèle suivent une loi Normale $\mathcal{N}(0, \sigma^2)$. En faisant l'hypothèse que la variance reste la même sur la plage étudiée, l'écart-type des résidus du modèle nous donne donc une estimation de la distribution de la variable phénoclimatique autour de la moyenne estimée. En multipliant par -0.84 ou 0.84 (d'après la table de la loi normale, Annexe C), on obtient une estimation respectivement des déciles 2 et 8 de la distribution de la variable

en $N + 1$. Nous noterons par la suite cette méthode *lm*.

Régression quantile linéaire

La méthode précédente se base sur une régression linéaire. Les paramètres de cette régression sont estimés en minimisant le critère des moindres carrés qui conduit à une meilleure estimation de la moyenne. Or, nous nous intéressons à une meilleure estimation d'un décile. La régression quantile a cette finalité (Koenker and Hallock, 2001). Pour cela, on utilise un critère d'ajustement asymétrique développé dans l'annexe D. Nous estimons la valeur du décile désiré en $N + 1$ en prolongeant la droite de régression quantile linéaire obtenue via le package *quantreg* de R (Koenker, 2012). Nous noterons par la suite cette méthode *rq*.

Combien d'années antérieures prendre en compte ?

Nous réalisons ces estimations de déciles en $N + 1$ suivant les deux méthodes précédentes en utilisant différentes antériorités des données sur lesquelles sont effectuées les régressions. Pour chaque station x date de semis x variable, on estime donc les déciles recherchés en utilisant les 5 à 30 années antérieures soit 26 estimations par triplet.

2.3.2 Évaluation de ces méthodes

Critère d'évaluation de la précision

On recherche une méthode décrivant le mieux la distribution des variables phénoclimatiques. L'estimation des quantiles doit donc être « bien placée » i.e. les observations doivent être bien réparties de part et d'autre du décile.

1. Estimation des valeurs de décile de 1986 à 2011 en utilisant à chaque itération les n années antérieures avec $n \in [3; 30]$.
2. Calcul du nombre d'observations y_i inférieures strictement au décile estimé à l'année i avec $i \in [1986; 2011]$. On le note *nbInf*. De même, on a *nbSup*, le nombre de y_i supérieures strictement au décile estimé à l'année i . Dans le cas d'une médiane par exemple, il y a autant de points inférieurs que de points supérieurs à celle-ci. De manière générale, notre *erreur* est l'écart à la répartition théorique :

$$\left(\frac{nbInf}{nbInf + nbSup} - q \right)^2$$

avec q la probabilité associée au quantile étudié soit 0.2 pour un décile 2, 0.5 pour une médiane et 0.8 pour un décile 8.

Méthodes d'analyse des valeurs de précision

On obtient une erreur d'estimation du décile associée à différentes combinaisons station x date de semis x variable x décile étudié x nombre d'années antérieures (Tableau 5). Les objectifs sont de comprendre dans un premier temps comment varie ce critère puis de déterminer le couple méthode x nombre d'années antérieures optimal (ou les couples optimaux) pour les calculs.

La variable à expliquer est la valeur nbInf. C'est l'agrégation d'une variable qualitative binaire qui prend les modalités :

- 1, l'observation est inférieure au décile estimé (succès),

nbInf	Station	Date de semis	Critère phénoclimatique	Méthode	Nombre d'années passées
0.2207	201	10/09	Epiaison.Epi1cm.sETR.D8	AF	5
0.2309	201	10/09	Epiaison.Epi1cm.sETR.D8	AF	6
0.2278	201	10/09	Epiaison.Epi1cm.sETR.D8	AF	7
:	:	:	:	:	:

TABLE 5 – *Erreur d'estimation obtenue pour différentes stations expérimentales (14), dates de semis (17), niveaux de risque de variables phénoclimatiques (20 associations variables x décile) et calculé suivant 3 méthodes, en utilisant entre 5 et 30 nombre d'années antérieures.*

– 0 l'observation est supérieure au décile estimé (échec).

On cherche quelles variables ont une influence sur cette répartition. On réalise ainsi des régressions logistiques pour expliquer la probabilité qu'une observation soit inférieure au décile estimé, sachant que nous étudions telle variable, avec tel décile, pour telle station et telle date de semis, avec telle méthode, en utilisant un tel nombre d'années antérieures. Les modèles mis en œuvre utilisent toutes les variables explicatives citées ci-dessus ainsi que leurs interactions d'ordre 2. Puisque cette probabilité de classement dépend par construction du décile étudié, on travaillera décile par décile.

Critère d'évaluation de la stabilité

Les estimations des niveaux de risque pour l'année $N + 1$ ne doivent pas être trop sensibles à l'année N . Nos estimations doivent être stables suivant les années c'est-à-dire qu'elles peuvent suivre une tendance linéaire du temps mais ne pas osciller autour de celle-ci. On fait en effet l'hypothèse qu'on ne peut pas mettre en évidence une tendance autre que linéaire au vu du faible nombre d'années observées et de la nouveauté du phénomène de changement climatique.

Afin de quantifier cette stabilité, on calcule les premières différences des estimations pour chaque station x date de semis x critère phénoclimatique. Les premières différences sont celles entre les niveaux de risque estimés à 2 années successives. Dans le cas d'une tendance linéaire parfaite, les premières différences sont toutes égales à la pente de la courbe. Ainsi, pour chaque station x date de semis x critère phénoclimatique, le critère de stabilité est l'écart-type des premières différences. Un écart-type élevé indique une faible stabilité des estimations suivant les années.

On cherche ensuite quelle combinaison méthode x nombre d'années antérieures utilisées dans les calculs est à la fois plus précise et stable. Pour cela, on calcule la racine de la moyenne des erreurs ainsi que la moyenne de l'écart-type des premières différences pour tous les couples méthode x nombre d'années antérieures.

2.4 Résultats

La figure 13 montre des exemples d'estimations d'un critère phénoclimatique obtenus via les trois méthodes présentées précédemment ainsi que différents niveaux d'antériorité. L'analyse des résultats vise à évaluer la performance des différentes méthodes testées dans le cadre de la problématique du choix des dates de semis. On cherche à comprendre comment varie la précision et la stabilité de nos estimations puis on tente de déterminer le couple méthode x nombre d'années antérieures optimal

(ou les couples optimaux) dans l'évaluation de nos critères phénoclimatiques qui déterminent le choix des dates de semis.

2.4.1 Analyse des valeurs de précision

Pour étudier les résultats de précision, on réalise des régressions logistiques. On notera qu'il aurait été préférable de réaliser un modèle *global* par décile (voire un modèle général). Nous devons cependant traiter beaucoup de données. Par manque de temps, nous n'avons pas pu nous approprier les outils nécessaires au traitement d'un tel jeu de données. Nous présenterons donc des "zooms" sur certains aspects.

Dans un premier temps, étudions un sous jeu de données relatif à une seule station. La table de la déviance de la régression logistique nous indique que la variable phénoclimatique explique au moins 50% de la déviance totale. Or, certains de nos critères phénoclimatiques sont très particuliers. En effet, les variables de gel hivernal étant très restrictives à des températures extrêmes, elles sont souvent constamment négatives toutes les années évaluées à savoir de 1986 à 2011. Les estimations de déciles sont donc constamment nulles et, par conséquent, toujours égales aux observations (100%). On réitère alors la régression logistique précédente en éliminant les données relatives à ces variables. La variable phénoclimatique explique encore la déviance de la variable à expliquer mais dans une plus faible mesure. Il semble donc intéressant de regarder variable par variable les variations de ce classement. Nous ne présentons les résultats que pour un exemple.

On réalise une régression logistique expliquant le terme *nbInf* du sous jeu de données restreint au critère Maturite.Epiaison.sRg étudié en médiane. Le modèle contenant tous les effets et leurs interactions d'ordre 2 significatives au seuil de probabilité 0.05 est :

$$\text{logit}(\pi(x)) = \frac{\pi(x)}{1-\pi(x)} = \beta_0 + \beta_{1,i} + \beta_{2,j} + \beta_3 \cdot x^{(3)} + \beta_{4,k} + \beta_{5,ij} + \beta_{6,i} \cdot x^{(3)} + \beta_{7,ik} + \beta_{9,jk} + \beta_{10,k} \cdot x^{(3)}$$

avec :

- i indice associé à la station (14 valeurs).
- j indice associé à la date de semis (17 valeurs)
- k indice associé à la méthode utilisée (3 valeurs)
- $x^{(3)}$ le nombre d'années antérieures prises en compte

On se fixe les contraintes $\sum_i \beta_{1,i} = 0$, $\sum_j \beta_{2,j} = 0$, $\sum_k \beta_{4,k} = 0$, $\sum_{i,j} \beta_{5,ij} = 0$, $\sum_i \beta_{6,i} = 0$, $\sum_{i,k} \beta_{7,ik} = 0$, $\sum_{j,k} \beta_{9,jk} = 0$ et $\sum_k \beta_{10,k} = 0$.

La table de la déviance de ce modèle (Tableau 6) nous indique qu'un tiers de la déviance est expliquée par l'interaction *Station : Méthode*. On a donc une méthode optimale par station et par critère phénoclimatique.

Or Arvalis utilise actuellement une unique méthode x nombre d'années antérieures pour tous les cas de figure. Par soucis de simplicité, nous décidons donc de ne retenir qu'un couple méthode x nombre d'années antérieures.

2.4.2 Étude globale de la précision

La précision moyenne présentée figure 14 est la racine carrée de la moyenne des erreurs de prédiction de tous les triplets station x date de semis x critère phénoclimatique par méthode x nombre d'années antérieures (voir tableau 5). La méthode la plus précise est la méthode *lm* dès lors que plus de 5 années antérieures sont utilisées dans l'estimation d'une année. Pour les méthodes *lm* et *rq*, les courbes ont des allures similaires avec un nombre d'années antérieures optimal de 12 et 6 ans respectivement. Pour ces méthodes, il semble que ces degrés d'antériorité correspondent à un compromis entre

	Df	Deviance	Resid. Df	Resid. Dev	Pr(> Chi)
NULL			19991	15073.6	
Station	13	1972.2	19978	13101.4	$< 2.2e - 16$
Semis	16	85.2	19962	13016.2	$1.874e - 11$
nbAn	1	540.4	19961	12475.8	$< 2.2e - 16$
methode	2	40.3	19959	12435.6	$1.817e - 09$
Station : Semis	208	748.0	19751	11687.5	$< 2.2e - 16$
Station : nbAn	13	1016.8	19738	10670.7	$< 2.2e - 16$
Station : methode	26	4495.7	19712	6175.1	$< 2.2e - 16$
Semis : methode	32	55.2	19680	6119.9	0.006683
nbAn : methode	2	686.7	19678	5433.3	$< 2.2e - 16$

TABLE 6 – *Table d’analyse de la déviance associée au modèle de régression logistique expliquant le critère Maturite.Epiaison.sRg étudié en médiane.*

une trop forte sensibilité aux années justes précédentes ($N, N - 1...$) et une prise en compte de trop d’années qui ne permet pas de mettre en évidence des tendances récentes par des régressions linéaires. La courbe de la méthode AF est quant à elle en escalier, avec des pics toutes les 5 années antérieures supplémentaires utilisées dans les estimations pas-à-pas. Cette allure se lisse lorsque l’antériorité augmente. Cette allure vient des résultats de précision moyens pour tous les critères phénoclimatiques étudiés en décile 8. En particulier, elle est déterminée par les variables discrètes de gel qui sont binaires.

2.4.3 Étude globale de la stabilité

La stabilité moyenne présentée figure 14 est la moyenne de l’écart-type des premières différences de tous les triplets station x date de semis x critère phénoclimatique par méthode x nombre d’années antérieures. Plus on a recours à de nombreuses années passées, plus nos estimations de niveaux de risque sont stables avec les années. Pour un même degré d’antériorité, la méthode AF est la plus stable, suivie de la méthode lm et enfin de la rq .

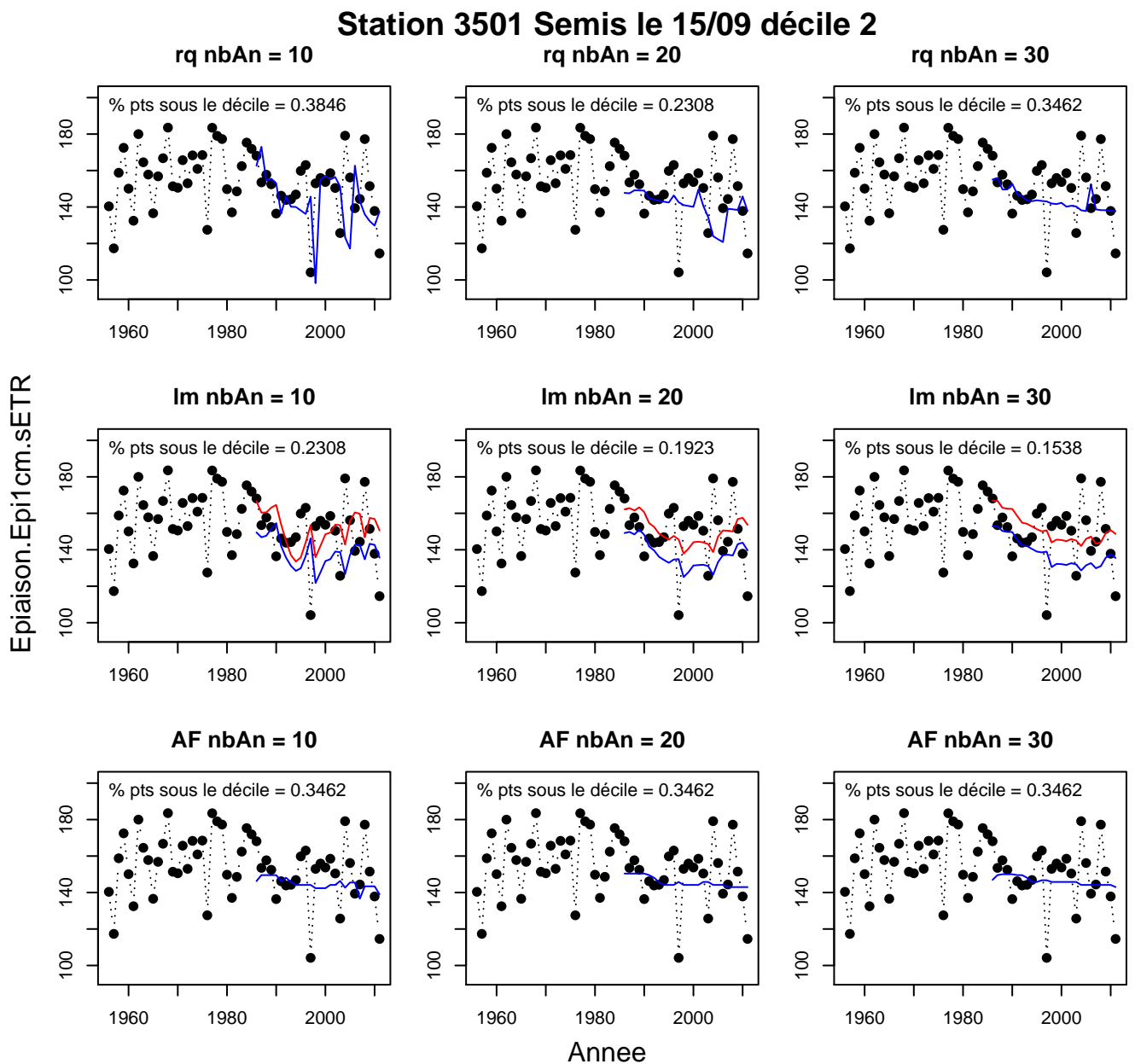


FIGURE 13 – Exemple des estimations du décile 2 (courbes bleues) de la variable *Epiaison.Epi1cm.sETR* pour les années 1986 à 2011 obtenues par les 3 méthodes avec 10, 20 et 30 années antérieures utilisées à chaque estimation. En rouge, l'estimation de la médiane par la méthode *lm*.

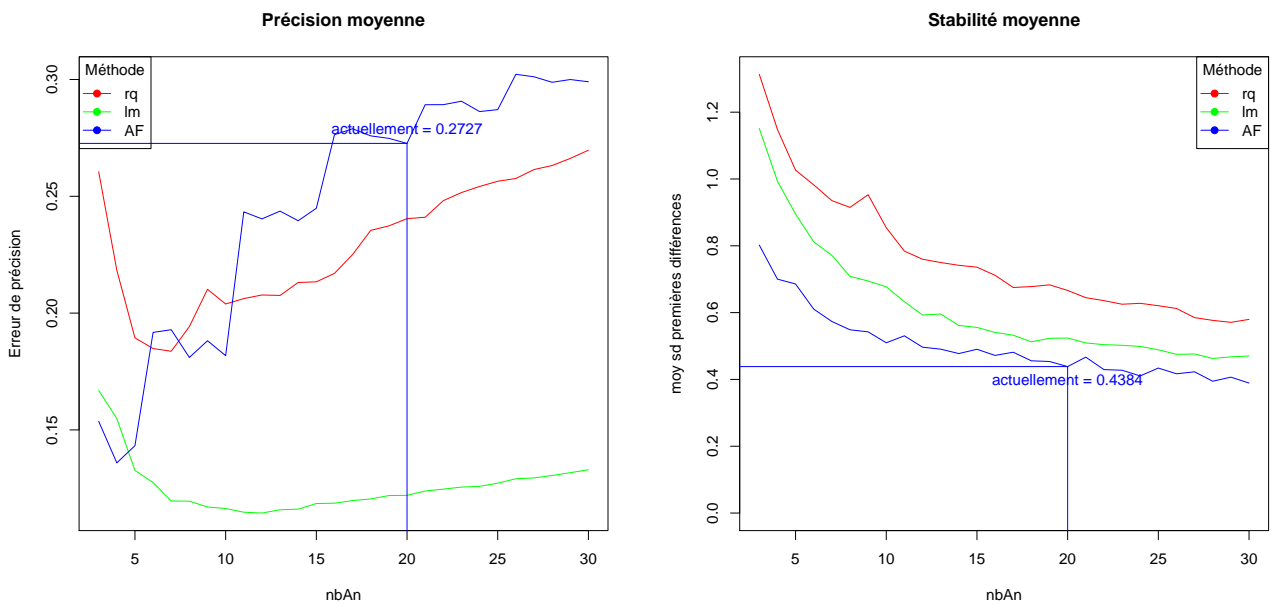


FIGURE 14 – Précision et stabilité moyenne pour chaque méthode suivant le nombre d'années antérieures utilisées

2.5 Commentaires des résultats

2.5.1 Compromis précision/stabilité

D'après les résultats, le couple méthode x nombre d'années antérieures optimal de façon globale correspond donc à un compromis entre précision et stabilité. Cette simple analyse des résultats nous conduit à choisir le couple *lm x 15ans*.

2.5.2 Les limites de la méthode *lm*

Bien que le couple *lm x 15ans* semble le plus adapté dans notre étude, il présente des inconvénients. Par construction, les résultats de la méthode *lm* sont particuliers. En effet, on estime dans un premier temps une moyenne sur les 15 années passées. Contrairement à une estimation de médiane, un point extrême influence grandement les résultats. Aussi, en ajoutant 0.84 fois l'écart-type des résidus pour l'estimation d'un décile 8 par exemple, on augmente à nouveau le poids des années extrêmes. La figure 15 montre la mauvaise estimation réalisée via la méthode *lm* dans le cas d'un gel destructeur.

De plus, les hypothèses sur les résidus de la régression linéaire ne sont toujours respectées.

2.5.3 Les avantages de la méthode *rq*

Puisque nous ne pouvons pas utiliser la méthode *lm*, nous choisissons de retenir le couple *rq x 20ans* qui semble être un bon compromis global entre précision et stabilité. Grâce la méthode de régression quantile, on prend alors en compte les potentielles tendances des variables phénoclimatiques qui peuvent avoir lieu avec ce climat changeant.

2.6 Bilan

Les différents couples méthode x antériorité ont été testés dans le cadre de la problématique du choix des dates de semis. Par souci de simplicité, nous avons décidé de ne retenir qu'un couple méthode x nombre d'années antérieures. Une étude, uniquement ponctuelle, des résultats de précision nous montre bien qu'un unique couple commun à toutes les stations et tous les critères n'est pas idéal (cf. partie 2.4.1). Le couple *rq x 20 ans* retenu correspond cependant à un bon compromis global entre précision et stabilité.

Bien que cette étude n'ait été réalisée que dans le cadre de la problématique du choix des dates de semis, elle montre qu'il existe des méthodes, alternatives à l'analyse fréquentielle, qui peuvent être plus performantes que celle-ci, et qui sont plus adaptées à l'estimation de niveaux de risque dans le contexte du changement climatique. L'institut devrait donc poursuivre cette étude en testant la performance de la régression quantile, calculée sur les valeurs prises par d'autres types de critères phénoclimatiques lors de 20 années antérieures.

Station 1801 Semis le 30/09 décile 8

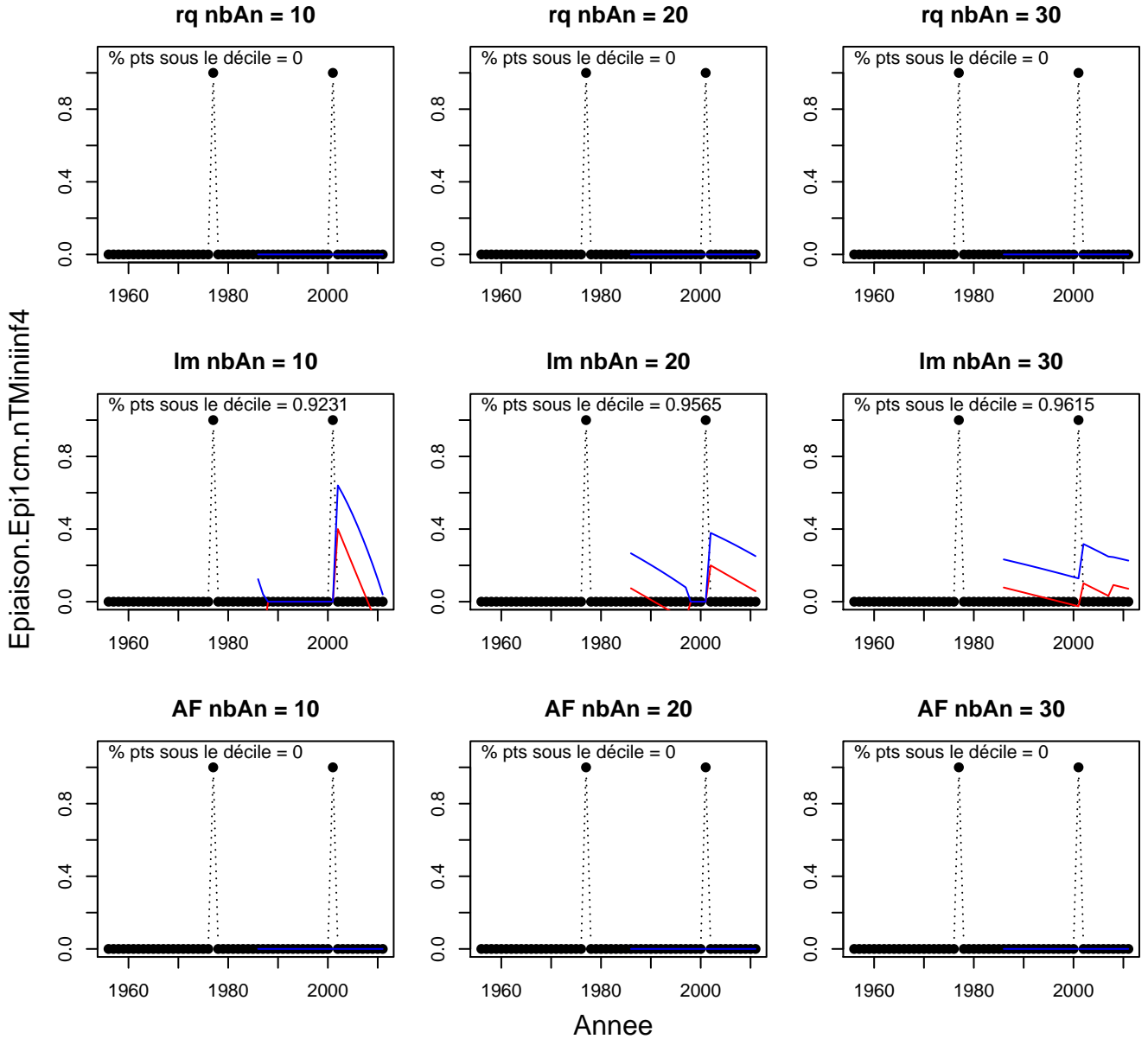


FIGURE 15 – Exemple des estimations du décile 8 de la variable *Epiaison.Epi1cm.nTminiinf4* illustrant les limites de la méthode *lm*

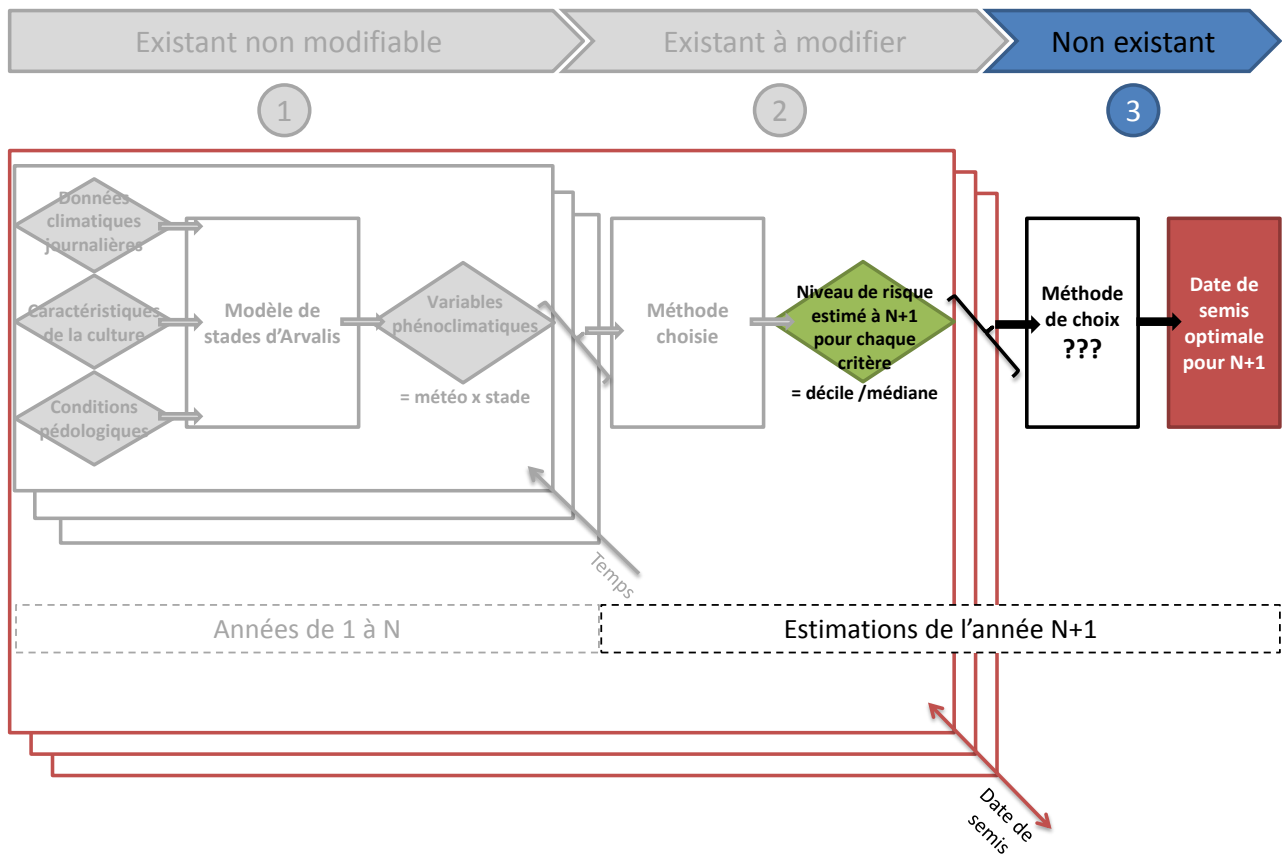


FIGURE 16 – Schéma résumé de la méthode de choix des dates de semis que l'on cherche à déterminer au chapitre 3. La boîte verte représente les données d'entrée de ce modèle qui sont des niveaux de risques relatifs à plusieurs choix de date de semis pour l'année suivante notée $N + 1$. En sortie (boîte rouge), on obtient une préconisation de semis pour l'année $N + 1$.

Chapitre 3

Préconiser des dates de semis à partir des estimations de critères phénoclimatiques d'intérêt

3.1 Introduction

Le choix des dates de semis se fait de façon indirecte. En particulier, on cherche à relier les dates de semis au rendement par l'intermédiaire de variables phénoclimatiques dont on sait qu'elles ont une influence sur le rendement. Ces variables phénoclimatiques ont été retenues par les écophysiologistes et ne sont donc pas absolues. Puisque les valeurs prises par ces variables sont très changeantes d'une année à l'autre, on cherche à décrire la distribution de ces variables. On estime notamment les niveaux de risque pour l'année $N+1$. Nous avons vu que cette estimation se fait originellement grâce à l'analyse fréquentielle mais on a montré qu'il est possible de mettre en œuvre d'autres méthodes plus adaptées.

Sur une station, les estimations de niveaux de risque pour l'année $N + 1$ sont effectuées pour tous les critères phénoclimatiques retenus et pour toutes les dates de semis. On étudie alors, dans un second temps (Figure 16), l'évolution de ces niveaux de risque en fonction des dates de semis. On cherche à minimiser ou maximiser les valeurs suivant la nature des variables phénoclimatiques étudiées (Tableau 4). On ne sait, pour le moment, pas de déterminer de limites d'acceptabilité relatives aux variables phénoclimatiques choisies. La date de semis retenue est donc celle qui optimise un maximum de critères phénoclimatiques. Ceci est effectué station par station puisque les dates de semis dépendent du climat.

L'objectif de ce chapitre est de mettre en place une nouvelle méthodologie qui permette de définir les dates de semis optimales correspondantes à des estimations de niveaux de risque phénoclimatique. Les résultats présentés ici sont basés sur *un* choix de critères phénoclimatiques. Ils seront donc déclinés pour l'exemple et doivent être regardés avec précautions. Nous chercherons toutefois à savoir si ces premiers résultats de préconisation de dates de semis sont cohérents avec l'expertise des écophysiologistes afin de voir si les choix, en amont, de critères phénoclimatiques ont été pertinents.

3.2 Données utilisées

Pour chaque station, on a un tableau dates de semis x critères phénoclimatiques qui contient, à chaque intersection d'une ligne et d'une colonne, le niveau de risque estimé, pour l'année $N + 1$, à l'aide de la méthode $rq \times 20$ ans.

3.3 Méthodes de choix des dates de semis à préconiser

3.3.1 Définition de fenêtres possibles par station

Nous avons déterminé des limites de risque acceptable uniquement pour certaines variables, à savoir les gels destructeurs (Tableau 4, page 20). On considère que leurs niveaux de risque (décile 8) doivent être nuls i.e. qu'il faut éviter ces risques 8 années sur 10. Dans un premier temps, on détermine alors, pour chaque station, les dates de semis qui permettent d'éviter "en général" ces risques destructeurs. Ceci correspond à la démarche utilisée historiquement par l'institut.

3.3.2 Optimisation du choix grâce à l'ensemble des critères phénoclimatiques

Un indicateur de stress global : la fonction de coût

Ensuite, on cherche les dates de semis optimales pour tous les autres critères. On crée alors un indicateur, appelé fonction de coût, dont la valeur est d'autant plus élevée que la date de semis s'éloigne de la date de semis optimale. On commence par transformer les variables à maximiser en les multipliant par -1 afin de minimiser toutes les variables. Puis, on centre et réduit chaque variable sur tous les couples station x date de semis dans le but de les comparer entre elles (Figure 17). On soustrait ensuite à chaque variable son minimum global pour rapporter leur minimum à zéro. Pour chaque variable, le zéro correspond alors à la valeur optimale i.e. à une absence de contrainte climatique.

Par station x date de semis, les variables ne sont donc pas réduites et n'ont pas forcément un minimum à zéro. Par station i x date de semis j , on définit alors la fonction de coût, notée $C_{i,j}$:

$$C_{i,j} = \sum_k var_{i,j,k}$$

avec $var_{i,j,k}$ le niveau de risque du critère phénoclimatique k , pour la station i et la date de semis j . Les dates de semis j optimales pour une station i sont celles pour lesquelles la fonction $C_{i,j}$ est minimum. Par construction, on peut comparer le coût obtenu d'une station à l'autre. La fonction de coût $C_{i,j}$ correspond à un niveau global de stress subi par la culture sur la station i , si on sème à la date j .

Relier les dates de semis à chaque critères phénoclimatiques

Il est alors intéressant de comprendre quels sont les critères phénoclimatiques minimisés en choisissant ces dates de semis pour une station. Une ACP¹ non normée, réalisée sur le tableau dates de semis x critères phénoclimatiques, permet alors d'étudier les effets de la date de semis sur les critères phénoclimatiques. Les colonnes ne sont volontairement pas réduites afin de structurer les dates de semis préférentiellement sur les critères ayant une variance élevée sur cette station. De façon plus globale, cette analyse nous permet de savoir quelles sont les dates de semis qui se ressemblent du point de vue des critères phénoclimatiques, ainsi que d'étudier les liaisons entre ces critères. Les résultats ont été obtenus grâce au package FactoMineR du logiciel R (Husson et al., 2012).

1. Analyse en Composantes Principales

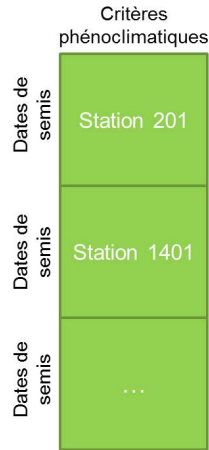


FIGURE 17 – Structure des données permettant de calculer le coût par date de semis pour chaque station. Chaque colonne de ce tableau est centrée/réduite puis son minimum est rapporté à zéro.

3.4 Résultats de préconisations

3.4.1 Quelles dates de semis optimales ?

Sur la figure 18 sont représentées par des carrés bleus les dates de semis pour lesquelles on évite tous les risques destructeurs par station. À l'exception de la station 1602, toutes les dates de semis permettent d'éviter ces critères ou, à l'inverse, aucune. Il apparaît donc que ces critères phénoclimatiques ou que leur seuil d'acceptabilité (décile 8 qui vaut zéro) ne sont pas bien choisis. Il faut donc revoir :

- soit les stades de sensibilité du blé tendre au gel
- soit la température minimale à partir de laquelle on considère qu'il y a un gel
- soit le seuil d'acceptabilité

Dans la suite de l'analyse, nous ne tiendrons donc pas compte de ces risques destructeurs.

On réalise les estimations des niveaux de risque pour 2012. L'évolution de la fonction de coût suivant les dates de semis est présentée pour les stations de Rennes (3501) et Lyon (6901) à la figure 19. La station de Lyon a des conditions plus stressantes que Rennes puisque, quelle que soit la date à laquelle le semis est réalisé, le coût est plus élevé. Pour Lyon, plus on sème tôt, plus on optimise nos critères phénoclimatiques. Dans le cas de Rennes, la fonction de coût est minimale pour des dates de semis intermédiaires.

3.4.2 Quelle optimisation ?

Afin de comprendre la cause de ces tendances, on réalise l'ACP décrite dans la partie 3.3.2, pour la station de Rennes en 2012. Les résultats sont présentés figure 20. Le premier plan factoriel regroupe 79.53% de l'inertie totale. D'après le diagramme des inerties, ce plan résume bien la variabilité de nos 17 dates de semis. Seules les variables bien représentées sur ce plan ont été tracées (\cos^2 supérieur à 0.5). Les variables de l'analyse sont les variables phénoclimatiques étudiées en décile 2, 5 et 8. Puisqu'une variable phénoclimatique est étudiée à la fois en décile 2 ou 8, et en médiane, la notation de la variable est suivie du chiffre 2, 5 ou 8 suivant le décile étudié. Le carré de la coordonnée d'une variable sur un axe nous renseigne sur la contribution de cette variable à la formation de la compo-

Carré = date de semis possible sur cette station en 2012

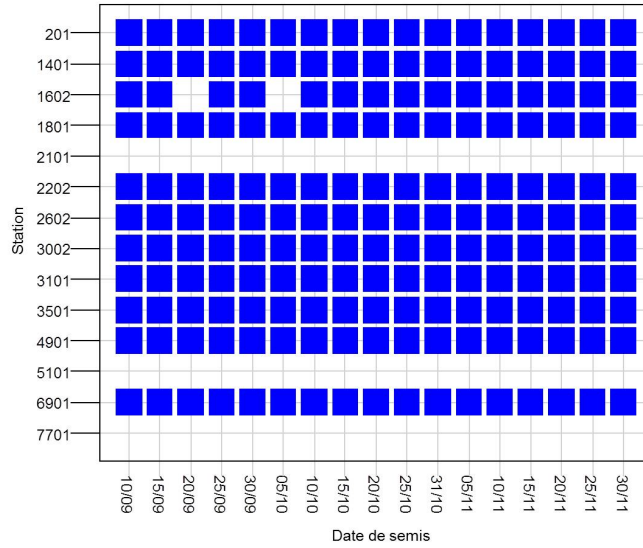


FIGURE 18 – Dates de semis permettant d’éviter les risques destructeurs de gel par station à partir des estimations réalisées pour 2012 avec le couple méthode rq x 20 années antérieures

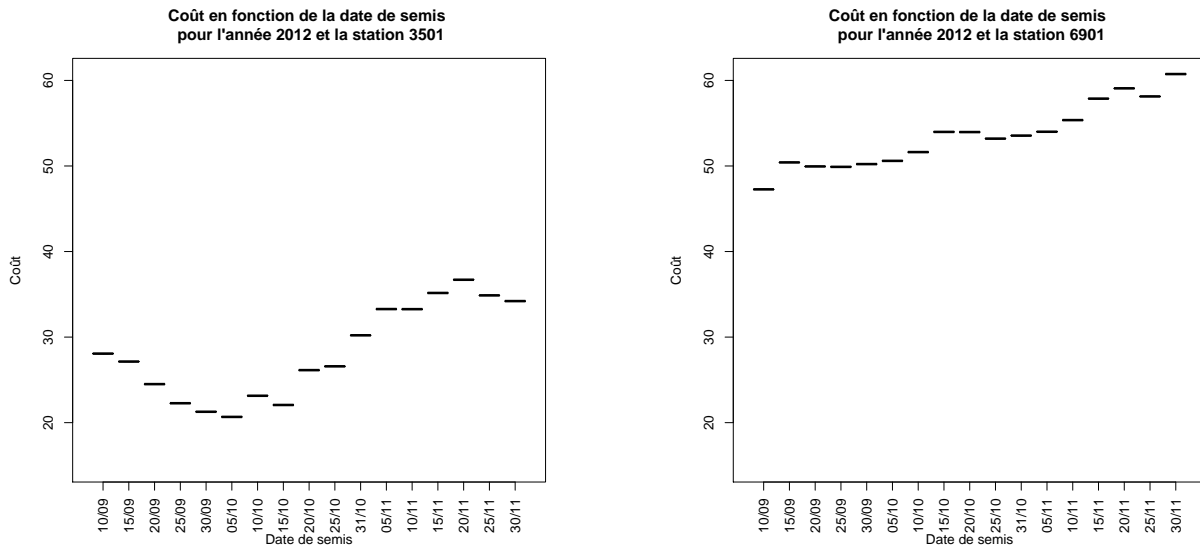


FIGURE 19 – Coût associé aux dates de semis les stations de Rennes (3501) et Lyon (6901) pour les estimations de 2012 calculées à l’aide la méthode rq x 20 années antérieures.

sante principale. Le premier axe est essentiellement dû aux variables $E_{piaison750C}$, $E_{piaison.sETR.5}$ et $Méiose5j$, $Méiose5j.nTminiint4.8$ qui totalisent 66% de son inertie. Ces critères phénoclimatiques structurent donc le plus nos dates de semis. La contribution des autres variables au second axe est plus homogène. Au niveau des dates de semis, la contribution des semis du 10 et 15 septembre au second axe factoriel est particulièrement élevée (69% de l'inertie de l'axe).

Sur le nuage des individus, on remarque que les dates de semis sont ordonnées chronologiquement. Les semis tardifs qui ont des coordonnées élevées sur l'axe 1, sont marqués par des risques gel à la méiose, de températures élevées au remplissage et de stress hydrique en fin de cycle plus élevés. A l'opposé, les semis précoces qui ont des coordonnées négatives sur cet axe, sont caractérisés par moins de rayonnement entre l'épiaison et la maturité. Les semis extrêmement précoces du 10 et 15 septembre que nous avons déjà identifiés comme atypiques ont des risques de gel à la méiose très élevés alors que leurs risques de stress hydriques depuis le stade épi 1 cm jusqu'à la fin du cycle sont faibles par rapport aux autres dates de semis. Ainsi, les risques évoluent suivant la date à laquelle le semis est réalisé.

Le coût associé à chaque date est presque équivalent à sa coordonnée sur l'axe 1. En effet, la fonction de coût et le première axe ont une corrélation de 0.96. Le coût que nous avons défini correspond quasiment à la première dimension de variabilité des dates de semis. Il est au contraire indépendant du critère $Méiose5j$, $Méiose5j.nRginf200.8$ (corrélation de -0.05) qui ne varie presque pas suivant les dates de semis.

Dans notre analyse, nous avons pris en compte nos variables phénoclimatiques du point de vue de leur tendance générale d'une part (exprimée sous forme de médiane ou décile 5) et du point de vue des valeurs extrêmes d'autre part (décile 2 ou 8 suivant la variable). On note que la liaison entre deux points de vues d'une même variable phénoclimatique est parfois faible, voire négative, comme par exemple la somme des rayonnement (saPAR) dont la corrélation entre la médiane et le décile 2 est de -0.23 . Cependant, les variables qui interviennent dans la fonction de coût ont un rôle d'autant plus important que la corrélation entre les deux points de vues est élevée. Ils font alors "double emploi". C'est par exemple le cas de la variable $E_{piaison750C}$, $E_{piaison.nTminisup15}$ dont la corrélation entre les estimations de médiane et de décile 8 vaut 0.88.

3.5 Commentaire des résultats

3.5.1 Le choix des limites de la plage de semis

La fonction de coût nous renseigne sur un classement des dates de semis. Ne connaissant pas les seuils d'acceptabilité de la majorité des critères phénoclimatiques, nous ne pouvons définir de limite à la plage de semis. Différentes méthodes sont envisageables.

Valeur maximale absolue de la fonction de coût

Une possibilité naturelle serait de choisir les dates de semis pour lesquelles la fonction de coût est inférieure à une certaine valeur. Cette méthode reviendrait à donner un seuil d'acceptabilité à l'ensemble des critères phénoclimatiques. Le choix de cette valeur absolue semble difficile. De plus, les valeurs de coût sont parfois très différentes suivant les stations (voir Rennes et Lyon pour 2012 figure 21). Une valeur unique à toutes les stations conduirait à accepter toutes les dates dans certaines stations ou bien les rejeter toutes dans d'autres. Tout comme les rendements attendus sont différents d'une région à l'autre, le coût maximum toléré est variable suivant les conditions du milieu.

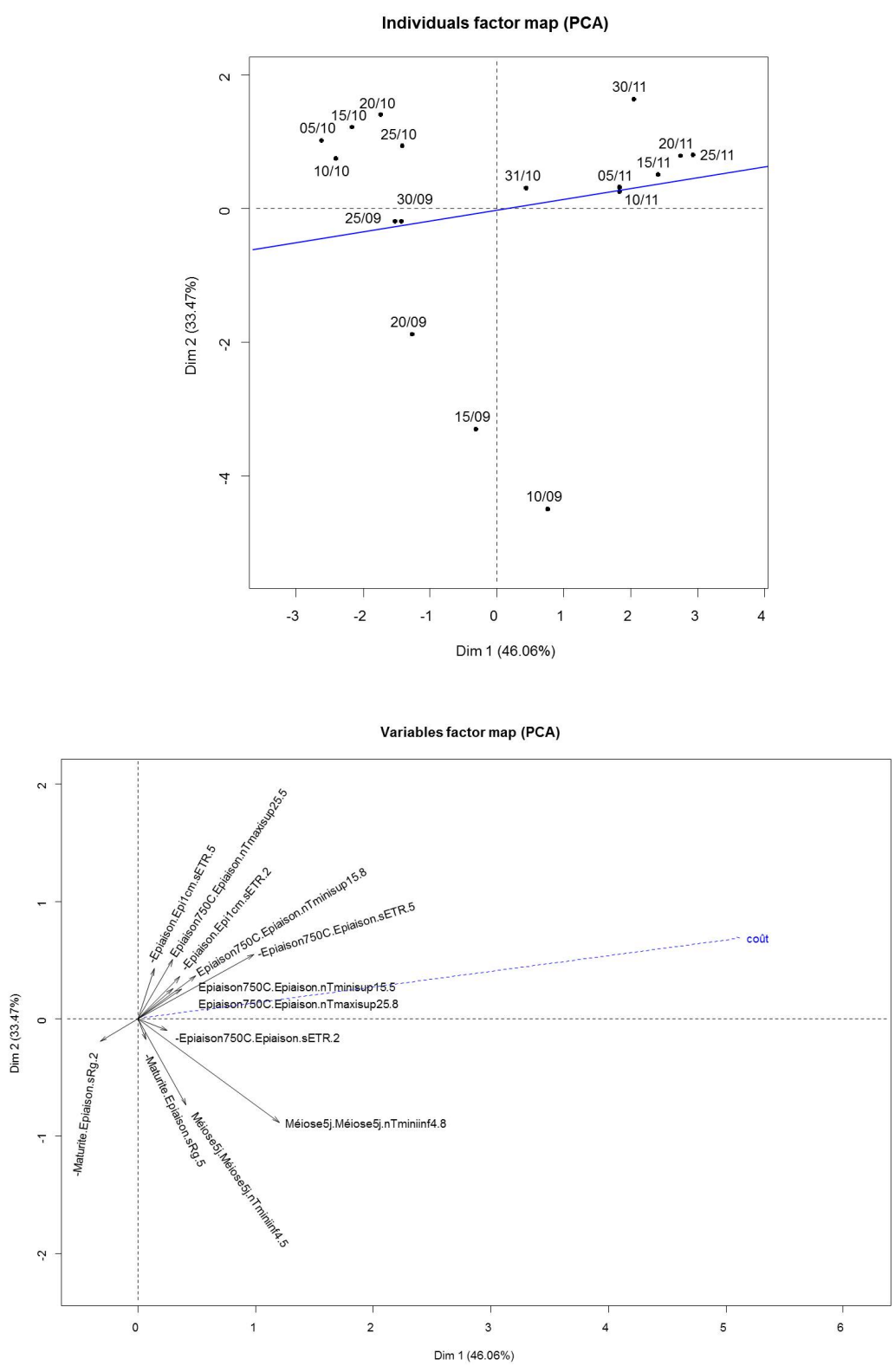


FIGURE 20 – Graphiques de individus et des variables de l'ACP décrivant la structure des dates de semis suivant les critères phéno-climatiques pour les station de Rennes (3501) et les estimations de niveaux de risque de 2012. Le signe "-" devant le nom de certaines variables indique qu'elles ont été multipliées par -1. Il faut donc lire la flèche dans le sens opposé.

Utiliser les coûts acceptés historiquement

On peut estimer la fonction de coût pour une année passée, par exemple 1997 (Figure 21). Afin de pouvoir comparer les valeurs de coût de 2012 et 1997, on a préalablement réduit les critères phénoclimatiques sur toutes les stations, dates de semis et sur les 2 années. Connaissant les dates de semis préconisées historiquement (Annexe H), on peut évaluer quel était le coût maximal qui était toléré historiquement. Pour Lyon, la plage de semis historique s'étend du 15 au 25 octobre ce qui correspondrait à un coût maximal accepté historiquement de 34.8. Premièrement, on remarque sur les estimations de 1997 que toutes les dates de semis pour lesquelles le coût est inférieur à 34.8 ne sont pas retenues. Ceci nous indique que les critères phénoclimatiques que nous avons retenus et/ou leurs poids ne correspondent pas à ceux retenus historiquement. Nous ne pouvons donc pas utiliser cette valeur de coût "historique". De plus, un tel coût nous mène aujourd'hui à un rejet de toute la plage de semis.

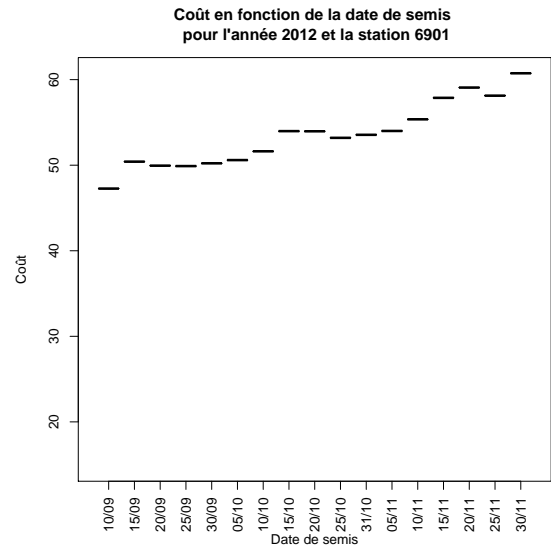
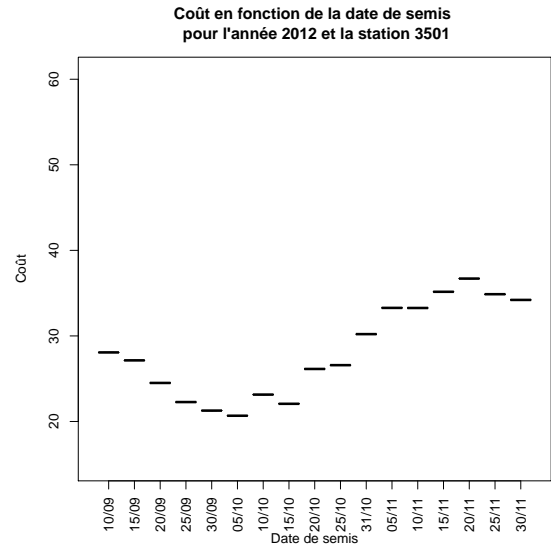


FIGURE 21 – Comparaison du coût associé aux dates de semis les stations de Rennes (3501) et Lyon (6901) pour les estimations de 1997 (à gauche) et celles de 2012 (à droite) calculées à l'aide la méthode $rq \times 20$ années antérieures.

Relier le coût à une perte de rendement

La fonction de coût correspond à un niveau de stress global subi par la culture. Il prend en compte de manière générale les différents stress qui s'appliquent sur une culture, un stress élevé pouvant être compensé par un autre plus faible. Nous pourrions essayer de relier chaque coût à une perte de rendement par rapport à un rendement maximum. Ainsi les plages de semis seraient délimitées pour ne pas perdre plus de x quintaux/hectare de rendement. Chaque lien coût - rendement dépendrait donc des pratiques culturales.

Prise en compte de nouveaux critères

Nous avons étudié ici que des critères physiologiques. La fonction de coût traduit donc l'optimisation des dates de semis uniquement du point de vue des besoins de la plante. Toutefois, d'autres critères interviennent dans le choix des dates de semis :

- le risque de développement d'adventices
- le risque de maladies
- la disponibilité de l'agriculteur
- l'accessibilité de la parcelle qui dépend de l'avancée des cultures précédentes
- la faisabilité des travaux du sol (préparation et semis) liée aux conditions climatiques

On peut donc essayer de coupler ces critères à notre optimisation physiologique afin de définir les limites de notre plage de semis. Cette dernière méthode est la seule réalisable.

3.5.2 Lien avec les préconisations historiques

Nous avons vu précédemment (partie 3.5.1) que les critères phénoclimatiques retenus ne correspondent pas à ceux utilisés historiquement dans leurs définitions et/ou dans leurs poids relatifs. L'évolution de la fonction de coût entre 1997 et 2012 nous indique cependant les modifications globales de nos critères.

La figure 21 représente conjointement la fonction de coût estimée pour les années 1997 et 2012. Si l'allure de la fonction de coût reste la même pour Lyon, celle-ci est modifiée dans le cas de Rennes. De plus, les conditions sont actuellement plus stressantes à Lyon, à l'inverse de Rennes. Dans le cas de Rennes, il est évident que les critères phénoclimatiques n'ont pas évolué tous conjointement puisque que l'allure est déformée. Il serait intéressant de réaliser des ACP sur les tableaux dates de semis x critères phénoclimatiques contenant les niveaux de risque estimés pour 1997 afin de comprendre quelle a été l'évolution des conditions stressantes.

3.6 Bilan

La méthode mise en place ici ne nous permet que de graduer les dates de semis suivant une optimisation globale de critères phénoclimatiques. Il serait nécessaire de prendre en compte d'autres types de critères afin de délimiter la plage de semis optimale par station. Nous devrions en outre revoir les critères phénoclimatiques qui interviennent dans le choix des dates de semis, définir des seuils d'acceptabilité et hiérarchiser les risques/besoins.

Chapitre 4

Discussion générale

Si les 3 chapitres précédents comportaient une section de commentaire des résultats, cette discussion générale a pour but de s'interroger sur la pertinence globale de notre travail et les limites de nos données, méthodes et résultats.

4.1 Discussion agronomique des résultats de préconisations de dates de semis

4.1.1 Quelle validité des préconisations antérieures ?

Nous avons vu précédemment que nous ne réussissons pas à retrouver les préconisations historiques (voir section 2.2 page 19 et section 3.5 page 35) en utilisant strictement :

- les variables phénoclimatiques historiques
- les données climatiques historiques
- la méthode d'évitement des risques phénoclimatiques par analyse fréquentielle (méthode utilisée par Arvalis)

On peut, dans un premier temps, supposer que les critères permettant de choisir les dates de semis ne sont pas les mêmes pour tout le territoire mais qu'il existe des adaptations réalisées "à la marge" pour chaque région. En effet, même si certains milieux sont plus stressants pour le blé tendre, la répartition actuelle des cultures en France montre bien qu'il est possible de cultiver du blé dans toute la métropole de façon économiquement viable. Cependant, ce sont les techniques culturales qui s'adaptent.

De plus, nous ne connaissons pas la formalisation exacte du raisonnement qui a été employé pour fournir ces préconisations (Annexe H). Historiquement, le modèle de stades d'Arvalis n'était pas aussi abouti et ne permettait donc pas de simuler, comme nous l'avons fait, les variables phénoclimatiques. Les raisonnements associés étaient plus succincts. On regardait dans un premier temps la date d'apparition du dernier gel par exemple. En particulier, on s'intéressait à la date à laquelle le dernier gel avait lieu 8 années sur 10. Puis, on calculait "à rebours" la date de semis qui permettait que le dernier gel ait lieu une fois quand le blé n'y est pas sensible, à savoir avant le stade épi 1 cm (Figure 1).

4.1.2 Le choix des variables phénoclimatiques, de leur seuil, des limites acceptables

Nos résultats sont une hiérarchisation des dates de semis basée exclusivement sur des considérations physiologiques. Ces résultats sont directement liés au choix :

- des variables phénoclimatiques. *Ex* : gel entre le stade semis 1 cm et l'épiaison
- de leur seuil. *Ex* : des températures minimales journalières inférieures à -4°C
- de leur limite d'acceptabilité. *Ex* : le décile 8 de la variable phénoclimatique est nul

Nous n'avons retenu que quelques critères par souci de simplicité. Bien que le choix de ces critères ait été subjectif, il repose sur des études antérieures qui constituent un guide. Il reste toutefois difficile de connaître l'importance relative de chaque critère. Les critères que nous avons retenus peuvent donc être critiqués, en particulier, aux regards des accidents climatiques de cette campagne (2011-2012). Les cultures ont en effet été touchées par des gels dus à des températures inférieures à -8°C entre le stade double-ride et le stade épi 1 cm. De plus, ces gels ont eu d'autant plus d'impact que la succession d'épisodes climatiques n'a pas permis aux cultures de s'endurcir préalablement au froid. En effet, la sensibilité du blé au gel peut être diminuée s'il est exposé à des températures basses pendant une certaine durée d'acclimatation lors de laquelle il va produire des protéines spécifiques (Gate, 1995). Il serait alors important de considérer non seulement une variable phénoclimatique définie entre deux stades S_i et S_{i+i} mais de prendre aussi en compte les événements précédents ayant lieu aux stades S_{i-1} , S_{i-2} ... Ceci permettrait de mieux caractériser les risques climatiques.

De même, la tolérance du blé à la température dépend de la fréquence des événements extrêmes. Enfin, il serait intéressant de coupler plusieurs critères. Par exemple, la sensibilité du blé à la température dépend, à certains stades, de la réserve en eau du sol disponible pour la plante (Porter and Gawith, 1999).

Une étude agronomique préalable approfondie serait ainsi nécessaire pour définir les critères phénoclimatiques à prendre en compte. Elle doit être réalisée en deux étapes :

1. Quels sont les facteurs pénalisant ?
2. Quelles pénalités y sont assorties ?

Bien que cette étude semble chrono-phage, elle reste plus intéressante que de réaliser une étude directe du lien rendement final - date de semis qui nécessite énormément de données d'essais.

4.2 Discussion générale sur le matériel utilisé et les méthodes mises en œuvre

4.2.1 Validité des variables phénoclimatiques estimées par le modèle de stades d'Arvalis

Les données utilisées dans cette étude ont été estimées à l'aide du modèle de stades d'Arvalis qui permet de prédire les dates d'apparition des stades de la plante à partir des données climatiques journalières (partie 2.2 page 18). Les conditions pédologiques que nous avons utilisées pour paramétrer le modèle, mériteraient d'être affinées par station. Nous avons, en effet, considéré un même type de sol et une même réserve hydrique quelle que soit la région.

Le modèle de stades d'Arvalis utilisé ici permet d'estimer la date d'apparition des stades de développement de la culture. Il a, en moyenne, une erreur d'environ 5 jours pour le stade épi 1cm, et 3 jours pour l'épiaison (Gate et al., 2007), ce qui est à l'origine d'incertitudes sur nos résultats

d'estimations des variables phéno-climatiques. De même, ce modèle a été paramétré pour certains climats. On ne sait donc pas si les estimations seraient valables pour des climats extrêmes dûs au changement climatique.

Nous n'avons travaillé que sur des données simulées pour la variété de blé tendre Soissons. Toutefois, le rythme de développement du blé tendre dépend de sa génétique et, par conséquent, de la variété (on parle de précocité variétale). Nos résultats ne sont donc valables que pour une précocité variétale de type "Soissons". Puisque nous travaillons sur des critères *phéno-climatiques*, les risques seront différents pour des variétés avec des précocités différentes. Par exemple, une variété dite *tardive*, semée tard, sera plus exposée à des risques liés à de fortes températures en été. Il est donc important d'accorder les dates de semis à des précocités variétales. En travaillant avec des données issues de simulation, on peut rapidement réaliser les mêmes analyses pour d'autres variétés. Il suffit de modifier les paramètres d'entrée du modèle. Il serait donc nécessaire de réaliser par la suite ce travail pour des variétés de précocités différentes.

4.2.2 Stratégie d'analyse

Notre méthode globale (soit les étapes 2 et 3 du schéma 3, page 5) consiste à estimer une date de semis optimale, pour l'année à venir, à partir de niveaux de risques phéno-climatiques i.e. des médianes et déciles et non des variables phéno-climatiques brutes. Le passage par l'intermédiaire de médiane et déciles permet de décrire la distribution des variables phéno-climatiques suivant la date de semis paramétrée dans le modèle de stades d'Arvalis. On peut donc étudier, d'une part, la tendance générale et, d'autre part, les valeurs plus extrêmes de ces variables pour différentes dates de semis.

Notre méthode globale aboutit à une date de semis optimale prédite pour l'année $N + 1$. Il serait intéressant de comparer les dates de semis optimales prédites à l'aide de notre méthode, aux dates de semis optimales, en termes de rendement, constatées dans les quelques essais expérimentaux disponibles. Ceci n'a pas été réalisé puisque le choix des critères phéno-climatiques à prendre en compte n'est pas abouti. De plus, les données expérimentales disponibles auront un intérêt limité puisque, pour un même essai, très peu de dates de semis sont mises en œuvres.

Notre stratégie a été de considérer indépendamment les dates de semis. Or, il existe un lien clair entre les variables phéno-climatiques pour des dates de semis successives (Figure 22). On pourrait imaginer une stratégie multidimensionnelle d'estimation des niveaux de risques qui utilise les variables *année* et *date de semis*.

Maintenant que nous avons mis au point une méthode globale d'estimation des dates de semis optimales pour une année $N + 1$, celle ci pourrait s'adapter à la définition d'une date de semis optimale a posteriori pour une année passée N . Pour cela, il suffit d'estimer les niveaux de risques phéno-climatiques pour l'année N , par exemple, en utilisant la valeur estimée à l'année N par une régression quantile d'une variable phéno-climatique en fonction des années. On utilise alors notre fonction de coût pour définir a posteriori la date de semis optimale de l'année N . Il serait ensuite intéressant de regarder l'évolution des dates de semis optimale avec les années.

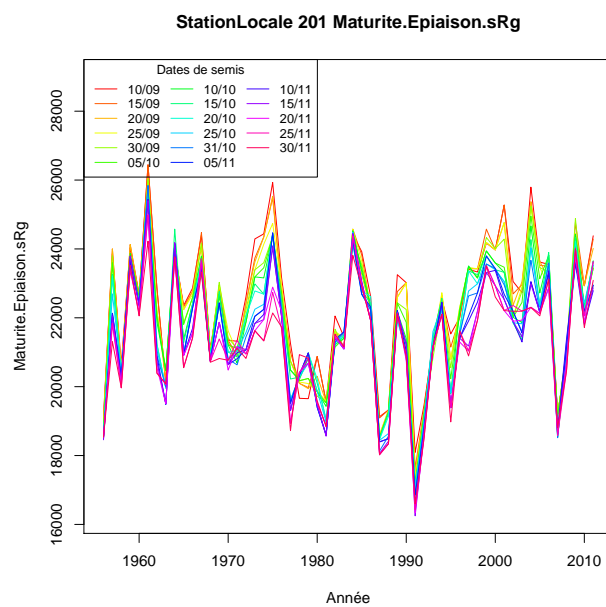


FIGURE 22 – Évolution de la variable *Maturite.Epiaison.sRg* suivant les années et les dates de semis pour la station 201.

Conclusion générale et perspectives

ARVALIS - Institut du végétal a un rôle de conseil auprès des agriculteurs, notamment, des dates de semis conduisant au plus haut rendement. De façon globale, l'institut cherche à préconiser les dates de semis qui permettent d'optimiser plusieurs variables phénoclimatiques. L'objectif de ce mémoire a été de repenser la succession de méthodes associées au choix des dates de semis, celles-ci ayant des lacunes ou bien étant inexistantes.

Ce travail a permis de faire avancer les connaissances actuelles d'Arvalis en terme de méthodologie à mettre en œuvre pour définir les dates de semis du blé tendre. Préalablement à ce stage, l'institut pensait que les préconisations ne nécessitaient qu'une mise à jour afin de prendre en compte notamment la dérive climatique actuelle. Nous avons montré que ce problème était plus complexe et nous avons apporté des voies de réponse possibles. Ce travail a pointé du doigt les erreurs commises par la méthode d'analyse fréquentielle. En outre, la méthode alternative proposée est plus pertinente pour analyser le changement climatique puisqu'elle tient compte des potentielles tendances des variables phénoclimatiques. Ceci est un enjeu fort pour l'institut puisque l'analyse fréquentielle est intégrée aux outils internes utilisés pour répondre à des problématiques très variées. Par ailleurs, une méthode d'optimisation des variables phénoclimatiques, préalablement inexistante, a été mise en place. Il existe maintenant une méthodologie complète de choix des dates de semis.

Enfin, ce travail a montré qu'il n'est pas possible de définir finement de dates de semis avec les connaissances physiologiques actuelles. Il est donc nécessaire de repenser au préalable le choix des variables phénoclimatiques avant d'appliquer notre méthode de détermination des préconisations. Les variables retenues ici constituent un point de départ intéressant car les premiers résultats de préconisation obtenus sont cohérents avec les dires des experts.

Bibliographie

- 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. *Field Crops Research* 119(1), 201–212.
- Deswarte, J.-C., I. Garcia, F. Cassignol, S. Dutriez, P. Gate, P. Giraudeau, C. Girousse, P. Senellart, J. Stragliati, S. Sunderwirth, G. Tcherkez, and D. Gouache (2011). Évaluation de variétés tolérantes à la sécheresse et aux fortes températures : analyse de la composition isotopique du carbone photosynthétique et de l’effet de chocs thermiques (résultats préliminaires). In *Actes de la rencontre scientifique du 13 janvier 2011*, pp. 67. GNIS.
- Gate, P. (1995). *Ecophysiologie du blé*. Paris : Lavoisier Editeur.
- Gate, P. (2007). Dossier changements climatiques - esquivé et adaptation : les solutions de demain. *Perspectives Agricoles* 336, 32–37.
- Gate, P. (2008). Date et densité de semis : une étape clé pour le succès d’une culture. *Perspectives Agricoles* 348, 46–49.
- Gate, P., N. Brisson, G. Charmet, D. Gouache, and F.-X. Oury (2009). Les causes du plafonnement du blé en france ? Congrès Arvalis INRA du 22/10/09.
- Gate, P., O. Deudon, and D. Gouache (2007). Changement climatique : impact sur le blé en france et pistes d’adaptation. *Perspectives Agricoles* 374, 54–59.
- Gate, P., M. Giban, A. Blondot, and P. Braun (2003). *Stades du blé*. Paris : ITCF.
- Gouache, D., X. Le Bris, M. Borgard, O. Deudon, C. Pagé, and P. Gate (2012). Evaluating agronomic adaptation options to increasing heat stress under climate change during wheat grain filling in france. *European Journal of Agronomy* 39, 62–70.
- Guillaume, S. (2011). *Adaptation d’un modèle de culture et conception d’un modèle de décision pour la gestion conjointe de l’irrigation et de la fertilisation azotée du blé dur*. Ph. D. thesis, INP Toulouse, Toulouse.
- Husson, F., J. Josse, S. Le, and J. Mazet (2012). *FactoMineR : Multivariate Exploratory Data Analysis and Data Mining with R*. R package version 1.18.
- Hyndman, R. J. and J. Fan (1996). Sample quantiles in statistical packages. *The American statistician* 50(4), 361–365.
- Koenker, R. (2012). *quantreg : Quantile Regression*. R package version 4.79.
- Koenker, R. and K. F. Hallock (2001). Quantile regression. *Journal of Economic Perspectives* 15(4), 143–156.

- Masse, J. (1987). Choix des variétés, des dates et des densités de semis : blé, orge et blé dur d'hivers. In *Implantation des cultures*, Blois - le Breuil, pp. 2–12.
- Olesen, J. E., M. Trnka, K. C. Kersebaum, A. O. Skjelvåg, B. Seguin, P. Peltonen-Sainio, F. Rossi, J. Kozyra, and F. Micale (2011). Impacts and adaptation of european crop production systems to climate change. *European Journal of Agronomy* 34(2), 96–112.
- Porter, J. and M. Gawith (1999). Temperatures and the growth and development of wheat : a review. *European Journal of Agronomy* 10(1), 23–36.
- Rieu, C. and P. Gate (1997). Choix de la date de semis : Quels éléments faut-il prendre en compte? *Perspectives Agricoles* 227, 50–55.
- Takeuchi, I., Q. V. Le, T. Sears, and A. J. Smola (2006). Nonparametric quantile regression. *Journal of Machine Learning Research* 7, 1231–1264.
- Team, R. D. C. (2012). *R : A Language and Environment for Statistical Computing*. Vienna, Austria. ISBN 3-900051-07-0.

Annexes

Annexe A : Description du jeu de données

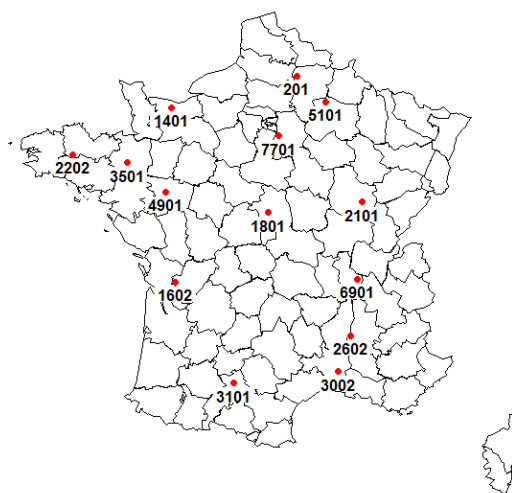


FIGURE 23 – Localisation des 14 stations du jeu de données. Ces stations correspondent à des stations de Météo France pour lesquelles Arvalis a accès aux données météorologiques journalières depuis 1956.

Variable phénoclimatique	Non gel 0	Gel 1	Nombre total
Epi1cm.Semis.nTminiinf12	11135	2193	13328
Epi1cm.Semis.nTminiinf15	12410	918	13328
Epiaison.Epi1cm.nTMiniinf4	12698	630	13328
Epiaison.Epi1cm.nTMiniinf5	13104	224	13328

TABLE 7 – Valeurs prises par les variables phénoclimatiques binaires pour toutes les stations, dates de semis et année de simulation soit 13328 valeurs.

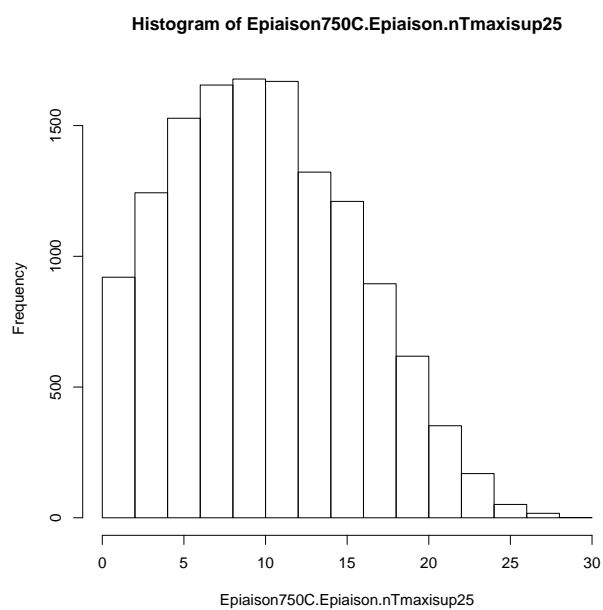
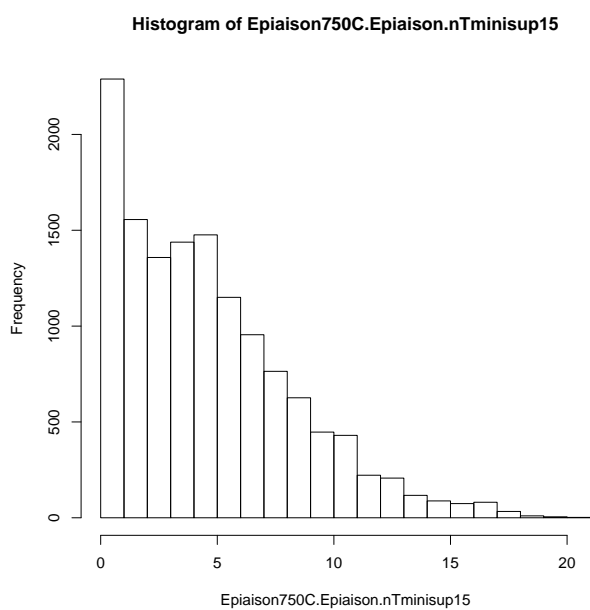
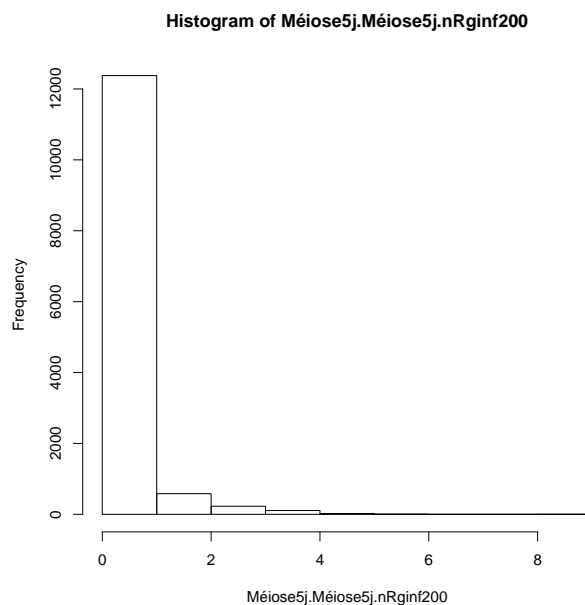
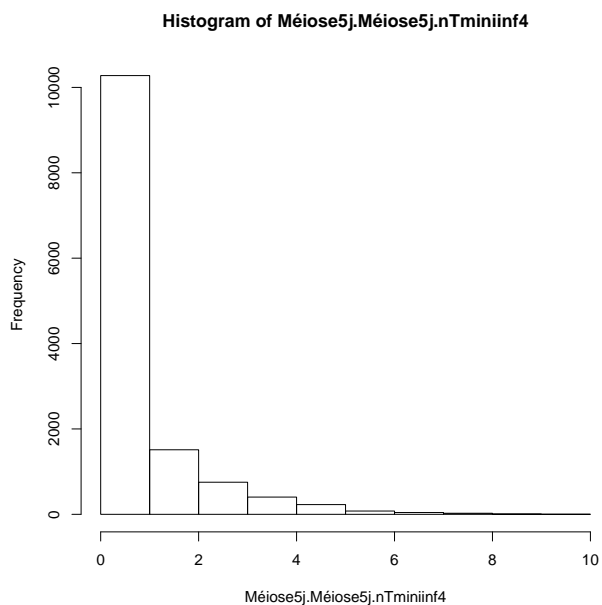


FIGURE 24 – *Distribution des valeurs prises par les variables phénoclimatiques non binaires pour toutes les stations, dates de semis et année de simulation soit 13328 valeurs.*

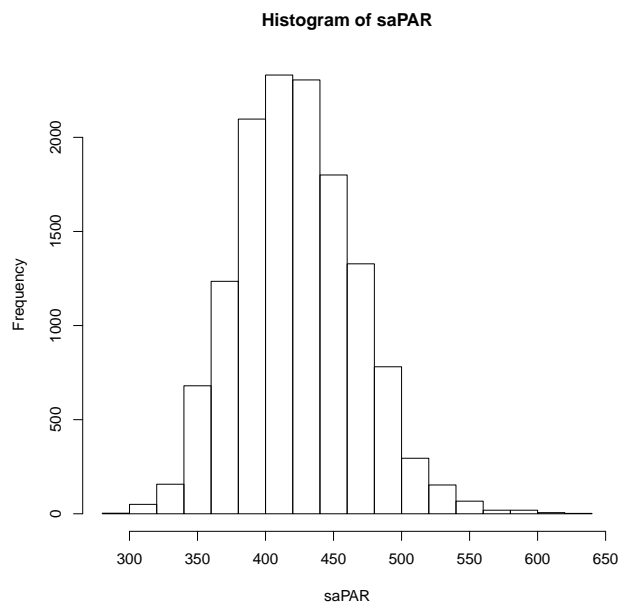
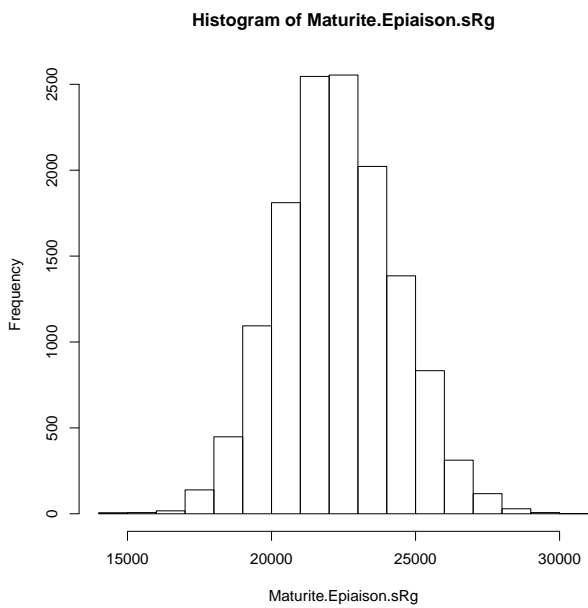
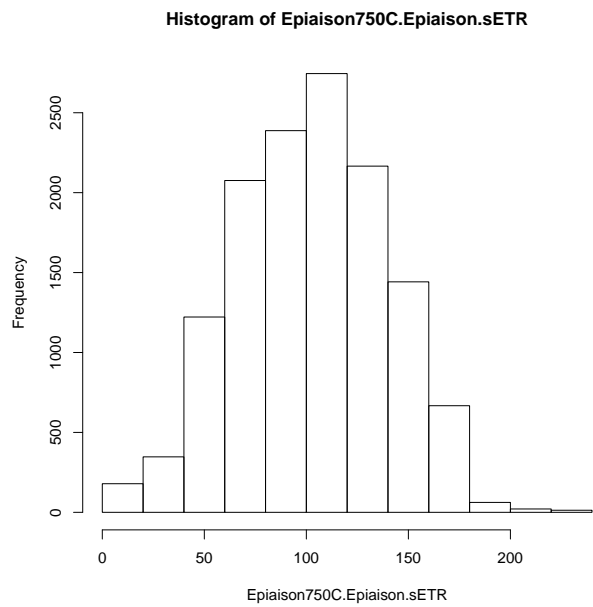
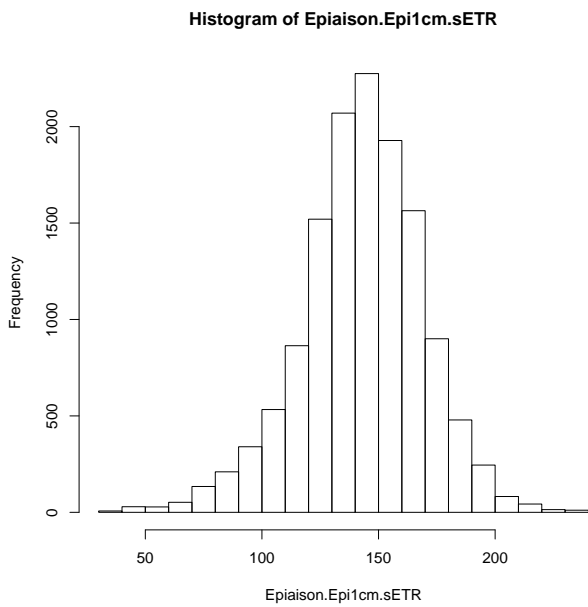


FIGURE 24 – *Distribution des valeurs prises par les variables phénoclimatiques non binaires pour toutes les stations, dates de semis et année de simulation soit 13328 valeurs (suite).*

Annexe B : Quelques algorithmes de calculs de quantiles d'un échantillon

Dans le cas d'une distribution théorique, connaissant la fonction de répartition, on a une valeur précise et unique de quantile. Dans le cas d'un échantillon, on ordonne ses valeurs. Par exemple, pour un échantillon de 10 valeurs x_i ordonnées, la médiane se situe dans $[x_5; x_6]$. Il existe ensuite différentes méthodes pour l'estimer. Notamment, l'estimation des quantiles se peut faire par interpolation linéaire. Ci-dessous, des exemples d'interpolation linéaire (Hyndman and Fan, 1996). Les noms des méthodes sont ceux utilisés dans la publication de Hyndman et Fan ainsi que dans l'aide de la fonction quantile du logiciel R (Team, 2012). La méthode de type 7 est celle utilisée par Arvalis quelle que soit la nature de la variable (discrète ou continue).

Notons N la taille de l'échantillon et p la probabilité associée au quantile Q_p recherché ($p = 0.5$ dans le cas de la médiane). De manière générale, soit g une fonction de poids, Q_p est la moyenne pondérée telle que :

$$Q_p = (1 - g) \cdot x_j + g \cdot x_{j+1}$$

Quantile de type 2 On cherche tout d'abord j , la partie entière de Np .

- Si Np est entier alors $g = \frac{1}{2}$ et le quantile Q_p est la moyenne des x_j et x_{j+1} .
- Sinon $g = 0$, $Q_p = x_j$

Dans le cas de l'exemple précédent, la médiane est la moyenne de x_5 et x_6 .

Quantile de type 7 Ici, j est la partie entière de $1 + (N - 1) \cdot p$ et g sa partie fractionnaire.

- Si $(N - 1)p$ est entier alors $g = 0$ d'où le quantile $Q_p = x_j$.
- Sinon, $Q_p = (1 - g) \cdot x_j + g \cdot x_{j+1}$

Quantile de type 8 j est la partie entière de $(N + \frac{1}{3}) \cdot p + \frac{1}{3}$ et g est sa partie fractionnaire.

- Si $(N + \frac{1}{3}) \cdot p + \frac{1}{3}$ est entier alors $g = 0$ d'où le quantile $Q_p = x_j$.
- Sinon, $Q_p = (1 - g) \cdot x_j + g \cdot x_{j+1}$

Exemple : Soit l'échantillon $\{1; 1; 3; 9\}$ dont on cherche le troisième quartile i.e. $p = 0.75$. Calcul de type 2 : $4 \times 0.75 = 3$ donc $Q_{0.75} = x_3 = 6$

Calcul de type 7 : $1 + 3 \times 0.75 = 3.25$ donc $Q_{0.75} = (1 - 0.25) \times 3 + 0.75 \times 9 = 4.5$

Calcul de type 8 : $Q_{0.75} = 6.5$

Annexe C : Table de la loi normale centrée réduite $\mathcal{N}(0; 1)$

t	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

TABLE 8 – Table de la loi normale centrée réduite $\mathcal{N}(0; 1)$

Annexe D : Principe de la régression quantile linéaire

Notations : y_i une observation de la variable d'intérêt, \hat{y}_i son estimation par le modèle, n le nombre d'observations, x_i l'ensemble des valeurs des variables explicatives associées à y_i , p la probabilité associée au quantile étudié.

La régression linéaire classique permet d'étudier la moyenne de la variable. En effet, les paramètres du modèle sont estimés en minimisant le critère des moindres carrés $\sum_{i=1}^n (y_i - \hat{y}_i)^2$. La régression quantile vise quant à elle à étudier la distribution de la variable d'intérêt en fonction de variables explicatives. On cherche alors le quantile conditionnel d'une variable Y par rapport à une (ou plusieurs) variable explicative X . Pour cela, le quantile Q_p d'un échantillon de n valeurs y_i est défini tel que Q_p minimise :

$$(p-1) \sum_{y_i < Q_p} (y_i - Q_p) + p \sum_{y_i \geq Q_p} (y_i - Q_p)$$

On a donc une pondération asymétrique (Takeuchi et al., 2006; Koenker and Hallock, 2001).

Le quantile de Y conditionnel à la variable explicative X suit le modèle :

$$Q_{p,i}(x_i) = \beta_0 + \beta_1 x_i = f(x_i, \beta)$$

On estime ensuite les paramètres β_0 et β_1 de ce modèle en minimisant :

$$(p-1) \sum_{y_i < \hat{Q}_{p,i}(x_i)} (y_i - \hat{Q}_{p,i}(x_i)) + p \sum_{y_i \geq \hat{Q}_{p,i}(x_i)} (y_i - \hat{Q}_{p,i}(x_i))$$

En posant ρ_p la fonction de perte asymétrique telle que :

$$\rho_p(x) = \begin{cases} px & \text{si } x \geq 0 \\ (p-1)x & \text{si } x < 0 \end{cases}$$

L'estimation des paramètres du modèle minimise alors $\sum_{i=1}^n \rho_p(y_i - f(\beta, x_i))$.

Annexe E : Distribution et variance des $D8_{i,n}$

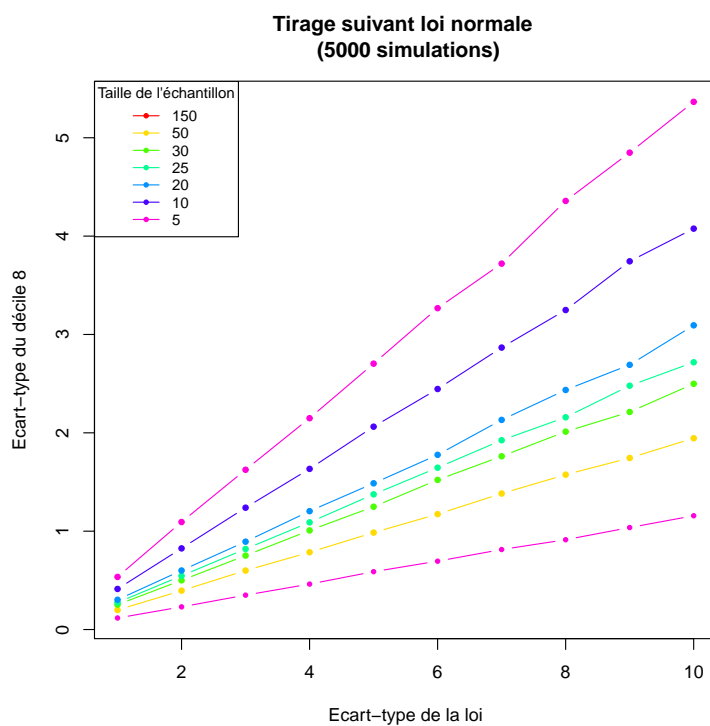


FIGURE 25 – Écart-type des $D8_{i,n}$ suivant la taille de l'échantillon et la variance de la loi Normale génératrice.

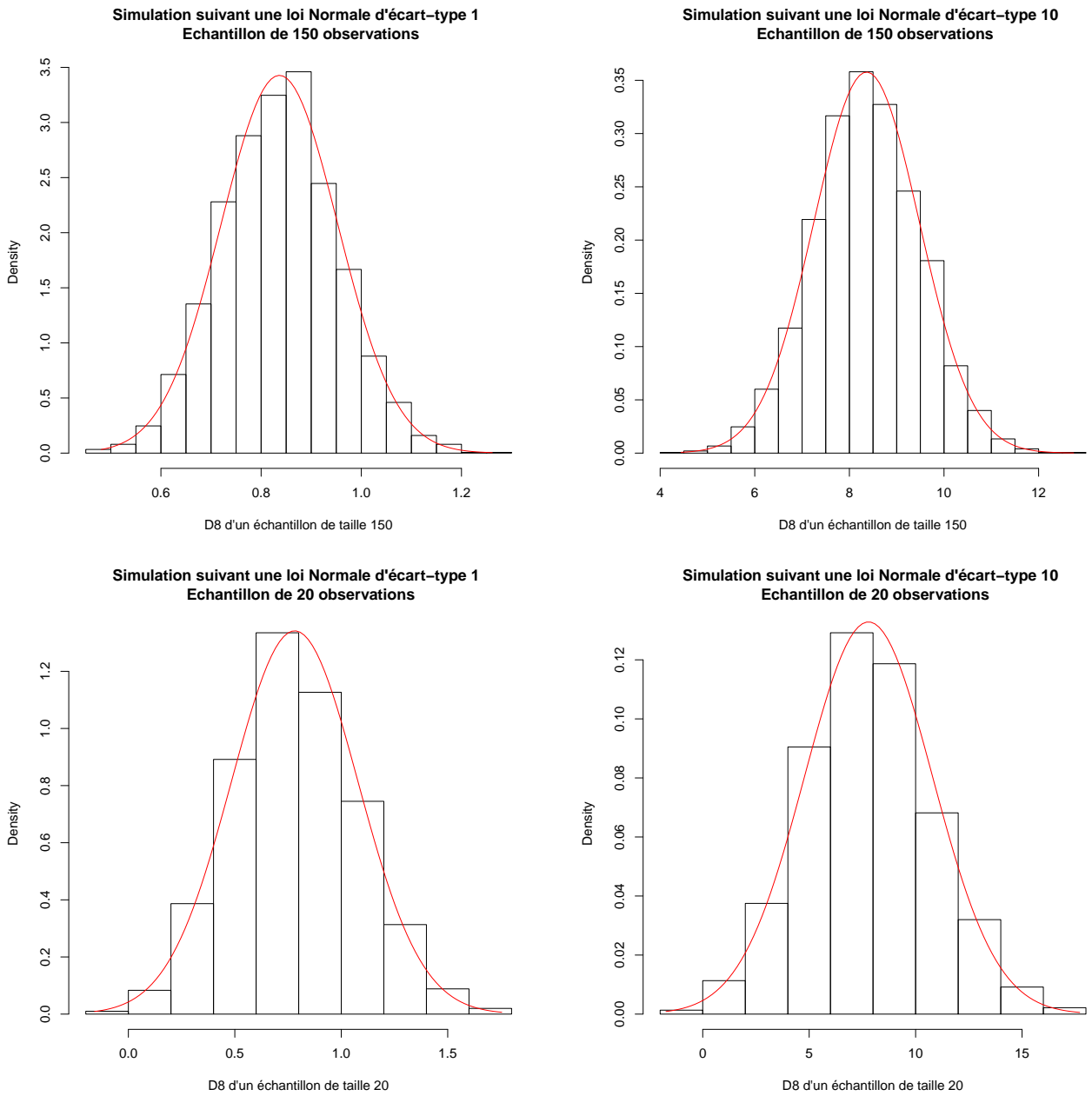


FIGURE 26 – Distribution des $D8_{i,n}$ calculés sur des échantillons de différentes tailles et issus de loi Normale de différentes variances.

Annexe F : Test de Student de comparaison des moyennes $D\hat{8}_n$ aux $D8_{th}$

tailleEch	sdLoi	p.value	tailleEch	sdLoi	p.value	tailleEch	sdLoi	p.value
5	1	7.8e-107	5	6	2.5e-96	5	11	2.6e-91
10	1	1.2e-53	10	6	3.1e-43	10	11	1.3e-61
20	1	2e-28	20	6	4.8e-37	20	11	3.1e-33
25	1	2.5e-26	25	6	1.2e-22	25	11	1.3e-27
30	1	6.8e-19	30	6	1.4e-22	30	11	1.3e-27
50	1	3.6e-10	50	6	4.3e-11	50	11	2.8e-17
150	1	0.0082	150	6	1.3e-08	150	11	0.0052
5	2	2.2e-93	5	7	1.5e-100	5	12	8.5e-100
10	2	1.1e-67	10	7	3.1e-60	10	12	6.1e-49
20	2	2.9e-29	20	7	1.6e-37	20	12	5.5e-26
25	2	6.1e-24	25	7	1.4e-27	25	12	2.1e-17
30	2	1e-22	30	7	7.2e-16	30	12	2.6e-19
50	2	5.3e-11	50	7	9.1e-09	50	12	5.6e-14
150	2	1.5e-06	150	7	0.00043	150	12	0.00024
5	3	7.6e-99	5	8	7e-86	5	13	2.3e-101
10	3	1.5e-58	10	8	4.5e-43	10	13	1.2e-51
20	3	2.4e-20	20	8	3.6e-38	20	13	1e-24
25	3	1.1e-16	25	8	1.4e-25	25	13	2.8e-24
30	3	5.1e-18	30	8	7.6e-17	30	13	4.9e-18
50	3	2.5e-13	50	8	4e-20	50	13	1.8e-16
150	3	0.067	150	8	7.9e-08	150	13	5.3e-05
5	4	1.7e-111	5	9	9.1e-98	5	14	3.7e-101
10	4	7.8e-56	10	9	1.7e-60	10	14	2.3e-54
20	4	3.4e-28	20	9	5.2e-27	20	14	2.1e-33
25	4	1.8e-27	25	9	3.6e-29	25	14	8.3e-23
30	4	1.2e-14	30	9	9.7e-25	30	14	2.7e-19
50	4	1.4e-13	50	9	1.5e-16	50	14	1.2e-11
150	4	0.0012	150	9	6e-05	150	14	1.8e-07
5	5	2.5e-109	5	10	2.3e-100	5	15	5.7e-92
10	5	1.9e-58	10	10	5.9e-60	10	15	2.2e-51
20	5	1.8e-25	20	10	4.9e-30	20	15	8.3e-21
25	5	2.2e-18	25	10	2.9e-29	25	15	2.2e-22
30	5	7.2e-23	30	10	4.1e-17	30	15	7.9e-20
50	5	1.2e-11	50	10	1.8e-11	50	15	2.2e-10
150	5	0.083	150	10	0.0056	150	15	5.9e-07

TABLE 9 – Test de Student de comparaison des moyennes des déciles 8 calculés sur de petits échantillons aux déciles 8 théoriques des distributions correspondantes. Les hypothèses de ce test sont $H_0 : D\hat{8}_n$ est égal à $D8_{th}$ et $H_1 : D\hat{8}_n$ est inférieur à $D8_{th}$

Annexe G : Analyse du modèle du biais $\Delta D8$

Le modèle retenu est :

$$\Delta D8(x, y) = \alpha_0 + \alpha_1 \cdot \frac{x}{y} + \alpha_2 \cdot x + \alpha_3 \cdot x^2 + \alpha_4 \cdot y + \alpha_5 \cdot x \cdot y + \varepsilon$$

avec :

- x = écart-type de la loi Normale génératrice
- y = taille de l'échantillon
- résidus du modèle $\varepsilon \hookrightarrow \mathcal{N}(0, 1)$ par hypothèse

Les résultats présentés dans cet annexe sont issus du modèle précédent ajusté sur les données de simulation.

	Estimate	Std.Error	t.value	p.value
Intercept	$-4.675e-02$	$1.737e-02$	-2.691	0.008029
div	$-9.844e-01$	$1.065e-02$	-92.429	$< 2e-16$
$I(\text{tailleEch}^2)$	$-1.650e-05$	$4.211e-06$	-3.918	0.000142
tailleEch	$2.699e-03$	$7.105e-04$	3.799	0.000220
sdLoi	$-1.368e-02$	$1.471e-03$	-9.302	$3.47e-16$
tailleEch : sdLoi	$8.413e-05$	$1.885e-05$	4.462	$1.70e-05$

TABLE 10 – Estimation des paramètres du modèle du biais

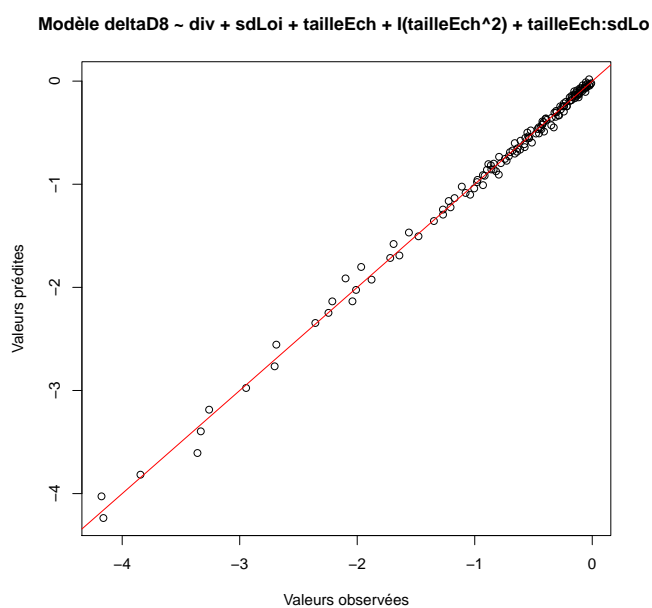


FIGURE 27 – Ajustement du modèle du biais $\Delta D8$

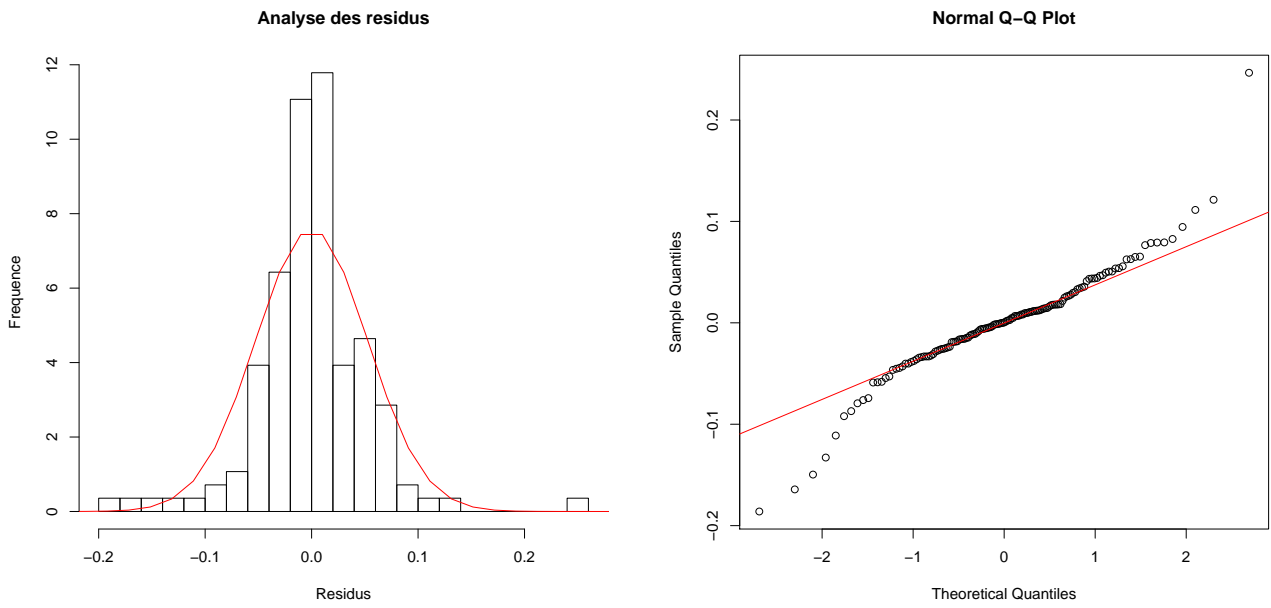


FIGURE 28 – Vérification de l'hypothèse de normalité des résidus du modèle du biais $\Delta D8$

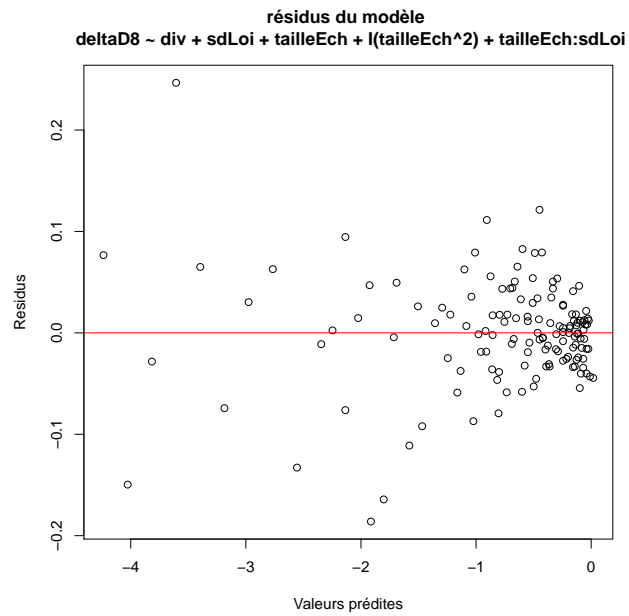


FIGURE 29 – Vérification de l'hypothèse d'homostédasticité des résidus du modèle du biais $\Delta D8$

Annexe H : Dates de semis préconisées historiquement

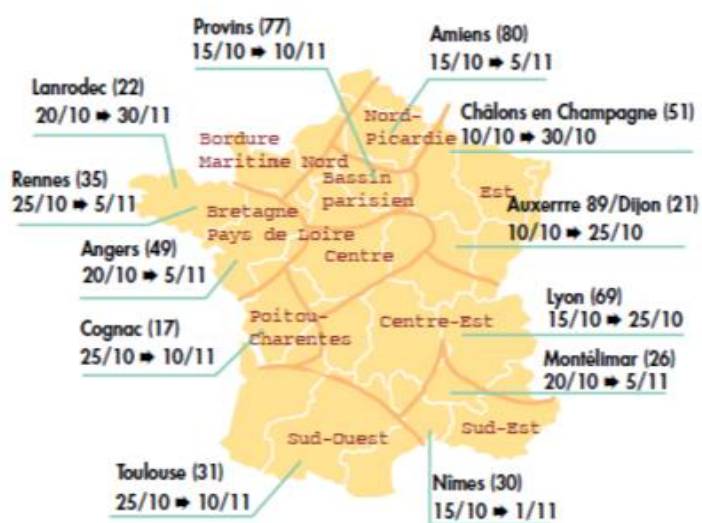


FIGURE 30 – Plages de semis préconisées historiquement d'une variété de céréales de précocité type Soissons. Source (Rieu and Gate, 1997).



Diplôme : Ingénieur Agronome
Spécialité : Statistique Appliquée
Enseignant référent : François HUSSON

Auteur : Marion GAUTHIER

Date de naissance : 23/03/89

Nb pages : 50

Annexes : 8

Année de soutenance : 2012

Organisme d'accueil :

ARVALIS – Institut du végétal

IBP – Université Paris Sud

Rue du Noetzlin – Bât. 630

91405 Orsay Cedex

Maîtres de stage : David GOUACHE

et Jean-Charles DESWARTE

Titre : Repenser les méthodes associées au choix des dates de semis du blé tendre

Title : To think over the methods leading to the choice of winter bread wheat sowing dates

Résumé

Le choix des dates de semis est un enjeu majeur car il détermine l'exposition de la culture aux aléas climatiques. Arvalis - Institut du végétal cherche à préconiser les dates de semis qui permettent de minimiser des risques ou maximiser la satisfaction des besoins phénoclimatiques i.e. qui interviennent à un stade de sensibilité de la plante. L'objectif de ce mémoire est de repenser la succession de méthodes associées au choix des dates de semis, celles-ci ayant des lacunes ou bien étant inexistantes. Nous proposons donc une méthode alternative pour tenir compte du changement climatique ainsi qu'une méthode qui permet de hiérarchiser les dates de semis en fonction d'un niveau de stress associé. Il serait cependant nécessaire de réaliser une étude préliminaire pour mieux définir les variables phénoclimatiques à utiliser et de prendre en compte des variables autres que physiologiques.

Abstract

The choice of sowing dates is a real challenge because it determines which minimize the climatic risks and maximize the needs of the crop which are described by phenoclimatic variables (climatic variables which occur when the plant is sensitive to it). The aim of this study is to think over the methods leading to the choice of winter bread wheat sowing dates. These methods are either unsuitable, either inexistent. We propose an alternative method which deals with climatic change and we built a new methodology which is able to sort sowing dates according to a level a stress for the crop. We still have to improve on the choice of the phenoclimatic variables and to use other variables to define clearly the best sowing dates.

Mots-clés : dates de semis, blé tendre, quantile, risques phénoclimatiques, changement climatique

Key Words : sowing dates, bread wheat, quantile, phenoclimatic risks, climatic change