THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (INP Toulouse)

Discipline ou spécialité :

Agrosystèmes, Écosystèmes et Environnement

Présentée et soutenue par :

Mme FETY ANDRIANASOLO

le vendredi 14 novembre 2014

Titre :

MODELISATION STATISTIQUE ET DYNAMIQUE DE LA COMPOSITION DE LA GRAINE DE TOURNESOL (HELIANTHUS ANNUUS L.) SOUS L'INFLUENCE DES FACTEURS AGRONOMIQUES ET ENVIRONNEMENTAUX

Ecole doctorale :

Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries (SEVAB)

Unité de recherche :

AGroécologie, Innovations, TeRritoires - AGIR

Directeur(s) de Thèse :

M. PHILIPPE DEBAEKE

M. PIERRE MAURY

Rapporteurs :

M. FRANCOIS GASTAL, INRA LUSIGNAN

M. PIERRE MARTRE, INRA CLERMONT FERRAND

Membre(s) du jury :

M. DAVID MAKOWSKI, INRA VERSAILLES GRIGNON, Président

M. ANDRÉ MERRIEN, CETIOM, Membre

M. PHILIPPE DEBAEKE, INRA TOULOUSE, Membre

M. PIERRE MAURY, INP TOULOUSE, Membre

Remerciements

Me voici arrivée à la fin de ce travail de thèse d'un peu plus de trois ans pendant lesquels j'ai beaucoup appris (la modélisation, maintenant j'adore ça) mais également été amenée à inter « agir » (ha ! ha !) avec de nombreuses personnes qui m'ont aidée à tenir le cap, à leur façon, et que j'aimerais chaleureusement remercier ici.

C'est avec émotion que j'exprime ma profonde reconnaissance envers mes directeurs de thèse, Philippe Debaeke et Pierre Maury, sans qui je n'aurais tout simplement pas pu réaliser cette thèse. Merci Philippe pour ta bienveillance et ta disponibilité, dans les moments calmes comme dans les moments de rush. Merci Pierre, pour ton soutien et ton écoute. Philippe, Pierre, nous nous connaissons depuis maintenant plus de 5 ans, et ce fut un honneur et un plaisir immense de travailler avec vous.

Merci aux membres de mon jury de thèse : François Gastal, Pierre Martre, David Makowski et André Merrien d'avoir accepté d'évaluer mon travail.

Je remercie également l'ANRT (contrat CIFRE N° 2010/1467 et projet SUNRISE 2012-19) ainsi que le CETIOM qui ont accepté de financer mon projet de thèse. Merci Luc Champolivier, pour ta disponibilité et ta bonne humeur ; merci d'avoir apporté ton regard critique sur une bonne partie de mon travail.

J'adresse mes remerciements à Pierre Casadebaig, qui a été présent du début à la fin, dans les réunions de travail, préparation d'articles scientifiques, et dont l'aide m'a été indispensable notamment dans le cadre de l'utilisation du modèle SUNFLO. Merci pour ta sympathie et ta disponibilité !

En fin de thèse tout s'accélère et le moindre petit obstacle devient un mur infranchissable ; dans ces moments où la patience tend à céder sa place à la panique, j'ai pu compter sur l'aide d'une personne pour finaliser mon modèle et atteindre mes objectifs : merci François Brun pour ton avis éclairé et ta patience, qui ont permis de gagner du temps et de l'énergie ! Merci d'avoir donné des cours de modélisation (Paris, 2012) sans lesquels je n'aurais jamais pu construire mon modèle.

Merci à Elie Maza d'avoir été disponible lors de mes questionnements d'ordre statistique.

Je remercie les membres non cités de mon comité de thèse : Luis Aguirrezábal, Jean-Michel Allirand, Monique Berger.

J'ai une pensée amicale pour Mercedes Echarte, enseignant chercheur à Balcarce (Argentine) et qui a fait un séjour d'un mois à VASCO en 2012. Merci Mercedes, pour nos échanges constructifs, pour ta douceur et ta disponibilité.

Merci aux équipes techniques sans qui mes expérimentations au champ et en serre n'auraient pu être mises en place : côté INRA, Laure Lagarrigue, Michel Labarrère, Didier Rafaillac, Nicolas Blanchet. Côté CETIOM, Richard Segura, Pascal Fauvin, ainsi que le personnel du laboratoire d'Ardon avec qui j'ai beaucoup interagit pour les analyses d'huile et de protéines des graines : Mohammed Krouti et Véronique Garnon. Merci à tous pour votre professionnalisme et votre gentillesse!

Je remercie Eric Justes qui a fait preuve de beaucoup d'empathie à mon égard lors des moments de doutes.

Merci à Hélène Tribouillois (j'ai adoré nos séances de bodyjam), Bochra Kammoun (merci pour ton écoute, ça fait du bien de se soutenir entre doctorantes) et Velonoromanalintantely Rabiafaranjato (Fara pour les intimes), pour nos sorties détente après nos journées difficiles de doctorantes. Merci à Mamy pour son amour inconditionnel, merci d'avoir vécu avec moi tous les jours les hauts et les bas de ma vie de thésarde, même à distance.

Merci à mes stagiaires Etienne et Zakia, ainsi qu'à Yannick (Dakaud pour les intimes), qui m'a aidé le temps d'un été pour des mesures et des prélèvements sur le terrain. Dakaud et Zakia, vous êtes des perles ! Ce fut une super ambiance au champ avec vous! Merci pour votre bonne humeur et votre amitié !

Enfin, et avec beaucoup d'émotion, j'ai une pensée attendrie pour mes parents, si loin là-bas, mais qui n'ont jamais arrêté de manifester leur soutien et leur amour. Merci Papa et Maman.

TABLE DES MATIERES

Remerciements 1

1. Le tournesol : contexte mondial, européen, français et débouchés principaux 6
2. La filière tournesol en France: acteurs, débouchés actuels et potentiels 8
3. En France, une régularité et une augmentation nécessaires des rendements est attendue par la filière 11

II. Synthèse bibliographique : modéliser l'élaboration de la qualité chez le tournesol ... 14

1. Avant-propos 14
2. L'élaboration de l'huile et des protéines chez le tournesol 14
   - 2.1. L'huile 14
   - 2.2. Accumulation des protéines de réserve 17
   - 2.3. Antagonisme entre accumulation d'huile et de protéines 17
   - 2.4. Déterminisme et facteurs limitants des teneurs en huile et en protéines chez le tournesol 18
3. La modélisation de la qualité 26
   - 3.1. Définition, objectifs et performances d'un modèle 26
   - 3.2. La modélisation de la qualité chez les oléagineux 31

III. Enoncé de la problématique : objectifs, démarche de la thèse et plan du manuscrit 38

1. Hypothèses et objectifs de la thèse 39
2. Questions de recherche 39
3. Stratégie de recherche 40
4. Plan du manuscrit 43

Chapitre I: Prediction of sunflower grain oil concentration as a function of variety, crop management and environment using statistical models. 46

1. Introduction 48
2. Materials and methods 52
   - 2.1. Dataset collection 52
   - 2.2. Simulation of oil concentration predictors 53
   - 2.3. Choice of putative predictors for oil concentration model 54
   - 2.4. Filtering USM and predictors 56
   - 2.5. Statistical models building 56
   - 2.6. Statistical models evaluation and diagnosis 58
3. Results 59
   - 3.1. Dataset diversity: cropping conditions and observed oil concentrations variability 59
   - 3.2. Yield threshold 61

3.3. Statistical models building 62
3.4. Comparative performances of statistical models 65
3.5. Patterns in response to management practices 68
4. Discussion 69
5. Conclusions 77
Bilan du chapitre I 78

Chapitre II: Analyzing oil and proteins accumulation through source and sink framework establishment in sunflower achenes 79

1. Introduction 82
2. Materials and methods 84
2.1. Site and experimental design 84
2.2. Simulation of water balance with SUNFLO crop model 85
2.3. Dynamic monitoring and model adjustment 85
2.4. Sink and source indicators at harvest 88
2.5. Statistical analysis 89
3. Results 89
3.1. Weather conditions, crop phenology and status 89
3.2. Year and nitrogen effects on source and sink indicators at harvest 91
3.3. Nitrogen, density and genotype effects on source and sink indicators at harvest 93
3.4. Analysis of dynamics of source and sink organs 95
3.5. Dynamic framework of source-sink relationships and orders of priority 101
4. Discussion 105
5. Conclusions 110
6. Acknowledgments 111
Bilan du chapitre II 112

Chapitre III: Effects of growth stage and leaf ageing on transpiration and photosynthesis response to water stress in sunflower 113

1. Introduction 116
2. Materials and methods 118
2.1. Experiments design 118
2.2. Modelling transpiration and photosynthesis response to soil water deficit and statistical analysis 120
3. Results 121
3.1. Vapor pressure deficit (VPD), transpiration and photosynthesis ranges 121
3.2. Regulation of leaf transpiration and photosynthesis in vegetative and reproductive growth stages 122

3.3. Regulation of leaf transpiration depending on leaf age 125
3.4. Upscaling from individual leaf to whole plant transpiration 127
4. Discussion 129
5. Conclusions 133
6. Acknowledgments 133
Bilan du chapitre III 134

Chapitre IV: A source-sink based dynamic model for simulating oil and proteins accumulation in sunflower grains 135

1. Introduction 138
2. Materials and methods 141
2.1. Model overview 141
2.2. Model formalisms description 143
2.3. Inputs and outputs 151
2.4. Calibration dataset and parameters 151
2.5. Evaluation dataset 152
2.6. Parameterization and evaluation tools 154
2.7. Sensitivity analysis 154
3. Results 159
3.1. Parameterization results 159
3.2. Evaluation results 160
3.3. Sensitivity analysis results 166
4. Discussion 168
5. Conclusions 172
Bilan du chapitre IV 173

Discussion générale, conclusions et perspectives 174

1. La modélisation statistique pour hiérarchiser et trier les facteurs importants de la teneur en huile175
2. Des expérimentations au champ et en serre pour décortiquer les effets des facteurs sur les déterminants des teneurs en huile et protéines 178
2.1. L'analyse des relations source-puits et des effets génotype, azote et densité 178
2.2. L'analyse de la réponse à la contrainte hydrique et les effets stade et génotype 179
3. La modélisation dynamique basée sur les relations source-puits 182

Références bibliographiques 187

Annexe 211

I. Introduction

1. Le tournesol : contexte mondial, européen, français et débouchés principaux

Le tournesol figure parmi les plantes les plus connues dans le monde, et est facilement reconnaissable avec ses fleurs jaunes et sa grande taille. Il est cultivé essentiellement pour ses graines, afin d'en extraire l'huile destinée à l'alimentation humaine (huile de table) et à d'autres débouchés non-alimentaires (biocarburants et oléochimie, Jouffret et al., 2011). En effet, l'akène ou graine de tournesol se distingue des autres graines oléagineuses par une teneur en huile élevée (44% en moyenne, contre 42% pour le colza et 18% pour le soja). Elle contient 18% de protéines, 15% de cellulose, 9% d'eau (PROLEA, 2009) et 14% de carbohydrates et minéraux (Roche, 2005). L'akène peut être séparé en ses deux constituants : la coque ou péricarpe, représentant entre 20 à 40% du poids de l'akène (Connor etHall, 1997 ; Lindström et al., 2007), et l'amande. La coque est une enveloppe ligno-cellulosique contenant peu de protéines et peu ou pas d'huile (Knowles, 1978) ; elle est constituée de plusieurs couches dermiques dont les plus extérieures sont sclérifiées et où le dépôt de phytomélanine donne sa couleur noire aux graines (Lindström et al., 2007). L'amande est composée d'une paroi, d'un endosperme et d'un embryon où sont accumulés 95 à 97% de l'huile de la graine (Izquierdo et al., 2008) et des protéines de réserve.

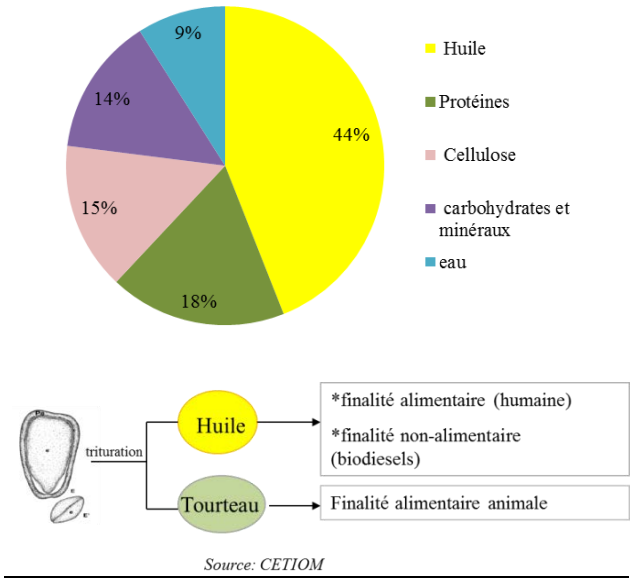


Fig.1. Représentation de la composition de la graine de tournesol et des débouchés principaux (sources : PROLEA et CETIOM, avec modifications personnelles)

L'huile est obtenue par un procédé de trituration des graines ; le reste des graines issues de la trituration constituent les tourteaux, résidus composés de fibres (cellulose) et de protéines et utilisés dans l'alimentation animale, notamment pour l'alimentation des bovins viande, lapins, truies, et poules pondeuses (Jouffret et al., 2011).

Ces deux principaux débouchés que sont l'huile et les tourteaux constituent à eux seuls plus de 90% de l'utilisation des graines de tournesol. Le reste concerne la consommation de tournesol de bouche et celle pour l'oisellerie (Borredon et al., 2011).

L'huile de tournesol se place en 4ème position mondiale (8%, sur une production d'huile végétale totale de 154 Mt en 2011) après celle du palme (33%), du soja (27%) et du colza (15%). Les plus grands pays producteurs sont les USA pour le soja (23% de la production mondiale en 2011), l'Indonésie pour le palme (48%), et l'UE27 pour le colza (39%). Côté tournesol, les plus grands pays producteurs d'huile sont l'ex-URSS (47%, 6 Mt), l'UE27 (21%, 2,7 Mt) et l'Argentine (11%, 1,5 Mt).

Selon les prédictions de la FAO (FAO, 2011), la demande en huile et en tourteaux continuera d'augmenter au cours des 7 prochaines années (horizon 2020/2021), face à une production d'oléagineux qui risque d'être déficitaire. Le développement de biocarburants à base d'huiles végétales a relancé l'intérêt des recherches sur le tournesol en 2005-2010, dans un contexte de plus forte tension sur les ressources environnementales (Pilorgé, 2010). Le déséquilibre entre offre et demande sera probablement accentué par le changement climatique (FAO, 2011) comme illustré par l'année culturale 2010/2011, où les rendements mondiaux du soja et du colza ont souffert de conditions climatiques adverses ; la même année, ceux du tournesol ont augmenté, mais les surfaces mondiales qui lui ont été dédiées sont moindres (24 millions d'hectares) par rapport au soja (100 millions d'hectares) et au colza (35 millions d'hectares, OIL WORLD, 2014). D'autre part, il est attendu que la demande mondiale en protéines d'origine animale double d'ici 2050 (FAO, 2014). Cela implique des besoins en protéines végétales considérables pour l'alimentation des animaux d'élevage.

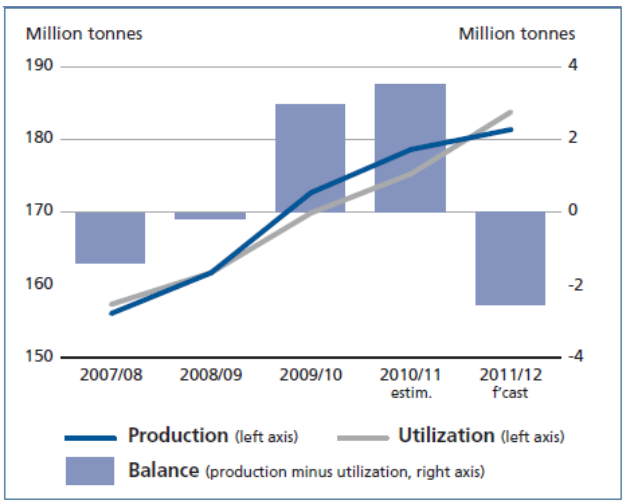


Fig.2 : Evolution de l'équilibre entre offre et demande en huiles végétales entre 2008 et 2012 (source : FAO, 2012)

Les deux principaux tourteaux en concurrence sur le marché des oléagineux sont ceux du soja et du tournesol (Jouffret et al., 2011). L'Ukraine est le plus grand exportateur d'huile et de tourteaux, suivi de l'Argentine et de la Russie (58%, 20% et 15% respectivement). L'UE27 est quant à elle la plus grande importatrice d'huile et de tourteaux (Jouffret et al., 2011), suivie de la Chine et de l'Inde. Les projections de la FAO montrent une augmentation de 2% par an de la consommation en tourteaux dans ces pays en développement (FAO, Agricultural Outlook 2011-2020).

En Europe, l'huile de tournesol se classe en 2ème place (18.5%, 2.6 Mt) après celle du colza (60%, 8 Mt) et du soja (15%, 2.2 Mt) (CETIOM, 2011). La France est le premier pays producteur d'huile de tournesol (671.000 t en 2011), suivi de l'Espagne (462.000 t), l'Italie et la Hongrie (~300.000 t). La France détient les meilleurs rendements moyens de tournesol (~24 quintaux par hectares) à travers le monde. Néanmoins, derrière cette apparente bonne performance se cache une forte variabilité inter-annuelle des performances de la culture.

2. La filière tournesol en France: acteurs, débouchés actuels et potentiels

La notion de « filière » en agronomie implique l'ensemble des participants –producteurs, instituts, organismes, industries- qui contribuent à la production, la transformation et la commercialisation d'un produit agricole (Goldberg, 1968). En France, la filière tournesol est très organisée (Fig.3). La production de tournesol est concentrée dans les régions Midi-

Pyrénées, Poitou-Charentes, Centre et Pays de la Loire (PROLEA, 2009), pour une surface totale de 769.653 hectares en 2013 (21.4 q/ha, Agreste, 2013).

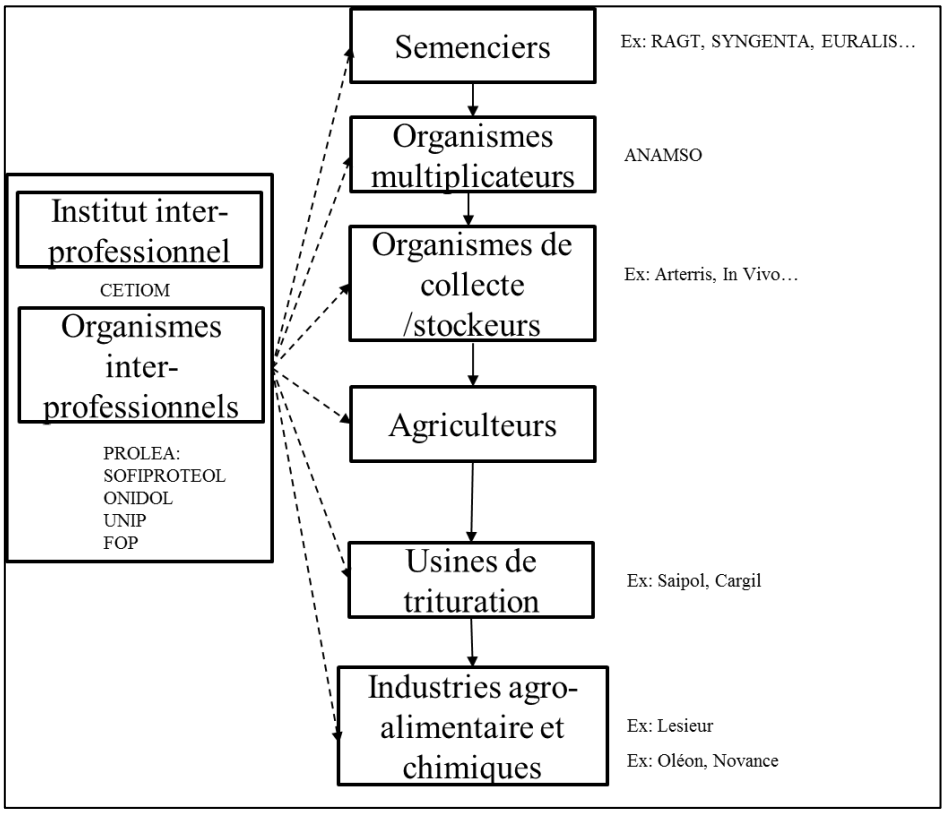


Fig.3. Schéma simplifié d'organisation de la filière tournesol en France. Les flèches en pointillés correspondent aux niveaux d'interaction possible des acteurs de la filière avec l'interprofession.

La notion de « qualité » sera appréhendée différemment par les acteurs selon leur place dans la filière. Pour les acteurs en amont (coopératives, agriculteurs), la qualité correspond à la teneur en huile (normes commerciales à 44% d'huile, 9% d'eau et 2% d'impuretés). En général, c'est l'organisme collecteur qui est soumis aux variations de prix liées à la qualité à l'échelle du silo, charge à lui d'en répercuter les conséquences positives ou négatives sur le producteur. Dans certains cas, la teneur en huile est directement répercutée en bonus/malus (par point d'huile à hauteur de 1.5% du prix) au producteur à partir d'échantillons de graines lors de la livraison. Ce seraient les difficultés techniques d'estimation rapide de la teneur en huile à la benne qui limiteraient le développement des paiements différenciés à la livraison du tournesol. En aval de la filière (triturateurs, industriels), la qualité peut englober plusieurs critères : teneur en huile, teneur en protéines, ou composition de l'huile (acides gras, tocophérols,..).

En France, l'huile est valorisée sous deux formes : l'huile conventionnelle, constituée majoritairement d'acide gras linoléique, qui a une finalité essentiellement alimentaire (huile de table et margarine, 46% des surfaces cultivées). L'huile oléique (seuil de teneur en acide oléique de 82% contractualisé) sert pour les huiles de mélange ou combinées, les huiles destinées à la friture, et surtout en remplacement des acides gras « trans » dans certaines préparations. Une partie minoritaire de l'huile oléique (5 à 8% de la production) est incorporée dans la fabrication de biocarburants (2 à 5% du mélange provient d'huile de tournesol) et dans l'oléochimie (peintures, encres, biolubrifiants) (Jouffret et al., 2011).

Les tourteaux de tournesol contiennent en moyenne 29% de protéines et 25% de cellulose brute (Borredon et al., 2011). Des processus de semi-décorticage, c'est-à-dire de dépouillement mécanique partiel des coques avant la trituration permettent d'augmenter la teneur en protéines des tourteaux. Ce processus a abouti à l'obtention de tourteaux « high pro » à 36% de protéines (en comparaison, le taux de protéines brutes dans le tourteau de colza est de 33.7% et celui du soja à 44%, PROLEA). Pourtant déjà bien implanté en Argentine et dans les autres pays de l'UE, le processus de semi-décorticage sur graine de tournesol n'a été mis en oeuvre que très récemment en France ; depuis 2013, des tourteaux « high pro » sont produits en France à l'usine de Bassens (Bordeaux) (Peyronnet et al., 2014). L'apparition des tourteaux « high pro » s'accompagne d'une diversification des débouchés en alimentation animale : le « low pro » (non décortiqué) est utilisé pour l'alimentation bovine et porcine, tandis que le « high pro » est destiné à la volaille. Les tourteaux peuvent être utilisés à des fins non-alimentaires : base de matériau biodégradable en horticulture, colle végétale à l'eau (Roche, 2005 ; Borredon et al., 2011). Les protéines végétales dans les tourteaux sont isolables pour l'alimentation (farine de tournesol, compléments pour produits carnés) (Roche, 2005).

Enfin, la fraction non-lipidique de l'huile de tournesol présente des composés mineurs à propriétés santé dont l'intérêt ne peut que croître dans les prochaines années. Ainsi, la présence de tocophérols, notamment l'α-tocophérol, confère une propriété vitaminique (Vitamine E) à l'huile de tournesol. La teneur en tocophérols dans l'huile varie de 0.05 à 0.18%. Les stérols ou phytostérols ont un pouvoir hypocholestérolémiant reconnu et une action anti-cancéreuse (Berger et al., 2010). Leur quantité est relativement élevée dans l'huile de tournesol par rapport à d'autres huiles : elle varie de 0.5 à 1% du poids de l'huile. (Fig.4)

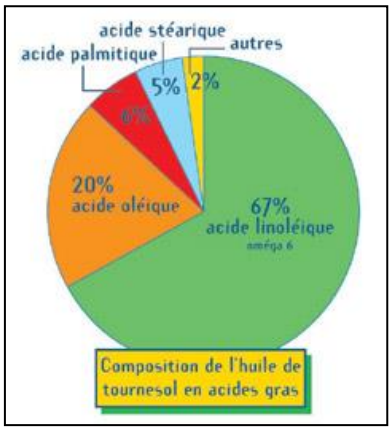


Fig.4 : Représentation de la constitution de l'huile de tournesol classique (source : PROLEA, 2009)

L'ensemble de ces débouchés actuels et potentiels font du tournesol une culture prometteuse et compétitive sur le marché des oléagineux. L'augmentation/stabilisation des teneurs en huile augmenterait l'attractivité de cette culture (plus-value santé et environnementale) d'une part, tandis que l'augmentation/stabilisation des teneurs en protéines en feraient une source non-négligeable de protéines pour l'alimentation animale face au colza et au soja, d'autre part.

Pourtant, la situation actuelle en France pose des questions sur l'avenir de la production de tournesol.

3. En France, une régularité et une augmentation nécessaires des rendements est attendue par la filière

En France, la production d'huile de tournesol a augmenté de 102.000 t entre 2001 et 2011, tandis que celle du colza a connu une hausse de 1,06 Mt entre les deux années (CETIOM, 2011). La production d'huile de soja a peu augmenté (8000 t). L'augmentation de la production d'huile de tournesol s'explique par une légère augmentation des surfaces qui lui sont dédiées (706.000 ha en 2001, 710.000 en 2010, 742.000 en 2011, ~769.000 en 2013 ; CETIOM).

La moyenne des rendements en grains n'a que peu ou pas évolué, et ce depuis 30 ans (Vear et al., 2003), malgré un progrès génétique régulier. En conditions non-limitantes, le rendement potentiel du tournesol a été évalué à 45 quintaux par ha en France (Debaeke, *communication personnelle*). Les facteurs limitant le rendement ont été diagnostiqués (Quere, 2004) : le stress hydrique lié à la sécheresse estivale, les maladies de fin de cycle, des structures de peuplement hétérogènes dans l'espace et dans le temps, des problèmes de structure du sol conduisant à des enracinements de mauvaise qualité et des carences azotées consécutives au

déficit hydrique. A ceux-ci s'ajoutent les effets des interactions entre génotype et environnement (Casadebaig, 2008) : dans un milieu où les ressources sont limitantes, les génotypes s'adaptent de façons différentes (Gallais, 1992b). La seule connaissance du potentiel d'un génotype n'est alors plus suffisante pour pouvoir prédire son comportement.

Les teneurs en huile nationales moyennes, exprimées aux normes, varient selon les années ; elles ont atteint 47.3% en 2013 et ont été au plus bas niveau l'année précédente (43.4%) sur la période 2008 à 2013 (Fig.5). En conditions non-limitantes, les teneurs en huile potentielles peuvent atteindre 55% chez certains génotypes (Champolivier, *communication personnelle*).

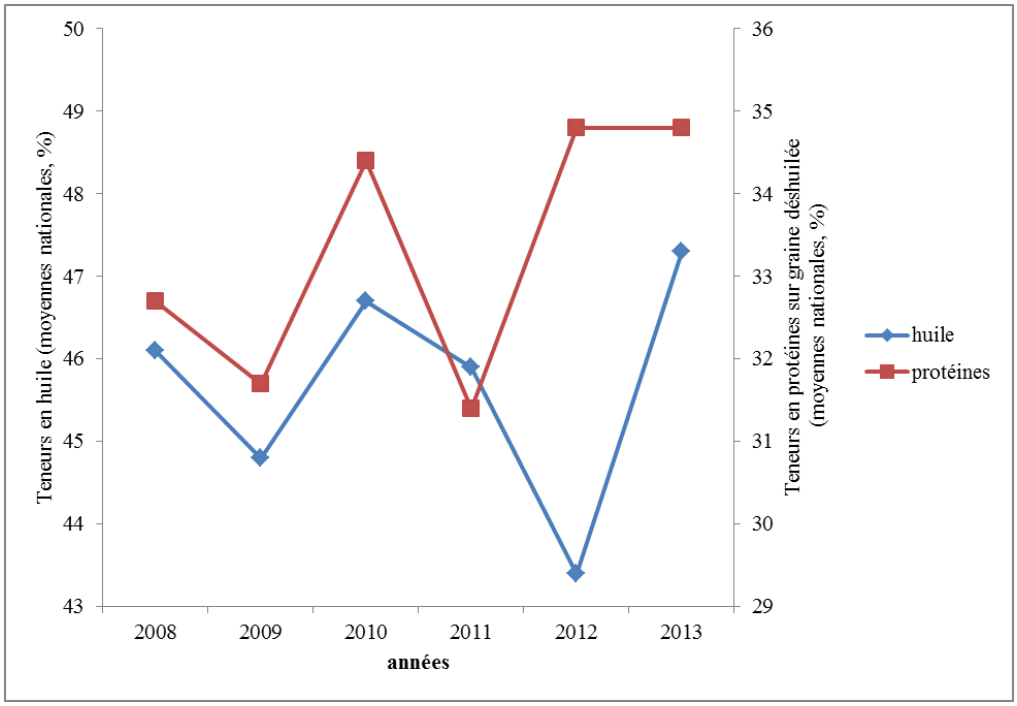


Fig.5 : Evolution des teneurs en huile et en protéines sur graine déshuilée (moyenne nationale, %) (source : CETIOM 2013, représentation modifiée)

Dans le cadre d'une enquête agronomique menée à l'échelle de deux bassins de collecte dans le Sud-Ouest de la France et sur deux génotypes à teneurs en huile contrastées, Champolivier et al. (2011) montrent que la variabilité liée à l'environnement et à la conduite est très supérieure à celle liée à la variété (environ 10 points d'huile contre 5 points d'huile pour la variété, Fig.6). Aussi bien pour le rendement en graines qu'en huile, les génotypes ont démontré une diversité de réponses, résultante de leur adaptation au milieu ; il ne suffit donc plus de connaître le potentiel en huile d'un génotype (fourni lors des phases d'inscription au catalogue) pour connaître sa performance réelle dans un milieu donné.

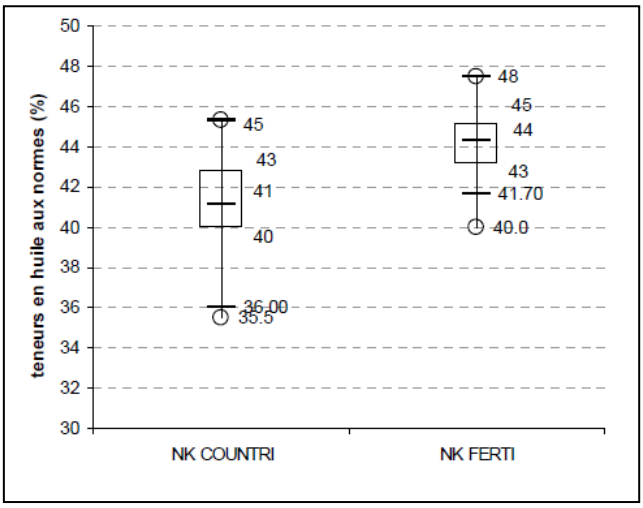


Fig.6 : Comparaison de la variabilité des teneurs en huile entre deux génotypes contrastés dans deux bassins de collecte du Sud-Ouest de la France (source : Champolivier et al., 2011)

Les teneurs en protéines ne sont pas évaluées à l'échelle nationale sur graine entière, mais des teneurs en protéines sur graine déshuilée sont disponibles (CETIOM) : les plus faibles ont été observées en 2009 (31.7%) et les plus fortes en 2012 et 2013 (~35%). (Fig.5). Peu d'études ont porté sur l'évaluation des facteurs limitant la teneur en protéines chez le tournesol ; celle menée par Merrien et al. (1988) montre que les facteurs pédo-climatiques influencent les redistributions azotées et l'absorption d'azote, directement en lien avec la teneur protéique.

La variabilité des teneurs en huile et des teneurs en protéines nous amène à nous interroger sur la meilleure stratégie à adopter afin de stabiliser/augmenter ces teneurs. Il s'agira soit d'augmenter la production, soit le rendenent.

L'augmentation des surfaces a été observée ces dernières années, mais elle est lente; le tournesol souffre de son passé (réforme de la PAC, 1992) où les surfaces ont diminué à cause d'un désintérêt des agriculteurs vis-à-vis de cette plante réputée tolérante mais à moins bonne marge économique que le colza. Elle a été ainsi confinée à des sols superficiels, à réserve hydrique limitée.

Une bonne marge d'amélioration peut être attendue via une optimisation du conseil variétal aux agriculteurs/producteurs. Elle devra tenir compte non seulement du génotype, mais aussi du comportement spécifique d'un génotype donné dans un milieu donné, plus généralement des interactions Génotype x Environnement x Conduite de culture (IGEC). Cela implique au moins 3 conditions : (i) une bonne connaissance des effets séparés du génotype et de l'environnement, (ii) une évaluation des « lieux » d'interaction, et (iii) une capacité à prédire des comportements donnés en fonction du génotype. Dans la suite de ce travail de thèse, nous faisons la démonstration que la modélisation permet de répondre à ces conditions.

II. Synthèse bibliographique : modéliser l'élaboration de la qualité chez le tournesol

1. Avant-propos

L'huile et les protéines sont les principaux débouchés actuels et futurs du tournesol. Afin de mieux prédire les performances de la culture, des modèles existent pour simuler le rendement en graines, mais très peu prennent en compte la qualité. On se limitera ici aux teneurs en huile et en protéines, mais la qualité englobe également les teneurs en composés spécifiques (acides gras, tocophérols,...) Après avoir rappelé les connaissances physiologiques sur l'élaboration de l'huile et des protéines, les modèles de simulation développés pour les oléagineux et le tournesol en particulier seront présentés, avec un focus sur la prise en compte de la qualité.

2. L'élaboration de l'huile et des protéines chez le tournesol

2.1. L'huile

2.1.1. Localisation dans la plante

L'huile est essentiellement localisée dans les graines, même si de petites quantités de lipides (10 à 30 g par kg) peuvent être trouvées dans tous les tissus de la plante, associées aux membranes cellulaires et sub-cellulaires (Connor et Hall, 1997).

2.1.2. Origine

Les acides gras composant l'huile sont stockés sous forme de triglycérides. Ils proviennent de la carboxylation dans le cytosol de l'Acétyl-CoA. Des réactions en chaîne de décarboxylation-réduction catalysées par des ACP synthases permettent l'allongement progressif des acides gras ; après 7 cycles, la dernière molécule formée (stéaoryl-ACP) est désaturée en acide oléique (C18 :1-ACP). Une partie des acides gras hydrolysés sont utilisés

pour la synthèse des phospholipides, tandis que d'autres sont désaturés dans le réticulum endoplasmique pour former des acides gras poly-insaturés (oléique, linoléique). Les triglycérides sont ensuite stockés dans les oléosomes, qui sont des extensions de la paroi du réticulum (Berger et al., 2010, Morot-Gaudry, 2012). Les tournesols à forte teneur en acide oléique sont obtenus grâce à une mutation de la désaturase impliquée dans la transformation de l'acide oléique en acide linoléique (Berger et al., 2010).

2.1.3. Accumulation de l'huile dans les graines

La dynamique d'accumulation de l'huile dans les graines suit une évolution sigmoïde ; la première phase d'accumulation (les 200 premiers degrés jours après floraison (stade R5.1, Izquierdo et al., 2008) ou les 7 à 10 premiers jours après le début de la floraison, Mantese et al., 2006) est lente et correspond à l'incorporation de lipides polaires dans les structures membranaires (vitesse de 0.02 $mg.g^{-1}.°Cjours^{-1}$). Ensuite, la teneur en huile augmente de façon linéaire (vitesse de 0.10 à 0.20 $mg.g^{-1}.°Cjours^{-1}$) pendant 200 à 250 degrés-jours. Comparativement, Goffner et al. (1988) observent que cette phase rapide d'accumulation d'huile a lieu entre 2 et 4 semaines après la mi-floraison. Enfin, la teneur en huile atteint une phase de plateau 30 jours après la fin de la floraison (Champolivier et Merrien, 1996). Le début du plateau marque une diminution de la vitesse d'accumulation de l'huile. Sa vitesse va se calquer sur la vitesse d'accumulation des autres composés de la graine. L'arrêt de l'accumulation d'huile peut avoir lieu en même temps que la stabilisation du poids des graines (atteinte de la maturité physiologique), ou avant (Chervet et Vear, 1989).

Afin d'illustrer la dynamique de la teneur en huile et de mettre en évidence la correspondance entre degrés-jours, jours et semaines après floraison (selon les références trouvées dans la la littérature), nous proposons de schématiser une dynamique moyenne « classique » de la teneur en huile à partir des durées et vitesses fournies par Izquierdo et al. (2008). Nous faisons ainsi l'hypothèse que pour une journée typique de la floraison à la maturation, la température journalière est de 20°C. On acquiert donc 14 degrés-jours par jour. De plus, nous appliquons la règle selon laquelle en dessous de 200 °C jours, la vitesse d'accumulation correspond à celle de la phase latente, et qu'après 450°C jours, la vitesse est à 0. La comparaison des différentes sources bibliographiques permet de montrer qu'il y a bien concordance entre les références, qu'elles se rapportent à des degrés-jours, des nombres de jours ou de semaines pour décrire l'accumulation de l'huile après floraison (Fig.8).

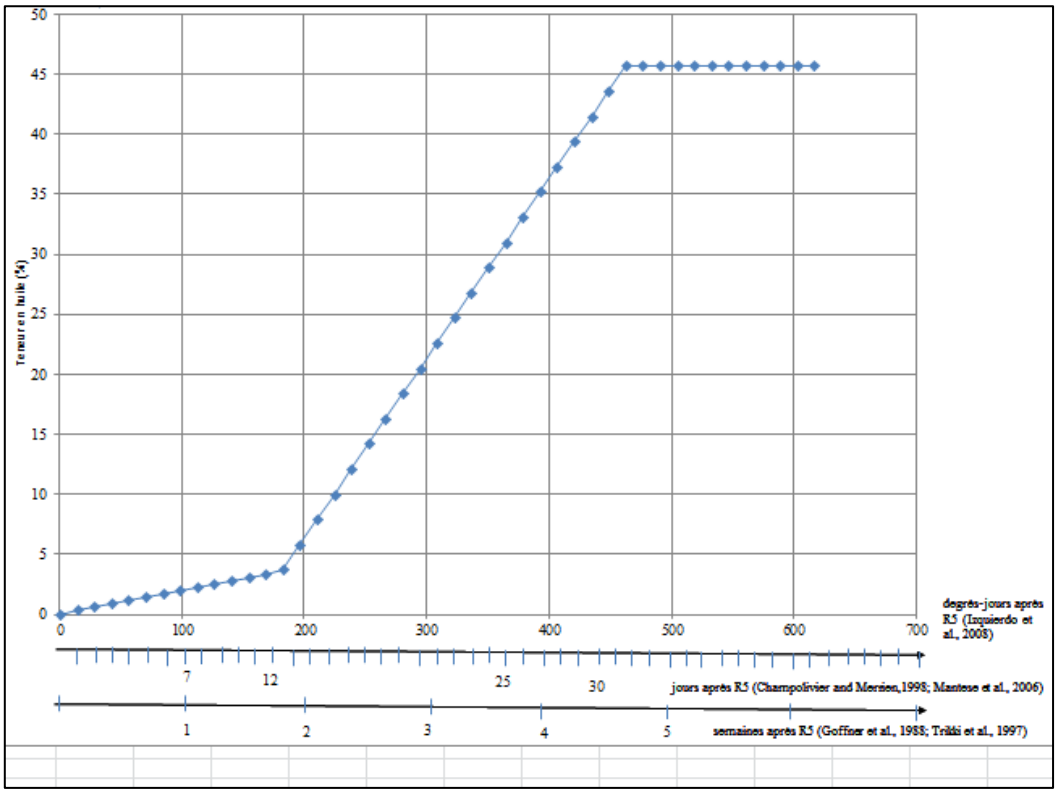


Fig.8 : Illustration de la dynamique d'accumulation de l'huile et correspondances entre degrés-jours, nombre de jours et nombre de semaines après R5 (représentation simplifiée).

Pour plus de clarté concernant les termes « début », « mi » ou « fin » floraison, nous proposons un tableau de comparaison des stades phénologiques selon l'échelle du CETIOM et celle de Schneiter et Miller (1981) (Table.1).

Table.1 : Correspondance des stades phénologiques peu avant la floraison jusqu'à la maturité entre échelle CETIOM et échelle de Schneiter et Miller (1981).

Phénophases			
	Echelle CETIOM	Echelle Schneiter	Description
	E1	R1	Apparition du bouton floral. Stade bouton étoilé
	E2	R2	Diamètre bouton compris entre 0.5 et 2 cm ; bractées visibles
	E4	R3	Bouton nettement dégagé (d :> 2 cm) + bractées déployées
Début floraison	F1	R4	Inclinaison et ouverture du bouton floral ; aperçu des fleurs ligulées perpendiculaires au plateau
	F3.2	R5	Début floraison des fleurs tubuleuses, par rangs de 3 cercles
Fin floraison	M0	R6	Fin floraison des fleurs tubuleuses ; les fleurs ligulées sèchent et tombent
Maturité physiologique	M2	R9	Le dos du capitule est jaune ; les bractées sont marquées de brun ; humidité des grains : 11% ; 75-80% de MS

2.2. Accumulation des protéines de réserve

2.2.1. Localisation dans la plante

Les protéines de réserve sont essentiellement localisées dans la graine chez le tournesol (Bauchot et Merrien, 1988).

2.2.2. Origine des protéines de réserve

Les protéines de réserve les plus rencontrées chez le tournesol sont les globulines (55 à 60%), albumines (17 à 23%), glutelines (11 à 17%) et prolamines (1 à 4%) (Bauchot et Merrien, 1988). Elles sont constituées d'acides aminés issus de phénomènes de translocation et de redistributions de l'azote. Le déchargement des acides aminés dans les embryons nécessite le franchissement de la paroi des graines, où a lieu une activité métabolique intense permettant la mise en place d'un système de transport spécifique et le stockage des acides aminés sous forme de corpuscules protéiques. Ces derniers sont synthétisés au niveau des ribosomes localisés sur le bord des membranes du réticulum endoplasmique.

2.2.3. Accumulation des protéines de réserve dans la graine

La vitesse maximale d'accumulation des protéines dans la graine s'observe entre le début et la fin floraison ; puis, les quantités continuent à augmenter à la même vitesse que les autres composés de la graine (Merrien, 1992). Les teneurs en protéines sont assez fluctuantes et se caractérisent par une dynamique en dents de scie (Goffner et al., 1988). Néanmoins, sur la période de remplissage, elles peuvent être considérées comme stables (15% ± 3%).

2.3. Antagonisme entre accumulation d'huile et de protéines

Il est souvent reporté une corrélation négative entre teneur en huile et teneur en protéines. Cela suggère qu'ils seraient en compétition dans le temps et/ou dans l'espace. Pourtant, Bauchot et Merrien (1988) démontrent que les métabolismes de l'huile et des protéines diffèrent en tous points : ils ne sont ni formés au même moment dans la graine (la protéogénèse précède la lipidogénèse), ni ne font intervenir les mêmes organites. De plus, les dépenses énergétiques nécessaires à leur synthèse sont différentes : 3 g de glucose pour 1 g de lipides, 2.5 g de glucose pour 1g de composés azotés. Merrien (1992) suggère néanmoins

qu'il doit exister une relative incompatibilité entre migrations d'assimilats protéiques (redistributions) et poursuite de l'assimilation tardive, et que c'est au niveau de l'Acétyl-CoA, précurseur commun entre synthèse d'huile et de protéines, que se joue l'équilibre entre protéogénèse et lipidogénèse. Cet équilibre serait influencé par les conditions hydriques et azotées.

2.4. Déterminisme et facteurs limitants des teneurs en huile et en protéines chez le tournesol

2.4.1. Déterminisme de la teneur en huile

L'huile provient majoritairement de la transformation d'assimilats photosynthétiques produits à partir de la floraison. Une partie complémentaire est fournie par les assimilats préalablement stockés à la floraison et remobilisés (Merrien, 1992). La durée de dépôt de l'huile dépendra essentiellement du maintien de la photosynthèse après la floraison. Or, l'activité assimilatrice des feuilles régresse à partir de la fin floraison (Blanchet et al., 1982) ; la sénescence s'accentue. Pour deux génotypes contrastés, Aguirrezábal et al. (2003) démontrent que c'est la somme de rayonnement intercepté entre 250 et 450 °C jours après la floraison qui détermine le poids des grains et la teneur en huile chez le tournesol. Les contributions relatives du carbone photosynthétique et du carbone assimilé avant la floraison dans la détermination de la teneur en huile n'ont pas été évaluées chez le tournesol. Sur le poids de graines, les résultats sont contradictoires (contribution significative de 15% à 27% dans des situations irriguées et non irriguées chez Hall et al., 1990 ; contribution non significative pour López Pereira et al., 2008)

Il est souvent reporté par ailleurs que les graines les plus grosses ont des teneurs en huile faibles car leur taux de coque est élevé (Denis et Vear, 1994). La part de la coque interfèrerait donc également dans la détermination de la teneur en huile finale.

2.4.2. Déterminisme de la teneur en protéines

Les protéines de réserve proviennent de phénomènes de translocations d'acides aminés depuis les feuilles les plus âgées et de redistributions de composés azotés à partir de la RubisCO (forme prépondérante de protéines solubles foliaires).

2.4.2.1. Absorption et assimilation de l'azote

L'absorption d'azote du sol s'effectue via les racines sous forme préférentielle de nitrate ($NO3^-$). La réduction en nitrite puis en ion ammonium s'effectue au niveau des feuilles (chloroplastes) chez le tournesol. Le $NH4^+$ sera incorporé sur des squelettes carbonés dérivés du cycle de Krebs -ayant tous comme précurseur l'AcétylCoA- (acides aminés, amides, …)- grâce à l'intervention de deux enzymes, la glutamine synthétase (dans le cytosol) et la glutamate déshydrogénase (dans le chloroplaste) (Morot-Gaudry, 2012). L'azote des feuilles est transporté sous forme de glutamine, glutamate et aspartate via le phloème jusqu'aux organes puits (Bauchot et Merrien, 1988).

2.4.2.2. Translocations et redistributions

A partir de la floraison, les feuilles matures du couvert ont fini leur expansion mais peuvent néanmoins continuer à accumuler de l'azote issu de l'absorption racinaire.Ce processus serait similaire à celui observé chez les plantes herbacées (Penhale et Thayer, 1980 ; Prado et al., 2008). Cet azote réduit est alors un surplus qui peut être immédiatement exporté vers des organes en demande d'azote plus éloignés que sont les graines en formation (Borum et al., 1989). Ce phénomène de transport « longue distance » de composés non utilisés sur le site de réduction, correspond à la translocation. Peu après, les protéines foliaires déjà présentes (essentiellement la RubiscO) sont hydrolysées et les acides aminés issus de cette hydrolyse seront remobilisés pour la synthèse d'autres composés azotés (acides aminés ou protéines), redistribués/remobilisés vers d'autres organes puits. Néanmoins, ni la durée potentielle de translocation d'azote après la floraison ni sa contribution relative dans la détermination de la teneur en protéines des graines, n'ont été évalués chez le tournesol.

2.4.3. Facteurs limitants de la teneur en huile

2.4.3.1. Génotype

La teneur en huile est une caractéristique génétiquement déterminée à forte héritabilité, mais influencée par les facteurs environnementaux (Fick, 1978). Son héritabilité à sens large, c'est-à-dire le ratio entre variance génotypique et variance phénotypique, a été estimée entre 65 à 72% à l'échelle de la graine entière (Shabana, 1974 ; Fick, 1975).

La teneur en huile (TH) de la graine peut se décomposer comme suit :

%TH graine= %coque * %TH coque + % amande * %TH amande

Denis et Vear (1996) ont regroupé 40 lignées recombinantes et 36 hybrides de tournesol selon le poids des graines, le pourcentage de coque et le taux d'huile : quelques génotypes rares avec un fort poids de graine et une faible teneur en huile, ainsi que des génotypes avec un faible poids de graines et néanmoins une faible teneur en huile, ont été identifiés. Les associations poids de graine et poids de coque et par conséquent poids de graine et teneur en huile ne sont pas systématiques. Le taux de coque est génétiquement déterminé (son héritabilité à sens large est de 27 à 32% Fick, 1978). Le progrès génétique a fortement contribué à améliorer les teneurs en huile des graines de tournesol (gain de 15 points d'huile entre les cultivars les plus anciens et les plus récents, Aguirrezábal et al., 2009). L'amélioration de la teneur en huile de graine s'est faite pour 2/3 grâce une amélioration de la teneur en amande (donc baisse du pourcentage de coque), et pour 1/3 à une augmentation de la teneur en huile dans l'amande (Connor et Hall, 1997 ; López-Pereira et al., 2000).

Des différences génotypiques au niveau du maintien de la surface foliaire verte après la floraison ont également été détectées (caractère « stay –green », de la Vega et al., 2011). Sadras et al. (1993) suggèrent que l'utilisation des pré-assimilats dépend de la quantité disponible à la floraison mais également d'une capacité proprement génotypique à prélever ces pré-assimilats. Ces caractéristiques génotypiques au niveau de la source ont été peu étudiées.

2.4.3.2. Température

Chimenti et al (2001) ont démontré que les très fortes températures (>34°C en base -1°C) entraînaient une baisse de la durée de remplissage de la graine se traduisant par une moindre quantité d'huile et une teneur en coque plus forte par rapport à une graine non stressée. Rondanini et al. (2003) démontrent que de fortes températures (>32°C) ayant lieu à mi-floraison diminuent les teneurs en huile via des diminutions de la teneur en amande et de la teneur en huile de l'amande, tandis qu'à mi-remplissage, elles affectent proportionnellement la teneur en huile de l'amande et la teneur en amande, amenant ainsi à des teneurs en huile de la graine stables. Enfin, Angeloni et al. (2012) ont identifié un seuil de température à partir duquel la teneur en huile chute de façon linéaire (17.2°C).

L'effet de la température sur la source est assez évident connaissant la dépendance des processus respiratoires (respiration de maintenance) et de la photosynthèse vis-à-vis de la température (coefficient Q10, Connor et Hall, 1997 ; Connor et Ferreres, 1999). L'effet combiné de la température et de la sécheresse de l'air entraîne une fermeture stomatique qui peut à terme altérer le système photochimique, limitant ainsi la capacité photosynthétique du couvert.

2.4.3.3. Eau

Le déficit ou stress hydrique a lieu lorsque, pour une période donnée, la demande en eau d'une culture est supérieure à l'offre disponible (faibles précipitations et/ou faibles réserves du sol ; Itier, 2008). De nombreuses études ont porté sur l'effet du stress hydrique sur le rendement (Hall et al., 1989 ; Hall et al., 1990 ; Sadras et al., 1993 ; Ebrahimi et al., 2008), mais peu d'informations sont disponibles sur la teneur en huile, au moins directement à l'échelle de la graine. En comparant 4 conduites hydriques (100% de la demande évaporative maximale ou ETM, 66%, 33%, et 0%), Santonoceto et al. (2002) ont montré que les teneurs en huile en non-irrigué étaient significativement différentes de celles conduites à 100% de l'ETM, notamment dans la partie finale d'accumulation de l'huile (phase de plateau). Anastasi et al. (2010) ont ainsi calculé qu'une augmentation de l'apport en eau de 0 à 100% de l'ETM permettait un gain de 13% sur la teneur en huile. Enfin, en situations agricoles, le stress hydrique est souvent associé aux fortes températures ou stress thermique (Tardieu et al., 2006). Nous pourrions supposer que les effets du stress hydrique sur la graine (remplissage plus court) sont similaires à ceux des fortes températures, comme démontré par ailleurs par Hall et al. (1989).

Le stress hydrique agit également sur le fonctionnement photosynthétique des feuilles avant et après la floraison ; dans les deux cas, il se manifeste par une régulation des flux transpiratoires via une fermeture progressive des stomates suivie d'une réduction de l'assimilation de CO_2 dans les feuilles (Connor et Hall, 1997 ; Tardieu et al., 2006). Avant la floraison, l'expansion des feuilles en est davantage affectée ; après la floraison, un stress hydrique intense et long peut causer une sénescence précoce suite à l'augmentation de la température des feuilles, affectant ainsi le système photochimique (Cechin et al., 2006) et créant un stress oxydatif (Grieu et al., 2008). Parallèlement, Hall et al. (1989, 1990) avaient décrit une plus forte contribution des carbohydrates de pré-floraison pour le remplissage des

graines en conditions de contraintes hydriques ; Blanchet et al. (1988) démontrent que les assimilats sont préférentiellement réorientés vers le capitule dans des cas de fort stress. Cette contribution des pré-assimilats en conditions de contrainte hydrique devrait être réévaluée sur une gamme contrastée de génotypes.

2.4.3.4. Azote

Il est souvent observé que les meilleures teneurs en huile sont obtenues dans des situations non fertilisées, comparativement aux situations bien pourvues en azote où les teneurs en huile sont les plus faibles (Connor et Hall, 1997 ; Diepenbrock et al., 2001 ; Zheljazkov et al., 2009). Ce phénomène est connu comme étant l'effet dépressif de la sur-fertilisation sur la teneur en huile (Merrien, 1992). Il pourrait s'expliquer par un effet de « dilution » de l'huile: en conditions non-limitantes d'azote comparées à une situation non fertilisée, toutes les composantes de la graine (coque, protéines, huile) sont quantitativement supérieures ; néanmoins, l'augmentation des quantités de coque et de protéines sont supérieures à celle de l'huile, aboutissant ainsi à des teneurs en huile plus faibles.

2.4.3.5. Densité de peuplement

Les résultats des travaux concernant l'effet de la densité de peuplement sur la teneur en huile sont assez contradictoires ; l'effet décrit est soit positif, négatif, ou inexistant selon les lieux, les climats et les génotypes (Gubbels et Dedio, 1986; Rizzardi et al., 1992; Diepenbrock et al., 2001). L'effet négatif de l'augmentation de la densité sur le poids moyen des graines est bien connu, mais il n'a pas d'impact systématique sur la teneur en huile. Lorsqu'il en a, l'effet densité pourrait jouer via des réductions de l'épaisseur des péricarpes (Lindström et al., 2006) menant ainsi à des teneurs en huile plus élevées. La variabilité de la réponse de la teneur en huile à la densité suggère des effets environnementaux et/ou génotype x environnement plus forts et surtout différents sur chacune des composantes de la graine qui devraient être explicités.

2.4.3.6. Maladies

Le phomopsis (*Phomopsis/Diaporthe helianthi*) et le phoma (*Phoma macdonaldii/Leptosphaeria linquistii*) sont les principales maladies cryptogamiques de fin de

cycle qui ont un impact sur la teneur en huile. Leur développement est favorisé par des conditions d'humidité élevées entre les stades bouton floral et fin floraison, liées au climat ou à la conduite de culture (peuplement élevé, fertilisation abondante, irrigation) (Debaeke et al., 2014). Concernant le phomopsis, les premiers symptômes apparaissent principalement sur les feuilles, puis progressent en direction de la tige causant une nécrose profonde (rupture de l'alimentation hydrique et casse de la tige). Des attaques tardives peuvent se manifester au niveau du capitule. Les attaques de phomopsis se traduisent par des effets négatifs sur le rendement et la teneur en huile si la contamination a lieu avant le début du remplissage. On estime la perte à 1 point d'huile et 2-3 quintaux par hectare de grain pour 10% de tiges portant au moins une nécrose encerclante sur tige (CETIOM, 2014). Les symptômes du phoma peuvent apparaître sur le collet, les feuilles, les pétioles ou les capitules. La nuisibilité du phoma, maladie très fréquente mais aux conséquences peu spectaculaires, est mal évaluée. Le phoma du collet crée un syndrome de dessèchement précoce de la plante, accélérant ainsi la sénescence et pénalisant potentiellement la teneur en huile (Seassau, 2010). Les attaques de phoma sur tiges sont responsables d'une sénescence anticipée mais les conséquences sur la production sont difficiles à mettre en évidence de manière significative.

Le sclérotinia du capitule (*Sclerotinia sclerotiorum*) est également une maladie fongique apparaissant pendant la floraison en conditions pluvieuses. Son évolution dépendra fortement de la pluviosité de fin de cycle et peut aboutir à de fortes pertes en graines, les plus importantes en France (CETIOM, 2014).

Enfin, la verticilliose (*Verticillium dahliae*) est une maladie récemment observée dans le Sud-Ouest de la France (2011), avec des symptômes qui peuvent apparaître à tous les niveaux de la plante (feuille, tige, capitule). Des microsclérotes sont conservés dans le sol pendant plusieurs années avant que leur germination ne soit stimulée par les exsudats racinaires du tournesol. Les premiers symptômes apparaissent généralement sur les feuilles à la floraison. Le mycélium formé attaque alors les tissus de la tige et les autres organes de la plante ; le diamètre du capitule peut s'en trouver fortement réduit ainsi que les composantes du rendement. Il arrive que les attaques de verticilliose amènent à des pertes de plus de 50% des plantes. La perte rapide de feuilles puis le déssèchement précoce de la plante pénalisent

potentiellement la teneur en huile, bien que l'impact de la verticilliose n'ait pas encore été quantifié sur la teneur en huile.

Il existe néanmoins des tolérances variétales à ces maladies, plus ou moins bien évaluées selon les cas, qui permettent de limiter l'impact de ces maladies en fin de cycle sur l'élaboration du rendement et de la teneur en huile.

L'analyse et la modélisation des impacts des maladies de fin de cycle sur le rendement et la composition de la graine ne seront pas abordées dans ce manuscrit.

2.4.4. Facteurs limitants de la teneur en protéines

2.4.4.1. Génotype

Peu d'études ont porté sur le déterminisme génétique de la teneur en protéines chez le tournesol. Stoyanova et Ivanov (1975) ont néanmoins montré sur des générations F1 issues de plusieurs types de croisement que leur teneur en protéines était intermédiaire et/ou tendait vers les teneurs en protéines parentales plus faibles. Cela suggère que contrairement à la teneur en huile, celle en protéines serait majoritairement influencée par l'environnement et/ou les interactions génotype x environnement, comme chez le blé (Baenziger et al., 1985 ; Aguirrezábal et al., 2009). Bien que peu étudiées chez le tournesol, des différences génotypiques au niveau de l'absorption de l'azote et de la remobilisation d'azote pourraient être attendues, comme chez le maïs (Pommel et al., 2006) ou l'orge (Dordas, 2012).

2.4.4.2. Azote

Que l'eau soit limitante ou non, de plus grandes quantités d'azote absorbées conduisent à de plus fortes teneurs en protéines. Ćupina et al. (1992) ont suggéré que les fortes doses d'azote inhibent la transformation du sucre en huile et stimulent la synthèse de composés azotés. Steer et al. (1984) montrent que chez le tournesol, de plus fortes doses d'azote, notamment si elles sont apportées autour de la floraison, prolongent la durée d'absorption de l'azote du sol et donc de l'accumulation des protéines. Il y a peu d'éléments concernant les contributions relatives de l'absorption d'azote et de la remobilisation dans la détermination de la teneur en protéines chez le tournesol.

2.4.4.3. Eau

Comme suggéré par Bauchot et Merrien (1988), les situations hydriques défavorables à la teneur en huile sont généralement favorables à la teneur en protéines. Selon Merrien (1992), les assimilats sont orientés vers des produits de moindre coût énergétique en conditions de contrainte hydrique (glucides et protéines). De plus, l'effet de la sur-fertilisation est moins marqué sur la teneur en huile si l'eau est disponible ; l'absorption d'azote dépend de la disponibilité en eau du sol mais l'absorption alimente également la demande du puits en carbone (biomasse des graines puis élaboration de l'huile). En cas de stress hydrique, l'argument d'Aguirrezábal et al. (2009) concernant une sénescence accélérée menant à une augmentation de la vitesse de remobilisation de l'azote depuis les feuilles et à de plus fortes teneurs en protéines chez les céréales, pourrait s'appliquer au tournesol.

Après cette liste non exhaustive des facteurs influençant l'élaboration de la teneur en huile et protéines chez le tournesol, nous proposons d'aborder l'aspect modélisation, en rappelant ses objectifs, puis la modélisation de la qualité chez les oléagineux et chez le tournesol en particulier.

3. La modélisation de la qualité

3.1. Définition, objectifs et performances d'un modèle

3.1.1. Définition d'un modèle

Un modèle de culture est une représentation simplifiée d'un système sol-plante-atmosphère, ou sol-couvert-atmosphère. Il décrit les relations entre ces composantes via des équations mathématiques, établies de façon empirique ou à partir de lois physico-chimiques (Wallach, 2006). La simplification du système se justifie soit par le fait qu'il y a un manque de connaissances et/ou de compréhension sur des processus biologiques seuls ou en interaction, soit parce qu'on accepte que l'on ne pourra de toute évidence pas représenter tous les phénomènes déjà connus et établis et qu'il faut faire un choix des phénomènes que l'on pense les plus importants et contribuant à décrire la variable d'intérêt.

Pour qu'un modèle fonctionne, les ingrédients de base sont une (ou plusieurs) variables explicatives et une (ou plusieurs) variables de sortie/d'intérêt. Les variables explicatives sont obtenues à partir d'expérimentations/de la connaissance bibliographique/ de bases de données. L'application d'un modèle à partir de la connaissance de variables explicatives pour une situation pédoclimatique donnée et qui mène à l'obtention d'une variable de sortie «virtuelle » s'appelle la simulation. En effet, l'objectif de la simulation via un modèle est de mimer le mieux possible le fonctionnement réel d'un système biologique.

Les variables explicatives peuvent être catégorisées en variables d'état (caractéristique du fonctionnement du couvert, par exemple) et variables d'entrée (variable souvent « extérieure », comme des précipitations, de la fertilisation azotée). Dans les variables d'entrée, on distingue les paramètres, quantités fixées pour un système donné et initialisées au début de chaque simulation. La nuance sera faite entre paramètres et constantes, car certains paramètres peuvent être spécifiques à un génotype donné et donc sont variables en fonction du génotype.

Selon le type de modèle, les variables d'entrée peuvent correspondre aux variables explicatives elles-mêmes ou être distinctes.

3.1.2. Types de modèles

Dans la modélisation du fonctionnement des cultures, on distingue généralement 2 catégories de modèles : les modèles statistiques et les modèles dynamiques (Boote et al., 1996). Il est à noter qu'un modèle dynamique n'exclut pas de modéliser certains processus de façon statistique. Généralement, un modèle statistique établit une ou plusieurs relations « figées » entre des variables explicatives et des variables de sortie à un instant t. Les modèles statistiques les plus souvent rencontrés en agronomie sont les modèles linéaires simples ou multiples, les modèles de régression hiérarchique (CART), les modèles de réseaux de neurones (neural network, NN) et les modèles additifs généralisés (GAM). La modélisation statistique se base sur l'établissement de relations entre les variables explicatives et les variables expliquées à partir de base de données, *i.e.* d'une collection souvent conséquente de résultats d'expérimentations obtenus sur plusieurs années culturales. L'aspect « statistique » peut entraîner une absence de relation de causalité - physiologiquement ou agronomiquement explicable- entre les variables.

Les modèles dynamiques sont des modèles beaucoup plus « souples » dans le sens où les phénomènes peuvent être modulés en fonction de la variation d'une ou plusieurs autres variables explicatives. Ces variations peuvent être décrites à l'échelle temporelle (ex : simulation journalière de l'accumulation de l'huile) ou spatiale (ex : effet de la variation du rayonnement sur l'activité photosynthétique de différentes zones d'une feuille)

Les modèles dynamiques sont souvent assimilés à des modèles dits mécanistes (« process-based » ou « process –oriented ») de par le fait qu'ils sont construits pour décrire un ou plusieurs mécanismes donnés, théoriquement universel(s) pour une espèce, et de la modulation par des facteurs environnementaux ou génétiques de ces mécanismes qui peut avoir lieu à différents moments et niveaux du mécanisme représenté.

Le choix entre modèle statistique/ dynamique et la représentation du niveau de complexité des phénomènes dépend de l'objectif du modélisateur.

3.1.3. Finalités d'un modèle

Depuis l'apparition des premiers modèles sol-plante-atmosphère dans les années 1960 (De Wit, 1959 ; 1966 ; 1970) puis 70 et 80 (Thornley, 1972 ; Sinclair et al., 1977) jusqu'aux modèles actuels de prédiction de performances agronomiques intégrant des relations génotype

–phénotype (modèles QTL, Hammer et al., 2006), les objectifs de la modélisation se sont diversifiés. Ils peuvent être classés en deux catégories, selon qu'ils sont construits pour comprendre, ou pour prédire.

3.1.3.1. Modélisation heuristique : intégration des connaissances et compréhension de la physiologie

Souvent, de nombreuses études dans une même discipline ou dans différentes disciplines sont effectuées séparément sans qu'un travail d'intégration des principaux résultats ne soit achevé.

Une manière efficace de valoriser ces résultats est de les intégrer via un modèle en identifiant des résultats similaires, ceux originaux, et ceux qui permettent de faire le lien entre différents processus et échelles/niveaux d'organisation. Bannayan et al. (2007) suggèrent par exemple d'intégrer les nouvelles connaissances acquises en biologie moléculaire et en génétique sur la phénologie, la photosynthèse et la réponse au stress hydrique dans des modèles éco-physiologiques afin de réduire l'incertitude liée aux paramètres « génétiques » établis à partir du phénotype. L'intégration de connaissances de différentes natures amènera à de nouveaux questionnements permettant ainsi au chercheur d'élaborer de nouvelles hypothèses sur le fonctionnement d'un système-plante-atmosphère (Boote et al., 1996). Le modèle peut être testé et comparé à des résultats expérimentaux nouveaux afin de vérifier la validité des hypothèses initialement posées et/ou de proposer de nouvelles hypothèses. Cette démarche « hypothèse-vérification » peut évidemment être réalisée via des expérimentations mais la possibilité de mettre en place ces expérimentations est souvent limitée dans la réalité agronomique ; le fait de simuler des performances à partir d'hypothèses permet d'avancer dans la compréhension des phénomènes/processus tout en réalisant un gain de temps important (Fig.9).

3.1.3.2. Prédiction, aide à la décision et conseil variété-conduite

Chaque valeur obtenue à la fin de la simulation d'une situation donnée est une prédiction ; de nombreux modèles ont été construits afin qu'ils prédisent le mieux possible (le plus proche de valeurs réelles) les performances d'une plante ou d'une culture. Lorsqu'un modèle est considéré comme robuste et fiable, il peut être utilisé pour optimiser un ou plusieurs itinéraires techniques afin d'optimiser les performances des plantes, qu'il s'agisse d'un choix de variété, de la date d'apport d'eau ou de fertilisant azoté ou de choix de densité de peuplement (Bock et Sikora, 1990 ; Boote et al., 1996 ; Bergez et al., 2001 ; Rinaldi et al., 2003). Les modèles dynamiques les plus récents, dans lesquels sont intégrés des paramètres

génotypiques à des niveaux détaillés de processus dynamiques, permettent de prendre en compte une multitude de possibilités d'interactions entre le génotype et l'environnement et d'établir des classements de variétés en fonction de leurs performances ; certains modèles pourraient ainsi assister l'évaluation variétale et contribuer à l'amélioration génétique en permettant de sélectionner des critères/traits vis-à-vis d'un processus donné (par exemple, caractériser la réponse au stress hydrique via un indicateur de résistance stomatique) mais également d'identifier les génotypes présentant les meilleures performances simulées et leurs caractéristiques (Boote et al., 1996 ; Chapman, 2008 : Debaeke et al., 2012). La modélisation permet alors de fournir un conseil variétal et variété-conduite aux sélectionneurs et aux agriculteurs.

3.1.4. Evaluation d'un modèle

L'évaluation d'un modèle, souvent appelée « validation », est une étape essentielle dans le processus de modélisation. Le terme « validation » semble néanmoins incorrect ; un modèle peut être invalidé mais ne peut être définitivement entériné (Wallach, 2006) car il n'est et ne restera toujours qu'une approximation de la réalité. L'évaluation consiste à comparer les valeurs prédites issues de la simulation avec des valeurs réelles, et à fournir un ensemble d'indicateurs de performance du modèle. Cette étape conditionne le niveau d' « acceptabilité » d'un modèle ; en effet aucun modèle ne sera parfait et plusieurs modèles sont possibles pour décrire un seul et même jeu de données. Ainsi, Bellocchi et al. (2008) proposent des ordres de grandeur basés sur les travaux de Stöckle et al. (2004) pour évaluer la simulation de l'évapotranspiration ; un modèle peut être considéré comme « très bon » si la RRMSEP est inférieure ou égale à 10% ; « bon » si la RRMSEP est entre 10 et 15% ; acceptable si elle est entre 15 et 20%. Des erreurs relatives de prédiction supérieures à 25% indiqueraient un modèle « mauvais ». Néanmoins, c'est au modélisateur de choisir le modèle qui lui conviendra et la crédibilité qu'il peut lui accorder selon les objectifs qu'il s'est fixé.

L'évaluation diffère du paramétrage/calibration ou ajustement du modèle. Ainsi, le paramétrage consiste à affecter des valeurs de paramètres établis dans un modèle afin qu'il soit le plus proche possible de valeurs réelles lors de sa construction. Les valeurs de paramètres sont obtenues soit à partir de connaissances bibliographiques, soit à partir de méthodes statistiques plus ou moins élaborées (méthodes de régression linéaire et non-linéaire, méthodes bayésiennes…) (Makowski et al., 2006). Le paramétrage et l'évaluation sont effectués sur des bases de données différentes. Concernant l'évaluation, différents indicateurs de « proximité » avec les valeurs réelles sont proposés par Wallach (2006)

(Table.2), dont les plus communs sont le biais, la RMSEP, la RRMSEP, l'efficience du modèle et le coefficient de corrélation. Kobayashi et Salam (2000) ont décomposé l'erreur carrée moyenne prédite (MSEP) en ses 3 composantes : le biais au carré (squared bias, SB), la différence entre les écart-types (square differences of the standard deviations, SDSD) et l'absence de corrélation pondérée par l'écartype (lack of correlation weighted by the standard deviation, LCS). Ces derniers indicateurs permettent d'identifier les points à améliorer dans la modélisation car leur utilisation aboutit à une quantification de l'erreur liée à une incapacité du modèle à simuler la magnitude (fort SDSD) ou la dynamique (fort LCS) de la variable expliquée. Si le but du modèle est de prédire, il n'y a pas de modèle valable sans évaluation.

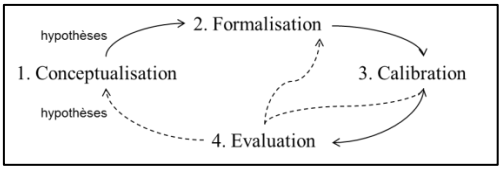


Fig.9 : Démarche de modélisation

Table.2 : Tableau d'indicateurs couramment utilisés pour évaluer les modèles en agronomie (source modifiée : Wallach et al., 2006). Yobs et Ysim représentent les variables observée et simulée, respectivement. N représente la taille de la population (nombre de situations simulées)

	indicateur	Définition	Equation
outils d'évaluation	Di	résidus/déviation	$Di = Yobs - Ysim$
	r	coefficient de corrélation	$r = \frac{covariance(Yobs, Ysim)}{écartype(Ysim) - écartype(Yobs)}$
	biais	-	$Biais = \frac{1}{N}\sum_{i=1}^{N} Di$
	RMSE(P)	root mean squared error (of prediction) racine de l'erreur carrée moyenne (de prédiction)	$RMSE(P) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Di)^2}$
	RRMSE(P)	relative root mean squared error (of prediction) racine de l'erreur relative carrée moyenne (de prédiction)	$RRMSE(P) = \frac{RMSE(P)}{moyenne(Yobs)} \times 100$
outils de diagnostic	MSE(P)	mean squared error (of prediction) erreur carrée moyenne (de prédiction)	$MSE(P) = SB + SDSD + LCS$
	SB	squared bias (biais au carré)	$SB = (Biais)^2$
	SDSD	squared difference between standard deviations différence au carrée des écartypes	$SDSD = (écartype(Yobs) - écartype(Ysim))^2$
	LCS	lack of correlation weighted by the standard deviations absence de corrélation pondérée par les écartypes	$LCS = 2 \times écartype(Yobs) \times écartype(Ysim) \times (1 - r)$

Les différences de performances entre modèles qui simulent une même variable de sortie (par exemple le rendement chez une espèce donnée) proviennent généralement soit du choix du type de modèle, des variables prises en compte, ou de la base de données utilisée.

Dans la prochaine section, nous allons passer en revue les modèles les plus connus, génériques puis spécifiques au tournesol; après une brève description des modèles, un focus sera fait sur l'aspect qualité (teneur en huile et teneur en protéines) dans ces modèles : le formalisme, les variables prises en compte et la qualité de prédiction.

3.2. La modélisation de la qualité chez les oléagineux

3.2.1. Modèles génériques

3.2.1.1. Le modèle STICS

Le modèle STICS (Simulateur mulTIdisciplinaire pour les Cultures Standard) est un modèle dynamique journalier qui décrit à travers différents modules l'éco-physiologie de plante (phénologie, production de biomasse aérienne et rendement en quantité et qualité), les fonctions du sol en interaction avec le fonctionnement des racines (bilan hydrique, bilan azoté) (Brisson et al., 2003). Il a été paramétré pour plusieurs espèces (ex : blé, maïs, soja, tomate, tournesol).

La teneur en huile de la graine dans STICS est considérée comme étant proportionnelle à la matière sèche de la graine. La quantité d'azote dans la graine est fonction d'un indice de récolte azoté proportionnel à la durée de remplissage. Malheureusement, nous n'avons pas trouvé à ce jour de travaux d'évaluation de la qualité de prédiction de cette teneur en huile ni de la teneur en protéines chez les oléagineux via STICS.

3.2.1.2. Le modèle CERES

Le modèle CERES (Crop Environment Resource Synthesis) comprend des formalismes mécanistes et statistiques qui permettent de simuler la croissance et le développement de la plante. Initialement développé pour le maïs (Jones et Kiniry, 1986), le modèle a été ensuite largement utilisé pour simuler le fonctionnement d'autres espèces notamment céréalières (blé, orge, sorgho, millet et riz) mais aussi oléagineuses (colza par exemple, Gabrielle et al., 1998). Dans CERES-Rape, l'azote entrant dans les siliques provient de translocations depuis

les feuilles, tiges et racines. La quantité d'azote entrante est limitée à 5%. Une répartition de l'azote a lieu entre les parois des gousses et les siliques. La teneur en huile est déduite de la teneur en azote des siliques et du poids des graines. Gabrielle et al. (1998) ont constaté après évaluation du modèle une surestimation de la teneur en huile (ordre de grandeur non fourni) ; elle serait liée à des erreurs sur la représentation de la teneur en azote de la plante.

3.2.1.3. Le modèle CROPGRO

Le modèle CROPGRO est un modèle mécaniste capable de simuler la croissance de la culture et des processus biologiques généraux tels que la photosynthèse et l'absorption d'azote (de Carvalho Lopes et al., 2011). CROPGRO a été paramétré notamment pour l'arachide, le haricot et le soja. La teneur en huile est estimée comme étant proportionnelle au poids de graines (Boote et al., 2003) ; aucune information sur la qualité de prédiction de l'huile ni des protéines n'a été trouvée.

3.2.1.4. Le modèle EPIC

Le modèle EPIC (Environmental Policy Integrated Climate Model) est un modèle construit par Williams (1990) dont le but est d'évaluer d'un point de vue agronomique et économique les effets de l'érosion du sol sur sa productivité. L'hydrologie, l'utilisation de l'azote et du phosphore, le devenir des pesticides et la croissance des cultures y sont simulés. Les processus sont décrits à l'échelle journalière (Izaurralde et al., 2006) et adaptables aux échelles régionales et globales. Le modèle EPIC a été par ailleurs paramétré pour plus de 100 espèces végétales, dont le tournesol (Kiniry et al., 1992). Le rendement en graines est simulé à partir de la biomasse totale et d'un indice de récolte modulé par le stress hydrique en phase de remplissage. Une version améliorée de ce modèle (EPICphase) a été proposée par Cabelguenne et al. (1999), dans laquelle l'indice de récolte peut être modulé par du stress hydrique et azoté avant, pendant, et après la floraison. La capacité d'adaptation du tournesol et du soja au stress hydrique (réduction des besoins en eau) est prise en compte. Ce modèle montre une meilleure correspondance entre rendement en grains simulé et observé (RMSEP=0.32 et 0.25 Mg par ha pour le tournesol et le soja, respectivement) par rapport à EPIC (RMSEP = 1.04 et 0.79 Mg par ha). Les teneurs en huile et en protéines n'y sont pas simulées.

3.2.2. Modèles spécifiques aux oléagineux

3.2.2.1. Colza

Le modèle Azodyn-Colza (Jeuffroy et al., 2006) est un modèle dynamique dérivé du modèle Azodyn-Blé, qui simule de façon journalière la croissance de la culture, l'expansion et la sénescence des feuilles, les flux d'azote et d'eau dans le système sol-plante ainsi que le rendement et la qualité des graines. Cette dernière est simulée de façon statistique à la récolte. La teneur en protéines est d'abord calculée à partir du poids des graines et d'une teneur en protéines potentielle permise par le statut azoté de la plante à la floraison (INN) ; la teneur en huile en est déduite en vertu d'une corrélation très étroite constatée avec la teneur en protéines. Plusieurs évaluations du modèle ont montré que les erreurs de prédiction pour la teneur en huile et en protéines étaient en moyenne de 3.1% et 4.5% respectivement.

3.2.2.2. Coton

Un modèle dynamique simulant au pas de temps journalier la croissance des capsules, les taux de protéines et d'huile chez le coton en Chine a été construit par Li et al. (2009). L'accumulation d'huile dépend d'une capacité à fabriquer des acides gras (nulle avant 6 jours et après 36 jours en équivalent de « temps physiologique de développement »), et de la demande journalière en huile (poids de capsule journalière multipliée par une teneur en huile maximale). Ces deux composantes sont modulées par des effets température et azote. Le calcul de la teneur en protéines se base sur une demande journalière en azote structural et en azote de stockage dans la graine, tous les deux fonctions de la biomasse de la capsule à un jour donné. L'accumulation de protéines est modulée par des effets température, azote, et position de la capsule. Le modèle démontre une qualité de prédiction de 2.2% et 2.7% pour les taux de protéines et d'huile, respectivement.

3.2.2.3. Palme

Enfin, des modèles statistiques, plus marginaux, ont été développés pour simuler la teneur en huile de palme en Malaisie, soit en fonction de la composition nutritionnelle des feuilles (Khamis, 2006), soit à partir de la disponibilité en eau dans le sol et l'âge de la plante (Keong et Keng, 2012). Néanmoins, aucune évaluation n'a été effectuée pour ces modèles.

3.2.3. Principaux modèles spécifiquement développés pour le tournesol

3.2.3.1. Le modèle QSUN

a) Description générale du modèle QSUN

QSUN est l'un des premiers modèles élaborés spécifiquement pour simuler les performances du tournesol dans les conditions sèches d'Australie (Chapman et al., 1993) . C'est un modèle mécaniste qui prédit la croissance (surface foliaire, sénescence et biomasse), le développement (phénologie : émergence, bouton florale, mi-floraison et maturité) et le rendement d'une culture de tournesol. Les variables climatiques prises en compte sont la température, la photopériode, le rayonnement et la pluviométrie. La production de biomasse est modulée par la quantité d'eau disponible dans le sol qui est également simulée. Le rendement est le produit de la biomasse aérienne par un indice de récolte qui augmente de façon linéaire à partir de la fin floraison. Il est corrigé par un coefficient de conversion afin de tenir compte du coût énergétique de l'huile.

b) Focus sur l'aspect qualité du modèle QSUN

Le modèle QSUN simule la teneur en huile de façon linéaire pendant 25 jours après l'anthèse ; elle peut atteindre un maximum de 45%. Chaque gramme d'huile est multiplié par un coefficient de conversion (2.24) pour obtenir des équivalents carbohydrates. Le modèle a été ensuite évalué sur des situations contrastées d'humidité du sol avec des variétés à précocité contrastées dans le Nord de l'Australie (Meinke et al., 1993) mais les performances concernant la teneur en huile n'ont pas été indiquées. La teneur en protéines n'est pas simulée dans ce modèle.

3.2.3.2. OILCROP-SUN

a) Description générale du modèle OILCROP-SUN

Le modèle OILCROP-SUN (Villalobos et al., 1996) est un modèle mécaniste et dynamique qui simule de façon journalière la phénologie, la biomasse de différents organes et leur teneur en azote, l'indice foliaire et la sénescence des feuilles, les bilans hydriques et azotés du sol et le rendement du tournesol. Ce dernier est séparé en ses deux composantes que sont le nombre de grains par m² (calculé à partir d'un paramètre génotypique de potentiel de nombre de grains par plante) et le poids de grains, lequel est alimenté par l'accumulation de biomasse

issue de l'interception du rayonnement selon des règles de priorité d'une part, et par la contribution de réserves pré-stockées mobilisées en cas d'insuffisance photosynthétique, d'autre part.

b) Focus sur l'aspect qualité du modèle OILCROP-SUN

Le modèle OILCROP-SUN simule l'accumulation d'huile dans la graine selon un taux fixe à partir de 13 jours après le début de floraison soit 4 jours après le début de remplissage des amandes, ces dernières s'accumulant 9 jours après les coques (début de floraison). Des paramètres spécifiques aux cultivars ont été introduits notamment au niveau de la phénologie, du nombre de grains potentiel, de la vitesse de croissance potentielle de l'amande et de la durée de remplissage, mais aucun paramètre génotypique n'a été appliqué sur l'huile. La teneur en protéines n'y est pas simulée. Le modèle a été ensuite évalué sur deux bases de données indépendantes, mais la performance sur la teneur en huile n'a pas été indiquée. Néanmoins, les auteurs ont suggéré sur ce point que de meilleures connaissances sur la durée du remplissage seraient nécessaires pour améliorer la modélisation.

3.2.3.3. Modèle de Pereyra-Irujo et Aguirrezábal (2007)

a) Description générale du modèle de Pereyra-Irujo et Aguirrezabal (2007)

Pereyra-Irujo et Aguirrezábal (2007) ont développé un modèle simple permettant de simuler le rendement et ses composantes ainsi que la qualité de la graine (teneur en huile) et de l'huile (teneur en acide oléique, acide linoléique, ratio tocophérols : acide linoléique). La phénologie et la croissance (surface foliaire) y sont simulées de façon journalière à partir de données climatiques, et de la densité de peuplement. Le nombre de grains, le poids de grains et la teneur en huile sont simulées à partir de relations pré-établies (Cantagallo et al., 1997 ; Aguirrezábal et al., 2003) de façon statistique.

b) Focus sur l'aspect qualité du modèle de Pereyra-Irujo et Aguirrezábal (2007)

Concernant le formalisme de la teneur en huile dans ce modèle, il correspond au minimum entre une teneur en huile potentielle maximale (50%) et un ratio impliquant la somme de rayonnement intercepté entre 250 et 450°C jours après floraison et la densité de peuplement. La teneur en protéines n'y est pas simulée. L'évaluation de ce modèle sur une base de

données indépendante avec 2 génotypes montre une qualité de prédiction de 1.4% (valeur de RMSEP) sur la teneur en huile. Ces situations correspondent à des situations non-limitantes en eau et en azote.

3.2.3.4. Le modèle SUNFLO

a) Description générale du modèle SUNFLO

SUNFLO (Casadebaig, 2008 ; Casadebaig et al., 2011) est un modèle décrivant de façon journalière la phénologie, la production de biomasse et son allocation dans la graine. Le modèle simule des bilans hydriques et azotés et prend en compte l'effet des stress hydrique et azoté sur l'expansion des feuilles, la production de biomasse et l'allocation. L'originalité de SUNFLO réside dans l'intégration de douze paramètres génotypiques établis selon des critères de variabilité génotypique, de stabilité dans plusieurs environnements, d'impact significatif sur les variables de sortie et de facilité de mesure. L'objectif appliqué du modèle est d'assister l'évaluation de variétés en réseaux d'inscription et post-inscription, en fournissant des informations complémentaires sur les comportements variétaux dans des conditions non prises en compte dans les essais, permettant ainsi de proposer les couples variété-conduite les plus pertinents (Debaeke et al., 2010). L'ensemble des variables d'état et de sortie sont simulées de façon dynamique sauf les variables d'allocation que sont l'indice de récolte et la teneur en huile, représentées de façon statistique.

b) Focus sur l'aspect qualité du modèle SUNFLO

La teneur en huile dans le modèle SUNFLO est calculée via une régression linéaire multiple avec 8 variables explicatives (prédicteurs) décrivant le fonctionnement du couvert en phase de post-floraison, des indices de stress azoté et hydrique et des caractéristiques génotypiques de teneur en huile potentielle et de date de floraison. Lorsqu'évalué sur une base de données où les conditions pédoclimatiques ont été plus ou moins bien caractérisées, l'erreur moyenne de prédiction sur la teneur en huile est de 3.7% (RMSEP). La décomposition de l'erreur montre une incapacité du modèle à mimer les variations créés par les différents stress (hydrique, azote) sur la teneur en huile (Casadebaig et al., 2011)

Table.3. Tableau récapitulatif des modèles existants chez les oléagineux en général et chez le tournesol en particulier. Lorsque le modèle n'a pas de nom, l'espèce concernée est indiquée entre parenthèses.

	nom modèle	référence	oléagineux	tournesol	teneur en huile (TH)	teneur en protéines (TP)	Evaluation qualité	qualité de prédiction TH	qualité de prédiction TP
modèles génériques	STICS	Brisson et al. (2003)	oui	oui	oui	oui	non	-	-
	CERES	Gabrielle et al. (1998)	oui	non	oui	oui	oui	-	-
	CROPGRO	Boote et al. (2003)	oui	non	oui	non	non	-	-
	EPIC	Cabelguenne et al. (1999)	oui	oui	non	non	non	-	-
autres modèles oléagineux	Azodyn-Colza	Jeuffroy et al. (2006)	oui	non	oui	oui	oui	3.1	4.5
	(coton)	Li et al.(2009)	oui	non	oui	oui	oui	2.7	2.2
	(palme)	Khamis (2006) Keong et Keng (2012)	oui	non	oui	non	non	-	-
modèles spécifiques tournesol	QSUN	Chapman et al. (1993)	oui	oui	oui	non	non	-	-
	OILCROP-SUN	Villalobos et al. (1996)	oui	oui	oui	non	non	-	-
	(tournesol)	Pereyra-Irujo et Aguirrezábal (2007)	oui	oui	oui	non	oui	1.4	-
	SUNFLO	Casadebaig et al. (2011)	oui	oui	oui	non	oui	3.7	-

III. Enoncé de la problématique : objectifs, démarche de la thèse et plan du manuscrit

L'exposé des travaux précédemment réalisés chez le tournesol sur la qualité des graines et sa modélisation a permis de montrer que :

- Concernant l'huile, des modèles empiriques et/ou statistiques existent; ces derniers sont soit non évalués, soit évalués dans des conditions non-limitantes avec une bonne qualité de prédiction (1.4 point d'huile), soit évalués dans des conditions limitantes avec une qualité de prédiction non satisfaisante (3.7 points d'huile).
- Concernant les protéines, aucun modèle de simulation ne tient compte des spécificités de l'akène de tournesol.

Nous avons par ailleurs démontré la nécessité de régulariser et d'optimiser les teneurs en huile et les protéines chez le tournesol:

- L'huile, parce qu'elle constitue une plus-value environnementale et nutritionnelle (vitamine E, acide oléique) pour la production de tournesol, et qu'elle peut être augmentée conjointement au rendement ; de plus, elle est au cœur d'un système économique de compensation/pénalisation des agriculteurs par les coopératives ;
- Les protéines, parce que face à une demande croissante de protéines végétales (ratio protéique) pour l'alimentation animale, le tourteau de tournesol représente une source potentiellement très compétitive de protéines, notamment depuis l'amélioration des processus de décorticage ayant abouti à des tourteaux de tournesol aussi riches en protéines que ceux du colza.

Pourtant, les teneurs en huile et en protéines sont déterminées à la fois par des facteurs génotypiques et environnementaux. La part relative de ces facteurs n'est pas fixe mais dépend des conditions de stress qui créent des interactions parfois complexes entre génotype et environnement. Il devient alors difficile de prédire une teneur uniquement à partir de la seule connaissance du potentiel d'un génotype. Or, pouvoir optimiser une teneur, c'est d'abord pouvoir la prédire.

1. Hypothèses et objectifs de la thèse

L'objectif de la thèse est double : (1) comprendre l'élaboration de l'huile et des protéines et (2) proposer un modèle de culture pour simuler les teneurs en huile et protéines chez le tournesol. Nous faisons ainsi l'hypothèse qu'une meilleure connaissance/compréhension (i) des déterminismes majeurs des teneurs en huile et protéines, de leurs poids relatifs (ii) de leur mode d'action et de leur(s) interaction(s) doit permettre de mieux prédire ces teneurs. Ainsi, la construction d'un modèle ayant à la fois un sens « physiologique » et une qualité de prédiction « acceptable » constitue également un objectif à part entière dans le cadre de la thèse.

2. Questions de recherche

Une meilleure compréhension des facteurs les plus déterminants conditionnant la construction d'un futur modèle, nous proposons de répondre aux deux questions de recherche suivantes :

(1) Quels sont les facteurs les plus importants influençant la qualité et quelle est leur hiérarchie ?

En effet, l'exposé des travaux sur le déterminisme de la teneur en huile et en protéines a permis de démontrer qu'il existe encore des zones d'ombre concernant l'importance relative des facteurs identifiés. C'est le cas par exemple des facteurs azote et stress hydrique qui ont fait l'objet de moins d'études et dont les effets sont moins bien compris. De plus il n'y a pas eu, à notre connaissance, d'analyse intégrée du poids des différents facteurs déterminants dans une seule et même étude ; il serait ainsi opportun d'établir une hiérarchie dans cette liste de déterminants potentiels en évaluant leur importance relative sur les teneurs.

(2) Sur quelles variables/composantes agissent les facteurs les plus importants, et comment ?

Une fois les facteurs/variables les plus importants sélectionnés, il s'agira ensuite de comprendre comment elles expliquent et/ou sont corrélées aux teneurs en huile et protéines : leur effet est-il positif, négatif, ou variable selon les situations (cas d'interaction) ? Est-il le même au cours du temps, ou bien existe-t-il des périodes plus sensibles aux facteurs déterminants que d'autres, comme démontré pour le rayonnement intercepté (Aguirrezábal et al., 2003) ? Existe-t-il une modulation différente des organes « sources » et « puits » par ces

différents facteurs ? Comment ces facteurs jouent-ils sur les processus clés déterminant les teneurs en huile et protéines ?

3. Stratégie de recherche

a) Utilisation de la notion de relations source-puits au sein de la plante

L'analyse bibliographique précédente nous a permis d'identifier un ensemble de facteurs potentiellement déterminants pour les teneurs en huile et en protéines ; les effets connus de ces facteurs portent parfois sur des processus au niveau d'organes « sources », « puits », ou les deux. Sont définis comme organes « sources » ceux qui exportent plus d'assimilats qu'ils n'en importent, tandis que les organes « puits » sont des consommateurs nets de photoassimilats (Merrien, 1992). Après la floraison et durant la phase de remplissage chez le tournesol, les organes « sources » sont les feuilles, les tiges et les réceptacles (Connor et Hall, 1997). Il est à noter que les réceptacles sont d'abord des organes « puits » peu après la floraison et jusqu'à mi-remplissage (López-Pereira et al., 2008) avant de jouer le rôle d'organes « sources » pour les graines. Les organes « puits » que sont les constituants des graines s'élaborent progressivement : d'abord les coques, puis les amandes où s'accumulent essentiellement l'huile et les protéines (Roche, 2005). On définira ainsi les relations source-puits chez le tournesol par les transferts d'assimilats depuis les organes « sources » vers les organes « puits » (Dordas, 2012).

La notion de relations source-puits peut également être décrite en termes de processus de remobilisations et assimilations carbonées et azotées ; ainsi, les organes végétatifs constituent des sources potentielles d'assimilats pré-stockés vers les organes « puits » via des processus de remobilisation, tandis que le maintien de l'activité photosynthétique et l'assimilation azotée via le processus d'absorption racinaire (Merrien et al., 1988) fournissent du carbone et de l'azote pour élaborer l'huile et les protéines. Les racines en tant qu'organes ne seront pas prises en compte dans la suite de ce travail ; nous nous focaliserons plutôt sur le processus même d'absorption racinaire.

Afin de mieux comprendre les effets des différents facteurs et leurs impacts sur les teneurs en huile et en protéines des graines, nous proposons ainsi dans le cadre de la thèse de décrire ces effets via les relations sources-puits, définis aussi bien en termes d'organes qu'en termes de processus en phase de post-floraison.

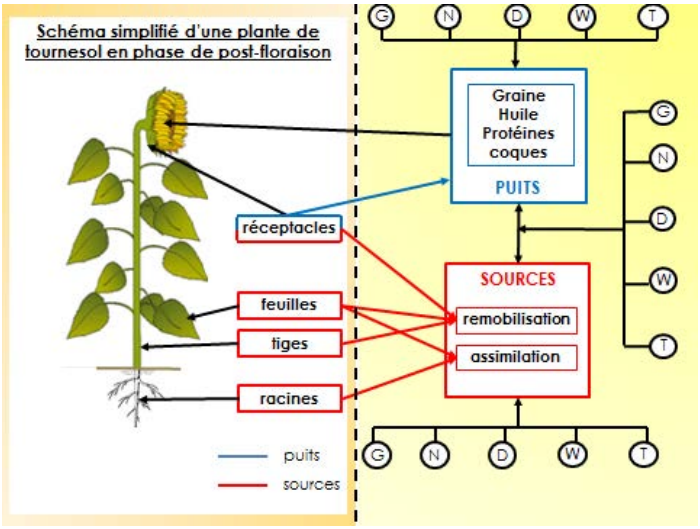


Fig.10. Représentation simplifiée des organes « sources » (en rouge), « puits » (en bleu) et des relations sources-puits chez le tournesol en phase de post-floraison. Ces relations peuvent également être décrites en termes de processus : remobilisation et assimilation. Les flèches noires indiquent les organes, les flèches rouges et bleues le type d'organes. Enfin, les facteurs déterminants des teneurs en huile et protéines les plus cités dans la littérature sont indiqués : génotype (G), azote (N), densité de peuplement (D), stress hydrique (W) et température (T). Ces facteurs peuvent agir au niveau des organes « sources » et/ou des organes « puits » et/ou de la relation entre organes « sources » et organes « puits ».

b) Choix de deux types de modélisation

La modélisation sera à la fois l'outil et la finalité de la thèse, puisqu'elle contribuera à l'avancée des connaissances et à prédire les teneurs en huile et protéines dans le cadre d'un conseil variétal auprès des agriculteurs/coopératives, et/ou d'évaluation de stratégies agronomiques les plus adaptées pour une situation pédoclimatique donnée. Deux familles de modèles seront utilisées (Fig.11) :

La modélisation statistique est retenue (1) pour identifier et hiérarchiser les variables déterminantes, et (2) pour disposer d'un outil rapide et simple de prédiction.

La modélisation dynamique permet de décrire et analyser de façon journalière l'accumulation de l'huile et des protéines. L'hypothèse est qu'un modèle dynamique serait mieux à même d'expliquer les effets différentiels des facteurs agronomiques qui sont à l'origine de potentielles interactions génotype-environnement sur les teneurs en huile et protéines (Merrien, 1992 ; Chapman, 2008 ; Wallach et al., 2013).

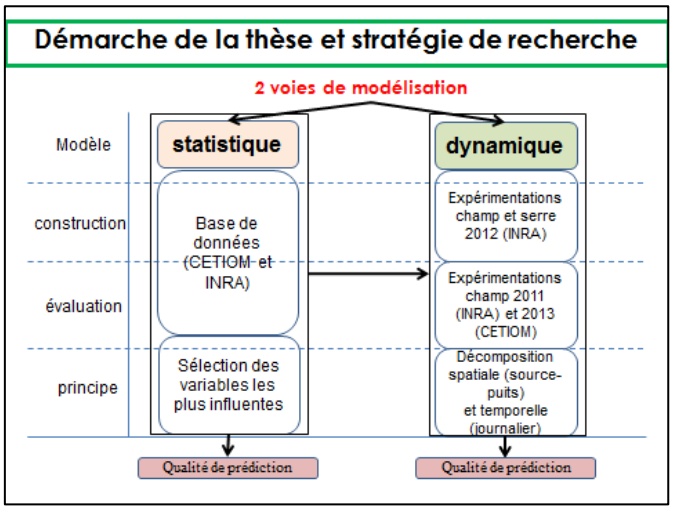


Fig.11. Illustration de la démarche de la thèse du point de vue de la modélisation. Deux types de modélisation seront abordés : statistique et dynamique. Les bases de données/expérimentations utilisées pour construire les modèles sont présentées dans les chapitres I (statistique), II et IV (dynamique), ainsi que celles utilisées pour la validation des modèles. Leur qualité de prédiction sera évaluée.

c) Un cadre de modélisation de référence : le modèle de culture SUNFLO

Enfin, l'ensemble de ce travail d'analyse et de modélisation sera réalisé en utilisant comme cadre le modèle de culture SUNFLO (Casadebaig et al., 2011) (Fig.12) ; d'abord comme cadre d'intégration des nouveaux modèles proposés pour les teneurs en huile et en protéines, mais également comme outil de simulation de variables non mesurées et/ou non accessibles. C'est le cas par exemple d'indicateurs avant la floraison (nombre de jours de stress hydrique, …), à la floraison (efficience d'utilisation du rayonnement,…), ou après la floraison (disponibilité hydrique et azotée du sol). L'organisation en modules du modèle SUNFLO permet de proposer de nouveaux formalismes pour certains processus sans modifier la totalité du modèle. L'un des critères attendus de notre nouveau modèle étant la bonne qualité de prédiction, notre objectif en ce sens sera d'améliorer le modèle de culture SUNFLO.

Ce dernier a été choisi comme cadre de la problématique de thèse en vertu d'un certain nombre de critères : (1) parce que le modèle issu de la collaboration étroite INRA/CETIOM (Casadebaig, 2008) est transparent et facilement utilisable; (2) parce qu'il a été paramétré pour un nombre important de génotypes de tournesol cultivés en France (>50) ; (3) parce qu'il peut être considéré comme représentatif des situations pédoclimatiques types car il a été validé sur une large gamme de conditions de culture en France et (4) parce qu'il est modulaire : il est possible de travailler sur des parties du modèle sans en affecter d'autres.

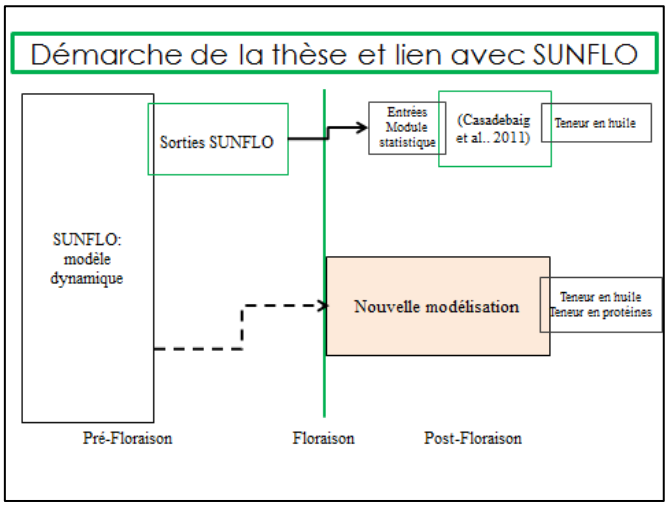


Fig.12. Illustration de la démarche de la thèse en lien avec le modèle de culture SUNFLO (Casadebaig et al., 2011). L'objectif étant de proposer une nouvelle modélisation de la teneur en huile et protéines pendant la phase de remplissage, le modèle SUNFLO sera utilisé pour simuler des variables non accessibles/non mesurées, à différentes périodes du cycle de culture. La modularité du modèle SUNFLO permettra de proposer de nouveaux formalismes qui pourront être connectés à d'autres « parties » du modèle global.

4. Plan du manuscrit

La première partie de la thèse (Chapitre I) concerne l'utilisation de la modélisation statistique, avec deux objectifs (1) amélioration de la qualité de prédiction d'un modèle existant (SUNFLO) et (2) évaluation de la contribution relative des variables référencées dans la littérature comme influençant l'élaboration de la qualité. Cette partie se limitera uniquement à la modélisation de la teneur en huile, par manque de données observées pour les protéines.

Trois types de modèles statistiques (multiple linéaire, additif généralisé (GAM), arbre hiérarchique (regression tree)) seront ainsi construits et comparés sur une base de données large (grand nombre de situations pédoclimatiques et de conduites culturales) et équilibrée (même nombre de situations pour chacune des modalités d'un même facteur). Le modèle de Pereyra-Irujo and Aguirrezábal (2007) sera également utilisé sur cette même base de données et sa performance sera comparée à celles des 3 autres modèles. Une liste de prédicteurs potentiels sera établie à partir de la littérature et du modèle SUNFLO (simulation de variables non mesurées au champ), puis elle sera réduite via des méthodes statistiques de simplification descendante (VIF et BIC). Enfin, la contribution relative des variables retenues dans chaque modèle pour la teneur en huile sera évaluée à l'aide de la méthode de partitionnement du coefficient de détermination « lmg » (« Lindeman, Merenda and Gold », Lindeman et al., 1980).

Les chapitres II et III traitent de l'analyse des effets de différents facteurs agronomiques (génotype, azote, densité de peuplement, eau) sur le puits (chapitre II) et/ou la source (chapitres II et III) à travers des expérimentations au champ et en serre.

Le chapitre II a pour objectif d'analyser les effets de la conduite de culture (choix de génotype, niveau d'alimentation azotée, densité de peuplement) sur la dynamique des « sources » (poids de tiges, feuilles, réceptacles) et des « puits » (poids d'huile, de protéines, de coques). Les dynamiques « sources » et « puits » issues d'expérimentations au champ (2011 et 2012) sont comparées (ANOVA) à l'aide d'un modèle simple bilinéaire à 4 paramètres, décrivant les vitesses d'accumulation (a1), de remobilisation (a2), les valeurs initiales des variables étudiées à la floraison (b1) et le temps thermique seuil correspondant à un changement significatif dans la dynamique d'évolution d'une source ou d'un puits (t1 : timing). Nous énonçons l'hypothèse sous-jacente selon laquelle les facteurs génotype et environnement devraient affecter différemment les sources et les puits, et que ce ne sont pas les mêmes paramètres qui sont affectés selon le facteur considéré.

Le chapitre III s'inscrit dans une démarche d'amélioration du modèle SUNFLO, s'agissant de la régulation génotypique des flux transpiratoires et de la photosynthèse en réponse à la contrainte hydrique. Ce paramètre génotypique est pris en compte dans le modèle SUNFLO mais mesuré sur des plantes en phase de préfloraison ; nous faisons l'hypothèse que le fonctionnement des plantes et leur réponse au stress hydrique en phase de post-floraison diffèrent de celles en préfloraison. Une meilleure connaissance de cette régulation en phase de remplissage est pourtant nécessaire car elle est déterminante pour l'élaboration de l'huile.

Une expérimentation en serre a été ainsi mise en place (2012) afin de mesurer la réponse au stress hydrique de 3 génotypes de tournesol en pré- et post-floraison (chapitre III). Une expérimentation antérieure (2009) portant sur les différences de régulation stomatique en fonction de l'âge des feuilles a permis de formuler l'hypothèse suivante : les différences de régulation transpiratoire avant et après floraison sont dues à des proportions différentes de types de feuilles dont le fonctionnement n'est pas le même en fonction de leur âge. De plus, nous souhaitions évaluer la sensibilité de la photosynthèse et de la transpiration à un déficit hydrique, en particulier lorsque celui-ci est appliqué pendant le remplissage des graines. Dans les deux expérimentations, la teneur en eau relative du sol (fraction of transpirable water, FTSW) est suivie de façon journalière, ainsi que le ratio entre les flux transpiratoires et la photosynthèse des plantes stressées (arrêt de l'irrigation) par rapport à celles maintenues sous irrigation.

Le chapitre IV sera consacré à la description du modèle dynamique d'élaboration de l'huile et des protéines chez le tournesol. C'est un modèle basé sur les relations source-puits issues des analyses des chapitres précédents. Le modèle décrit les flux de carbone et d'azote depuis les « sources » (tiges, feuilles, réceptacles) vers les « puits » (réceptacles, huile, protéines, coques), avec des règles de priorité définies par des fenêtres de temps thermiques et/ou d'un ordre prédéfini (chapitre II) régi par la loi du minimum entre source et puits (Tabourel-Tayot et Gastal, 1998a). Après une première étape de paramétrage du modèle (données 2012), nous procéderons à son évaluation (données 2011 à 2013) ainsi qu'à une analyse de sensibilité par la méthode de Morris (Morris, 1991 ; Campolongo et al., 2007).

Enfin, la dernière partie du manuscrit sera consacrée à un bilan des résultats importants de chacun des chapitres, et à la comparaison (avantages et inconvénients) des modélisations statistique et dynamique abordées durant la thèse.Tous les calculs et analyses statistiques, ainsi que la construction des modèles sont effectués sous R (R version 3.0.2).

Chapitre I: Prediction of sunflower grain oil concentration as a function of variety, crop management and environment using statistical models.

Article accepté le 5 Décembre 2013 dans la revue European Journal of Agronomy.

Andrianasolo, F. N., Casadebaig, P., Maza, E., Champolivier, L., Maury, P., & Debaeke, P. (2014). Prediction of sunflower grain oil concentration as a function of variety, crop management and environment using statistical models. *European Journal of Agronomy*, *54*, 84-96.

Prediction of sunflower grain oil concentration as a function of variety, crop management and environment using statistical models.

Fety Nambinina Andrianasolo[1,3,4], Pierre Casadebaig[1,4], Elie Maza[2,4], Luc Champolivier[3], Pierre Maury [1,4 a] , Philippe Debaeke [1,4 a]

[1]*INRA, UMR AGIR, BP 52627, 31326 Castanet-Tolosan cedex, France*

[2]*INRA, UMR GBF, BP 52627, 31326 Castanet-Tolosan cedex, France*

[3]*CETIOM, Centre INRA de Toulouse, BP 52627, 31326 Castanet-Tolosan cedex, France*

[4]*Université de Toulouse, INP, ENSAT, BP 52627, 31326 Castanet-Tolosan cedex, France*

[a] *Co-advisors of the first author PhD thesis*

Correspondence: fandrian@toulouse.inra.fr

Abstract

Sunflower (*Helianthus annuus* L.) raises as a competitive oilseed crop in the current environmentally-friendly context. To help targeting adequate management strategies, we explored statistical models as tools to understand and predict sunflower oil concentration. A trials database was built upon experiments carried out on a total of 61 varieties over the 2000-2011 period, grown in different locations in France under contrasting management conditions (nitrogen fertilization, water regime, plant density). 25 literature-based predictors of seed oil concentration were used to build 3 statistical models (multiple linear regression, Generalized Additive Model (GAM), regression tree (RT)) and compared to the reference simple one of Pereyra-Irujo and Aguirrezábal (2007) based on 3 variables. Performance of models was assessed by means of statistical indicators, including root mean squared error of prediction (RMSEP) and model efficiency (EF). GAM-based model performed best (RMSEP=1.95%; EF=0.71) while the simple model led to poor results in our database (RMSEP=3.33%; EF=0.09). We computed hierarchical contribution of predictors in each model by means of R^2 and concluded to the leading determination of potential oil concentration (OC), followed by post-flowering canopy functioning indicators (LAD2, MRUE2), plant nitrogen and water status and high temperatures effect. Diagnosis of error in the 4 statistical models and their domains of applicability are discussed. An improved statistical model (GAM-based) was proposed for sunflower oil prediction on a large panel of genotypes grown in contrasting environments.

Keywords: GAM, genotype by environment interaction, regression models, sunflower oil concentration

1. Introduction

Worldwide vegetable oil consumption is expected to grow by 2% per year as a result of increasing edible oil and renewable energy demands (FAO, 2012). In the 2011/2012 campaign however, oilseed grains production was greatly reduced because of adverse cropping conditions, then leading to a negative balance between supply and demand. The use of deemed tolerant oilseed crops, such as sunflower (*Helianthus annuus* L.), should be thus given consideration. The latter shows some agronomic and industrial potentialities (Ayerdi-Gotor et *al*., 2008; Aguirrezábal et *al.*, 2009; Pilorgé, 2010) as a promising competitive oilseed crop.

Sunflower cultivation could be particularly improved in France, where it is often grown in limited, shallow soils, non-irrigated and poor-nutrient sites (Debaeke et *al*., 2006; Casadebaig, 2008). In those situations, genotype x environment x management interactions were evidenced (Grieu et *al*., 2008) since genotypes do not exhibit the same strategies to cope with stress in restrictive conditions (Gallais, 1992; Denis and Vear, 1994).

Obtaining higher-oil concentration varieties appeared to be an alternative track for enhancing sunflower production, and could become a plus-value for French producers (Roche, 2005). Sunflower oil concentration was reported to be a conservative genetic component (Fick and Miller, 1997; Ruiz and Maddonni, 2006); however, recent studies highlighted differential responses of sunflower genotypes in contrasting cropping conditions; greater variability of oil concentration was whether linked to management and environmental conditions (Champolivier et *al*., 2011), or to genotypic and environment interactions (Andrianasolo et *al*., 2012). In both cases, a good understanding of oil concentration elaboration and effects of genotype and environmental factors raised to be essential for proposing convenient management strategies targeting both grain yield and oil content.

Sunflower oil is composed of 98% fatty acids (Berger et *al*., 2010; Echarte et *al.*, 2010), which are produced from two potential sources; main originates from post-flowering photosynthetic carbon (Merrien, 1992), supplemented with carbon assimilates stored in vegetative parts before flowering that will be remobilized thereafter (Hall et *al*., 1990; Merrien, 1992). Plant parts that provide carbon after flowering are considered as “source”

(source pool: leaves, stems) whereas those requiring carbon at this period are denoted "sink", namely grains. Reported determinants of sunflower oil concentration are genotype and environmental factors (Connor and Hall, 1997; Champolivier et *al*., 2011), among which intercepted radiation, nitrogen availability, high temperatures and water stress are often cited. These factors could play on both source and sink components, though only few studies explicitly separate effects on source and sink or make the link with oil concentration.

Genotype effect –*i.e* genotypes with intrinsic high or low-oil concentration- was described to play through kernel to hull proportion (López-Pereira et *al*., 2000; Izquierdo et *al*., 2008). At source level, genotype effect could play through contrasting strategies in mobilizing pre-flowering and post-flowering available carbon (Sadras et *al*., 1993).

Cumulative intercepted radiation between 250 and 450 degrees days after flowering was found to be the main determinant of oil concentration (R^2~80%) among sunflower hybrids in Argentine (Aguirrezábal et *al*., 2003). Higher plant densities could have a positive effect on source before flowering (Ferreira and Abreu, 2001) and on sink after flowering; Diepenbrock et *al*. (2001) suggested that the variation of oil concentration could be partly linked to negative impact of higher plant densities on final grain weights. However, Rizzardi et *al*. (1992) observed genotype x plant density interactive effects on final oil concentrations when comparing two contrasting genotypes.

Nitrogen effect is often described through the negative relationship between oil and protein concentration (Connor and Sadras, 1992); highest oil concentrations were met in non-fertilized treatments. Nitrogen doses that are brought during vegetative period permit to optimize dry matter at flowering (Hocking and Steer, 1983) thus potential quantity of mobilized pre-flowering assimilates during grain-filling.

High temperatures after flowering were reported to shorten grain filling duration; depending on authors, we identified various temperature thresholds: 30°C (Aguirrezábal et *al*., 2009), 34°C maximum temperatures (Chimenti et *al*., 2001; Rondanini et *al*., 2003) or 17°C mean temperature (Angeloni et *al*., 2012).

Little evidence exists about the effect of water availability on oil concentration; Santonoceto et *al*. (2003) observed significant differences in oil concentration in the final phase of oil accumulation under water stress, with an obvious lower rate of grain oil accumulation for non-irrigated modality. Before flowering, water stress could affect leaf expansion

(Casadebaig et *al*., 2008), while it could limit green leaves photosynthesis and duration in post-flowering period (Aguirrezabal et *al*., 2009).

Literature-based knowledge about sunflower oil concentration determination is illustrated in a schematic conceptual framework (Fig.1).

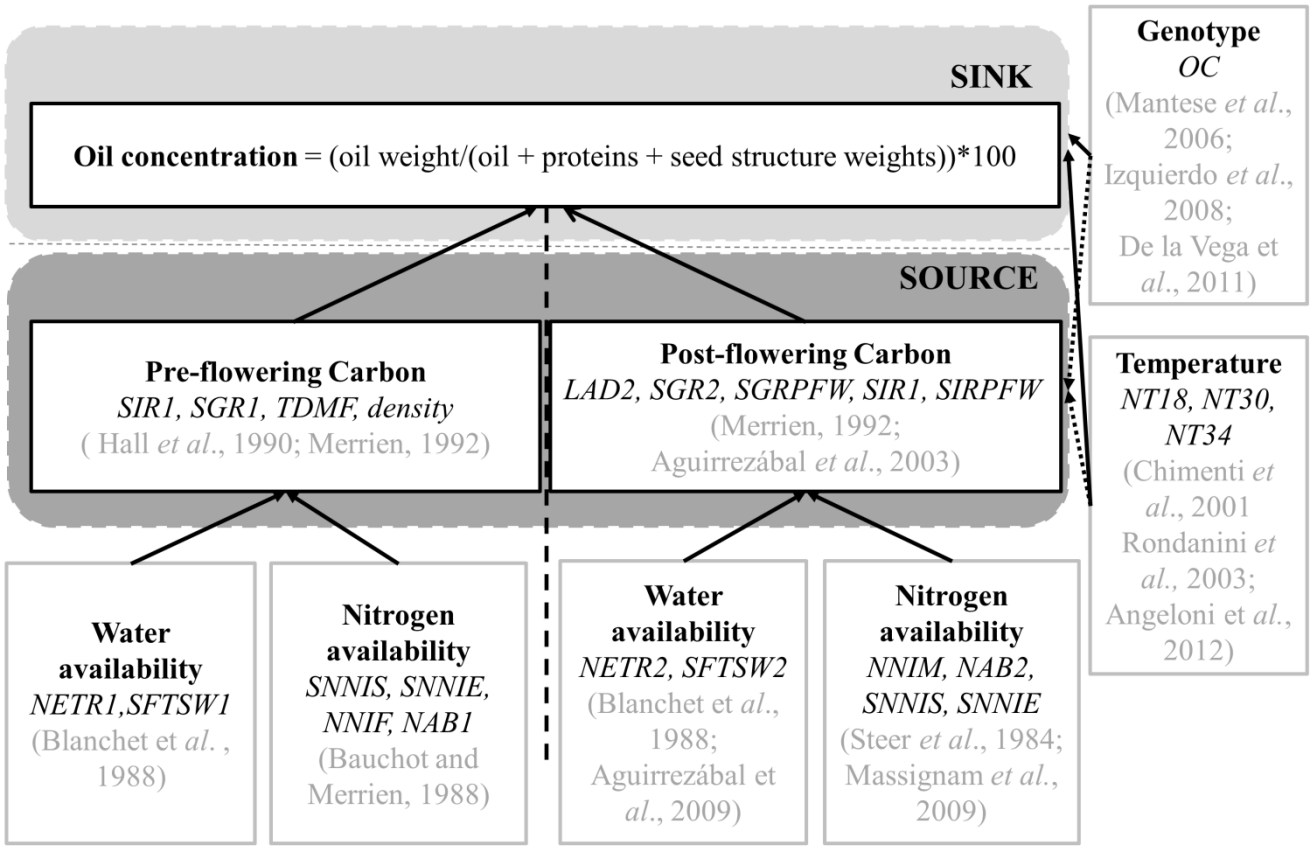


Fig.1. Schematic framework of sunflower oil concentration elaboration as described in section 1 and relative selected predictors used for statistical modeling. Meanings of abbreviations are given in Table.2. Continuous arrows indicate literature reported relationships which were used to compute the selected predictors. Dotted arrows indicate known relationships that were not used in this study.

To help understanding crop physiology and yield determinism, crop models are tools that are increasingly developed. These can be used for multiple purposes, either to describing complex biological systems, or to interpreting experimental results, making a diagnosis of limiting factors and providing advices and predictions towards farmers for better crop and policy management (Boote et *al*., 1996). Statistical/empirical models, particularly, have been of great use in the history of science. Their easiness of computing and usability enhanced their attractiveness among decision-makers and practitioners (Razi and Athappilly, 2005), while they allow highlighting relative importance of variables when much is uncertain (Lobell et *al*., 2005; Tittonell et *al*., 2008; Tulbure et *al*., 2012). Statistical models could be divided into two main subgroups: parametric and non-parametric. Parametric models (ex: simple or multiple linear regression) have the advantage to be quantifiable, and assessable, but the form of the relationship between dependent and independent variable(s) should be known *a priori* to avoid misleading results; non-parametric ones (ex: GAM, regression trees, neural networks) do not assume neither any *a priori* model structure nor any formal distribution of the data. They permit to bring out non-linear relationships but often lead to heavy parameterized models. Wullschleger et *al*. (2010) used non-parametric models to establish equations of parametric ones for switchgrass yield prediction. Other non-parametric models (regression trees, Breiman et *al*., 1984) were utilized to analyze yield variability in maize (Tittonell et *al*., 2008), wheat (Lobell et *al*., 2005), soybean (Zheng et *al*., 2009), sugarcane (Ferraro et *al*., 2009) or switchgrass (Wullschleger et *al.*, 2010; Tulbure et *al*., 2012).

Few statistical models exist for seed oil prediction; those existing are mostly parametric. For instance, multiple linear regressions were used to model palm oil (Khamis et *al.*, 2006; Keong and Keng, 2012), though their predictive performances were not assessed. For sunflower, a non-linear empirical model was established by Pereyra-Irujo and Aguirrezábal (2007) relating actual oil concentration to genotypic oil concentration, radiation cumulated during the post-flowering specific period (Aguirrezábal et *al*., 2003) and plant density. However, the model was parameterized in sites where nitrogen was non-limiting, and where water stress could be likely moderate or non-existing.

For specifically predicting oil concentration, the crop model SUNFLO (Casadebaig et *al*., 2011) uses a multiple linear regression model linking oil concentration with some simulated genotype, environmental stress and post-flowering canopy functioning indicators. Following oil model evaluation on an independent dataset, it was hypothesized that the acceptable

though improvable RMSEP (Predictive root mean squared error: ~4 oil points) was due to the narrowness of ranges of situations represented in the database, and the choice of predictors that failed to take into account physiologically-based responses of sunflower.

Therefore, the objectives of this paper are the following: (1) build statistical models based on physiologically-sound predictors and compare their predictive performance for sunflower grain oil concentration on a large dataset; (2) highlight essential features of grain oil elaboration by assessing variable importance and unraveling interactions; (3) compare the performance of these statistical models with the reference one from Pereyra-Irujo and Aguirrezábal (2007). The latter was chosen as reference model given its simplicity (low number of variables, simple equation), easiness of use (variables that can be simulated by pre-existing model SUNFLO) and physiological-basis relevance of variables.

We proceeded similarly to Casadebaig et *al*. (2011) by providing model inputs to obtain simulated predictors, and include the latter into different regression models of sunflower oil concentration, while following principle of parsimony simplification approach (Crawley, 2012).

2. Materials and methods

2.1. Dataset collection

We collected sunflower oil concentration data from various French experiments conducted from 2000 to 2011 by CETIOM and INRA institutes, covering South-West to Middle-East French regions with 18 experimental sites and 61 commercial varieties in total. The whole dataset comprised 418 units of simulation (USM): each USM corresponds to one plot describing a site (soil), a growing season, a crop management and a genotype. Based on the factors studied, we established 6 categories of trials: nitrogen fertilization trials (N.trials), water regime trials (W.trials), plant density trials (D.trials), variety assessment trials (V.trials) and trials where factors were combined: nitrogen and water (NxW.trials) and nitrogen x plant density (DxN.trials). A trial was considered as a combination of experimental treatments in a given site x year. In all trials, sunflower oil concentration was measured by MNR (Magnetic Nuclear Resonance) on a subsample of seeds and expressed at equivalent 0% moisture. Information about trials, number of USM and agronomic factors is summarized in Table.1.

Table.1. Summary table of dataset trials types and corresponding number of units of simulation (USM), genotypes and sites, within the dataset. N., W., D., and V. correspond to nitrogen, water, density and variety respectively.

Trials type	Number of USM	% of whole dataset	Modalities range and number	Number of genotypes	Number of sites
N.trials	63	15	from 0 to 160 kg N per ha 7 modalities	6	7
W.trials	6	1	rainfed and irrigated (160 and 200 mm) 3 modalities	1	1
D.trials	24	6	from 3 to 8 plants per m² D1 to D6: 6 modalities	2	2
V.trials	273	65	from 8 to 20 varieties per site	61	8
N. x W.trials	24	6	from rainfed x 0N to irrigated x160N 12 combinations modalities	2	2
D. x N.trials	28	7	from 4.8 to 6.8 plants per m² and 0N to 160N 10 combinations modalities	2 to 8	1

2.2. Simulation of oil concentration predictors

SUNFLO model was used to simulate indicators that constituted our putative predictors for modeling. These predictors were simulated by using the previous database as input data. Requested inputs for running SUNFLO dynamic model were available in most of the trials; where appropriate, experts' advice was followed when missing data. These concerned less than 10 USM.

Climatic weather stations located within a 15 km distance from trials, provided the following meteorological data: rainfall (mm), minimum and maximum temperatures (T, °C),

evapotranspiration (ET, mm) and global radiation (GR, $MJ.m^{-2}$). Soil water availability, as a function of soil deepness and stoniness, and residual nitrogen amounts at sowing were measured by experimenters in fields. Genotypic information in SUNFLO included phenology, canopy architecture, water stress response, potential harvest index and potential oil concentration (Debaeke et *al*., 2010). Particularly, potential oil concentration was measured in an independent set of trials from CETIOM (Champolivier, *personal communication*) and computed as the maximum observed oil concentration for a given variety on a range of sites and years. Dates and rates of N fertilization and irrigation were provided for all trials, as well as planting density at emergence. Pests and diseases were adequately controlled in all experiments.

We illustrated the variability of observed oil concentration as related to genotypes, environments and management practices diversity (Fig.2).

2.3. Choice of putative predictors for oil concentration model

Based on physiological processes and determining factors identified in literature, we chose indicators describing pre- and post-flowering periods that were related to environmental resources, canopy general functioning, nitrogen and water-linked indicators of plant state, and specific genotype characteristics (Fig.1). Most of them were simulated by SUNFLO model since they were not measured in past field experiments. Assuming that intermediate and final variables simulated by SUNFLO have been already evaluated and considered as acceptable (Debaeke et *al.*, 2010; Casadebaig et *al*., 2011), we used our 25 indicators as putative predictors for sunflower oil concentration statistical model. Information about indicators is provided in Table.2.

Table.2. List of predictor variables used to build statistical models for sunflower oil concentration, selected according to their literature-relevance characteristics, and simulated by SUNFLO model. Ranges of variation in the dataset and variables units are provided.

Categories of predictors	Predictors	Meaning	Range	Units
Environmental resources				
	SGR1	sum of global radiation during vegetative period	686 to	MJ/m²
	SGR2	sum of global radiation during reproductive	509 to 916	MJ/m²
	SGRPFW	sum of global radiation between 250 and 450 degree days after flowering	202 to 313	MJ/m²
Environmental constraints				
Water stress	NETR1	number of days with water stress (real to maximum evapotranspiration ratio lower than	0 to 27	days
	NETR2	number of days with water stress (real to maximum evapotranspiration ratio lower than 0.6) during reproductive period	0 to 38	days
	SFTSW1	sum of 1 – (fraction of transpirable soil water)	7 to 34	-
	SFTSW2	sum of 1 – (fraction of transpirable soil water) during reproductive period	20 to 39	-
Nitrogen stress	NAB1	sum of nitrogen quantities absorbed by plant in vegetative period	20 to 172	kg/ ha
	NAB2	sum of nitrogen quantities absorbed by plant in reproductive period	7 to 60	kg /ha
	NNIF	nitrogen nutrition index at flowering	0.39 to	-
	NNIM	nitrogen nutrition index at the beginning of grain filling	0.39 to 1.42	-
	SNNIE	Integration of nitrogen nutrition index when the latter exceeds the value of 1, computed on the	0 to 34	-
	SNNIS	Integration of nitrogen nutrition index when the latter is lower than 1, computed on the whole	0 to 39	-
Thermal stress	NT18	Number of days during which seed filling period mean air temperature is higher than 18°C	21 to 39	days
	NT30	Number of days during which seed filling period maximum air temperature is higher than 30°C	0 to 26	days
	NT34	Number of days during which seed filling period maximum air temperature is higher than 34°C	0 to 13	days
Canopy functioning				
	LAD2	Leaf area duration in reproductive period	24 to 122	m² days/m²
	SIR1	Sum of intercepted radiation during vegetative	244 to 454	MJ/m²
	SIR2	Sum of intercepted radiation during reproductive period	149 to 352	MJ/m²
	SIRPFW	Sum of intercepted radiation between 250 and 450 degree days after flowering	62 to 139	MJ/m²
	MRUE2	Mean radiation use efficiency during	0.02 to	g/MJ
	MRUEPFW	Mean radiation use efficiency during 250 to 450 degree days	0.06 to 0.73	g/MJ
	TDMF	Total aerial dry matter at flowering	311 to 713	g/m²
Management				
	density	Plant density at emergence	3 to 8.2	plants/m²
Genotype				
	OC	Potential oil concentration	47.7 to 60.8	%

2.4. Filtering USM and predictors

2.4.1. Yield difference threshold USM filtering

Before starting oil concentration modeling, we checked the goodness of fit between simulated and observed grain yields in our dataset. We assumed that in situations where SUNFLO model lacked precision to simulate yield, indicators would suffer the same imprecisions. Therefore, we decided to exclude units of simulation where the difference between observed and simulated yields was beyond a given threshold. This threshold (10 quintals per ha) was set according to the observed variability in yield differences in the dataset, and described in results section (Fig.2).

2.4.2. Reducing multi-collinearity by deleting some putative predictors

We drew particular attention in detecting possible multi-collinear variables among our predictors, which would impact the reliability of our statistical models (Dormann et *al*., 2013). Following the method suggested by Zuur et *al*. (2010), we computed the variation inflation factor (VIF) and applied a stepwise deletion of predictors according to decreasing VIF values, until a threshold of 2. We assumed that remaining predictors contained essential information so that the dropped ones were only redundant predictors.

2.5. Statistical models building

For practical purposes, we numbered the statistical models that were progressively built from 1 to 4: Pereyra-Irujo and Aguirrezábal (2007) adjusted model, multiple linear regression using BIC-stepwise selection, GAM-wised transformed and regression tree model.

2.5.1. Model.1: Pereyra-Irujo and Aguirrezábal (2007)-adjusted model

We used the equation provided in Pereyra-Irujo and Aguirrezábal (2007) for simulating sunflower oil concentration, and applied their formula to our dataset. 3 predictors were used: SIRPFW (sum of intercepted radiation between 250 and 450 °Cd after flowering), OC (potential oil concentration), and plant density:

$$OCobs = \min\left\{a + b \times \left(\frac{SIRPFW}{density}\right), OC\right\},$$

where *a* and *b* are model parameters and correspond to intercept and slope of the linear part of the equation, respectively. Pereyra-Irujo and Aguirrezábal (2007) potential oil concentration was set at a maximum value of 50%, which was shown to be valid for many sunflower hybrids. However, 50% was quite low regarding our potential oil concentrations range, so we re-estimated model parameters by the use of *nls* (non-linear least squares) function of basic R to adjust to our data.

2.5.2. Model.2: Multiple linear regression (MLR) and stepwise selection by BIC

Following the method of Casadebaig et *al*. (2011), we built an additive multiple linear regression model (MLR) with the non-dropped predictors. We then carried out a stepwise forward variables selection based on BIC (Bayesian Information Criterion) with the help of *stepAIC* function from "MASS" package in basic R (R-3.0.1 software version).

2.5.3. Model.3: Generalized Additive Model (GAM) and predictors transformations

Generalized additive models (GAM) are non-parametric models that fit to data by means of smoothing functions based on local regression splines (Wood, 2003). They are generally used to visualize possible non-linear relationships between dependent and independent variables and to check the improvement in predictive performance in case non-linear relationships were detected (Wood, 2004; Wullschleger et *al*., 2010). We fit our smallest current statistical model with the *gam* function of R "mgcv" package (Wood, 2004).

We went further into investigation by checking possible equations that matched the transformations of predictors suggested by GAM –*i.e*, parameterizing the model. For this, we used Formulize Eureqa version 0.98 Beta software (Schmidt and Lipson, 2009; Schmidt and Lipson, 2013). Various possible fitting curves were obtained; we chose equations with goodness of fit (R^2) to data higher than 98%. In any case of having several possible equations with $R^2 > 98\%$, we chose the one with the less parameters. Parameters values were proposed by the software, which we used as initial starting guesses parameters for *nls* regression in R.

2.5.4. Model.4: Regression tree model

Regression tree (RT) is a non-parametric model that splits hierarchically continuous dependent variable into nodes in a binary way (Breiman et *al*., 1984). Splits are obtained using a recursive partitioning algorithm, where predictors appear from the one most contributing to the variance of the response variable to the least contributing. We used regression tree in order to (1) assess relative importance of variables with no assumption of linearity, (2) identify possible interactions, which we did not willingly include in our previous statistical models. *rpart* function or R "rpart" package was used for fitting RT (Breiman et *al*., 1984).

2.6. Statistical models evaluation and diagnosis

Performances of models were evaluated and compared according to their goodness of fit to data, predictive quality and adequacy in simulated patterns for some agronomic trials. We also computed relative variable (predictor) contribution to simulated oil concentration in each model.

2.6.1. Goodness of fit: R^2, EF

All statistical models were first evaluated for their goodness of fit to data, by computing coefficient of determination (R^2) and model efficiency (EF).

2.6.2. Predictive performance and error diagnosis: RMSEP, SDSD, SB, LCS

Then, statistical models were evaluated for their predictive performance (RMSEP) by launching leave-one-out cross-validation (LOOCV) for linear models, using *cv.lm* function of "DAAG" package of R (Maindonald and Braun, 2010). LOOCV involves using a single observation from the whole dataset as the validation set, and the remaining observations as the training set; the process is repeated such that each observation in the dataset is used once as a validation set. For GAM, ML (maximum-likelihood) method was used for model fitting, and GCV (Generalized Cross Validation) for model evaluation (Wood, 2003). For regression tree, cross-validation was used as a standard method for evaluating predictive performance with the help of *xpred.rpart* function of "rpart" package (Breiman et *al*., 1984). For the non-

linear model adjusted from Pereyra-Irujo and Aguirrezábal (2007), we ran a LOOCV with the help of *cross.val* function of "R330" package (Lee and Roberston, 2012).

We then split global MSE into components (Kobayashi and Salam, 2000) that could bring more information on understanding of the model type of error. Components were SDSD (Squared Difference between Standard Deviations), SB (Squared Bias) and LCS (Lack of Correlation between Standard deviations). High SDSD (magnitude) and LCS (pattern) values would suggest that a given statistical model fails to simulate the variability of measurements around the mean. High SB values originate from systematic behavior of the model errors.

2.6.3. Comparing response patterns to varying management practices

Simulated patterns of oil concentrations responses to varying management practices (such as nitrogen fertilization and plant density) were compared to observed ones for each statistical model. These concerned D. and N.trials.

2.6.4. Variable importance computation

We assessed relative variable importance in each statistical model by using *calc.relimp* function of R "relaimpo" package (Grömping, 2006). This function computes coefficient of determination of each variable by partitioning total model R^2 while averaging over orders. For regression tree, variable importance was automatically computed by decomposing variance and then scaled to 100%. Model.1 variable importance was assessed by calculating Sobol indice (Sobol', 2001) with Monte Carlo Sobol sensitivity analysis (*sobol* function in sensitivity package; Saltelli et *al*., 2000).

3. Results

3.1. Dataset diversity: cropping conditions and observed oil concentrations variability

We illustrated dataset richness and diversity by computing ranges of variations of predictors describing cropping conditions and crop states (Table.2), as well as observed oil concentrations (OCobs) variations (Fig.2). Sum of global radiation during post-flowering period varied from 509 to 916 MJ/m² in the dataset. Regarding environmental constraints,

water stress days indicator (NETR2) ranged from 0 to 38 days, while nitrogen nutrition index at flowering (NNIF) varied from 0.4 to 1.4. High temperatures stressing days (NT34) reached up to 13 days in some trials.

Minimum and maximum values of observed oil concentrations (OCobs) for the modalities we selected were 40.8 and 56.9% respectively, but reached up to 59.4% in a non-illustrated modality (Fig.2). For nitrogen, density and genotype modalities illustrated here, OCobs amplitudes were 7.1, 6.8 and 10.9 oil points respectively. Per variety, OCobs range was from 0.32 to 3.72 oil points. Potential oil concentrations (OC) varied from 47.7 to 60.8% (Table.2).

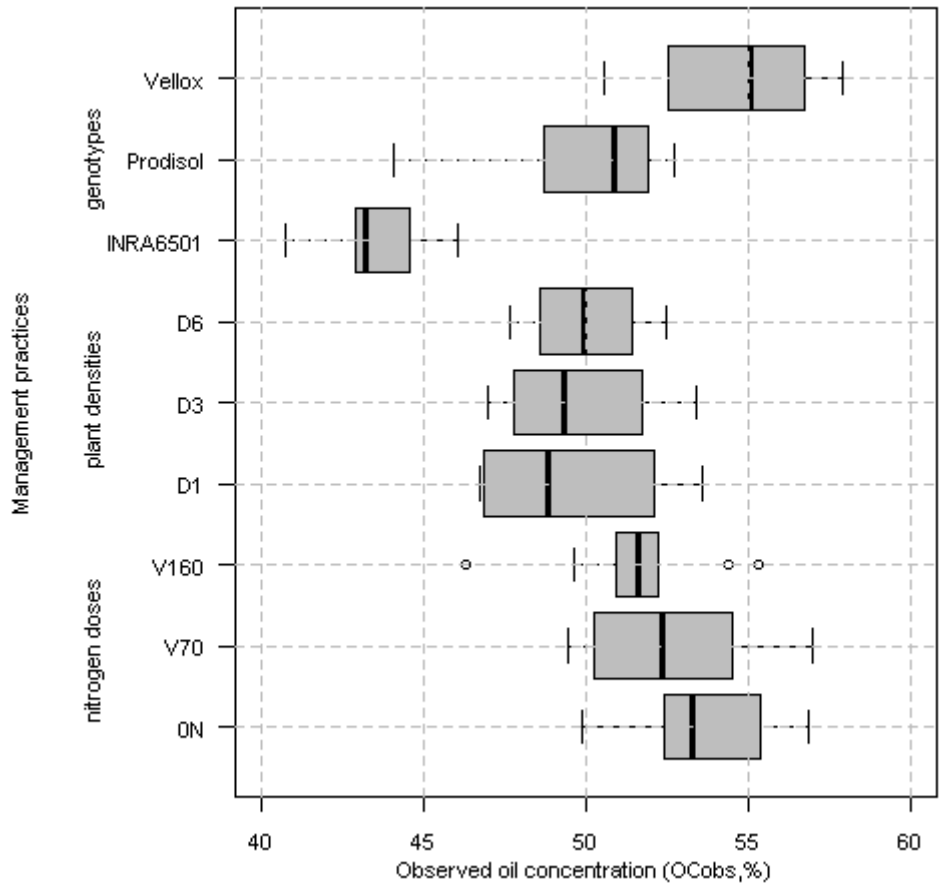


Fig.2. Variability of observed oil concentration (OCobs) related to variability of management practices (choice of genotypes, plant densities and nitrogen doses) in dataset. 3 most contrasted modalities were picked in each management practice for illustration purposes. Nitrogen doses varied from 0N (no fertilization) to V160 (160 kg N per ha brought during vegetative stage). Plant densities range was 3 (D1) to 8 plants per m² (D6). We selected genotypes (INRA6501, Prodisol and Vellox) based on their contrasted potential oil concentration (OC).

3.2. Yield threshold

Differences between SUNFLO simulated and observed grain yields varied from 0.002 to 23.58 q per ha (Fig.3). Though, there were only few USM that were concerned by high differences (higher than 10 quintals per ha). These corresponded to about 10% of total dataset. We then decided to exclude all USM which yield difference was equal or higher than this threshold. Remaining USM totalized 374.

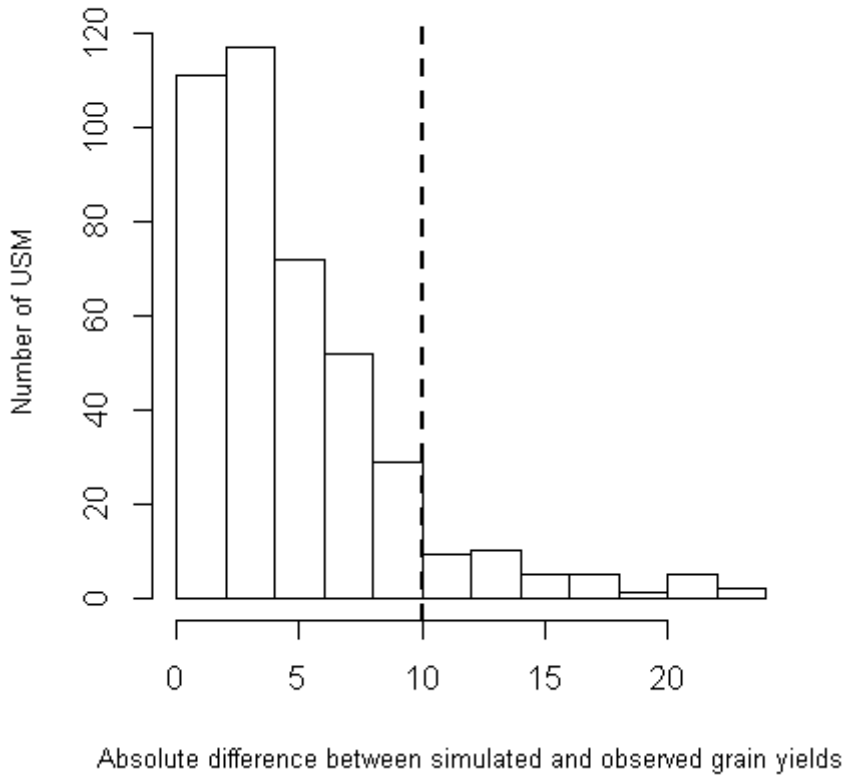


Fig.3. Histogram of number of units of simulations (USM) as a function of absolute differences between SUNFLO simulated and observed grain yields. Dashed vertical line corresponded to the threshold (10 quintals per ha) chosen for excluding some USM from the dataset.

3.3. Statistical models building

3.3.1. Model.1: Pereyra-Irujo and Aguirrezábal (2007) adjusted model

Using Pereyra-Irujo and Aguirrezábal (2007) equation, we re-estimated the parameters *a* and *b* which initial values were 36.4 and 0.5 respectively. This led to the following adjusted values: $a = 48.06$ and $b = 0.17$.

3.3.2. Model.2: Multiple Linear Regression (MLR) and stepwise selection by BIC

There remained 12 predictors (out of 25) after VIF stepwise method for deleting multi-collinear variables. After BIC stepwise model selection, 9 predictors were retained. These were potential oil concentration (OC), water stress indicators (SFTW1, SFTSW2), nitrogen status indicators (SNNIE, NAB2), thermal stress (NT34), canopy functioning after flowering (LAD2, MRUE2) and management practice (density) predictors. Coefficients values were 0.08, 0.11, -0.05, -0.19, 0.03, 0.02, 27.6, 0.65 and 0.97 for SFTSW1, SFTSW2, NAB2, SNNIE, NT34, LAD2, MRUE2, density and OC respectively, and -17.96 for intercept.

3.3.3. Model.3: Generalized Additive Model (GAM) and predictors transformation

The previous 9 predictors-model, being the smallest one we got from a stepwise deletion process, was used in GAM to be compared to the linear one. Notation "s()" corresponds to the transformed values of each predictor (Fig.4). We first extracted transformation equations by the help of Formulize Eureqa software before plotting observed oil concentration (OCobs) with each predictor and their corresponding transformed values.

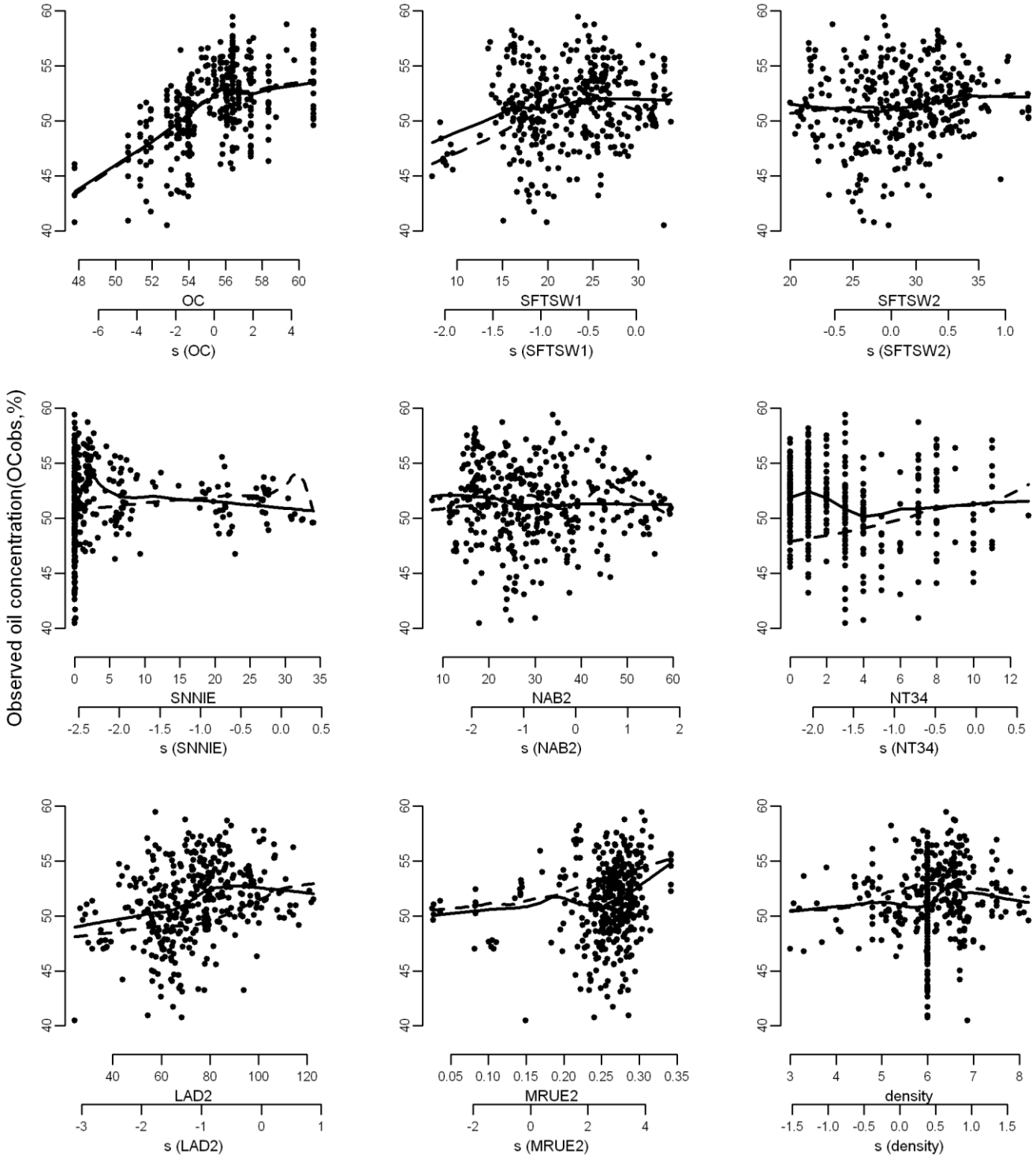


Fig.4. Relationships between observed oil concentrations (OCobs) with each raw and GAM-transformed (prefixed "s ( )") predictor respectively. Predictors were those that were selected by stepwise process in the multiple linear regression model (Model.2). Dots correspond to raw data. Continuous black (raw data smoothing) and dashed grey (transformed data smoothing) lines were obtained using *lowess* functions of R. Upper (raw data) and lower scales (transformed data) are indicated.

3.3.4. Model.4: Regression tree model

Regression tree is illustrated in Fig.5. The tree was highly branched (up to 8 splitting nodes) and demonstrated relatively high levels of predictors' interactions in explaining observed oil concentrations in our dataset.

The main splitting knot was linked to potential oil concentration (OC); OCobs variability of varieties having their OC lower than 54.4% (left part of the tree) was mostly linked to OC, MRUE2, SGR1, SFTSW1 and density. For those displaying higher OC and low values of SGR1 (< 763.2 MJ/m²), OCobs depended on OC, SGR1 and LAD2. If else, OCobs depended on interactions between cited predictors and MRUE2, SGRPFW, SNNIE, NAB2, and SFTSW1. 19 groups of OCobs dependencies were obtained at lower branches of the regression tree.

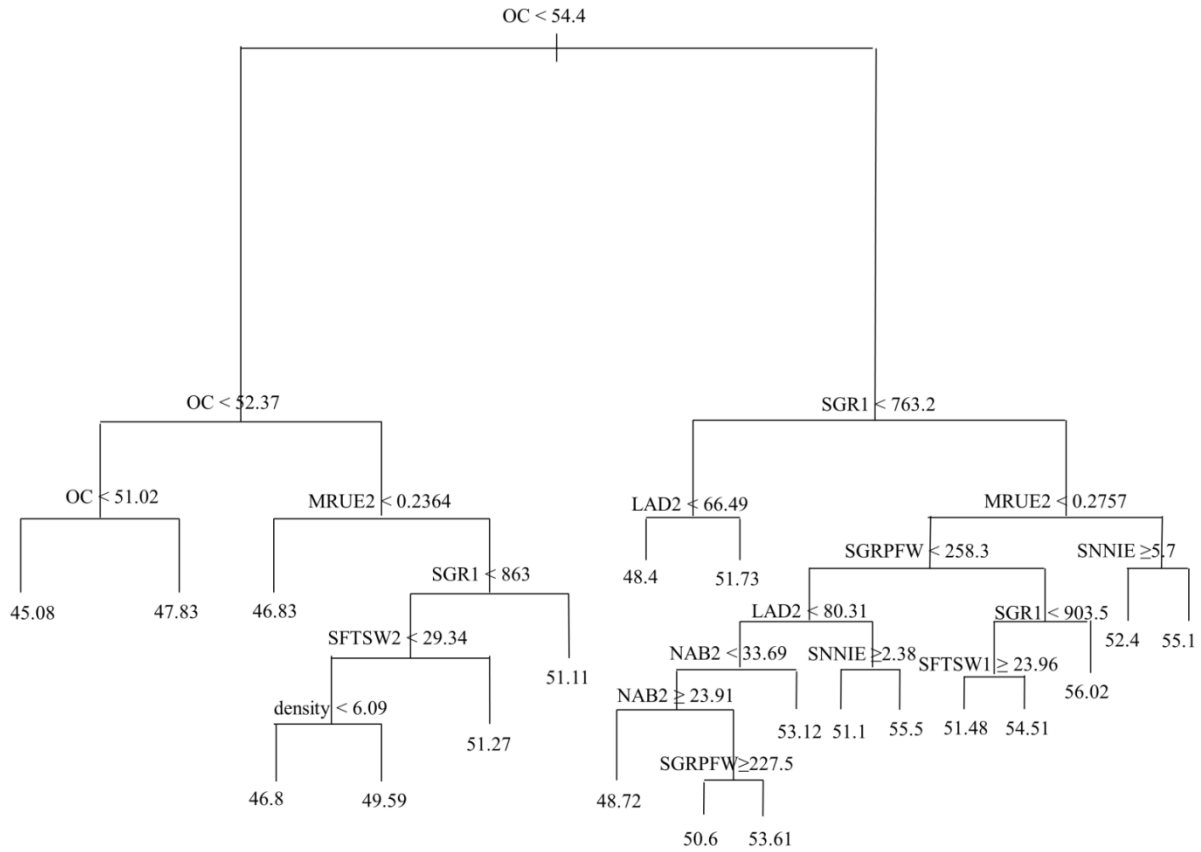


Fig.5. Regression tree model of observed oil concentration (OCobs) as related to its most contributing predictors from the non-stepwise BIC selected initial model (12 predictors). Mean values of OCobs are represented at final ends of lower-branches. Predictors are hierarchically positioned along the branches; nodes correspond to thresholds splitting values in binary way.

3.4. Comparative performances of statistical models

3.4.1. Goodness of fit, predictive performances and error diagnosis

Table.3 is a summary table of fits and predictive performances. Best fits and predictive performances were obtained with Model.3 (R^2 =71%; RMSEP=1.95 oil points). Model.1 was the less efficient regarding its EF value (10%) and highest RMSEP (3.33 oil points). Multiple linear regression (Model.2) performed worse than Model.4 for goodness of fit to data (R^2 = 53% against 70% respectively), but better than regression tree for predictive performance (2.41 and 2.54 for RMSEP values respectively). Models biases values were all close to 0, despite being negative in Model.1 (Bias=-0.16). Graphical illustrations of simulated and observed oil concentrations relationships are provided in Fig.6. Referring to first bisector, points of Model.1 were located on an horizontal line, while those of Model.2 and 4 were scattered along the bisector line. Model.3 displayed closest scatterplot to the 1:1 line.

Error was found to be linked to LCS component in all models (contribution varying from 86 to 95% of total mean squared error), except in Model.1 where it was rather linked to SDSD (contribution of 71%), and little to LCS (28%). Highest LCS was obtained in Model.4 (regression tree). SDSD contribution to error was relatively low in other models, but its contribution increased from Model.4, 3, to 2.

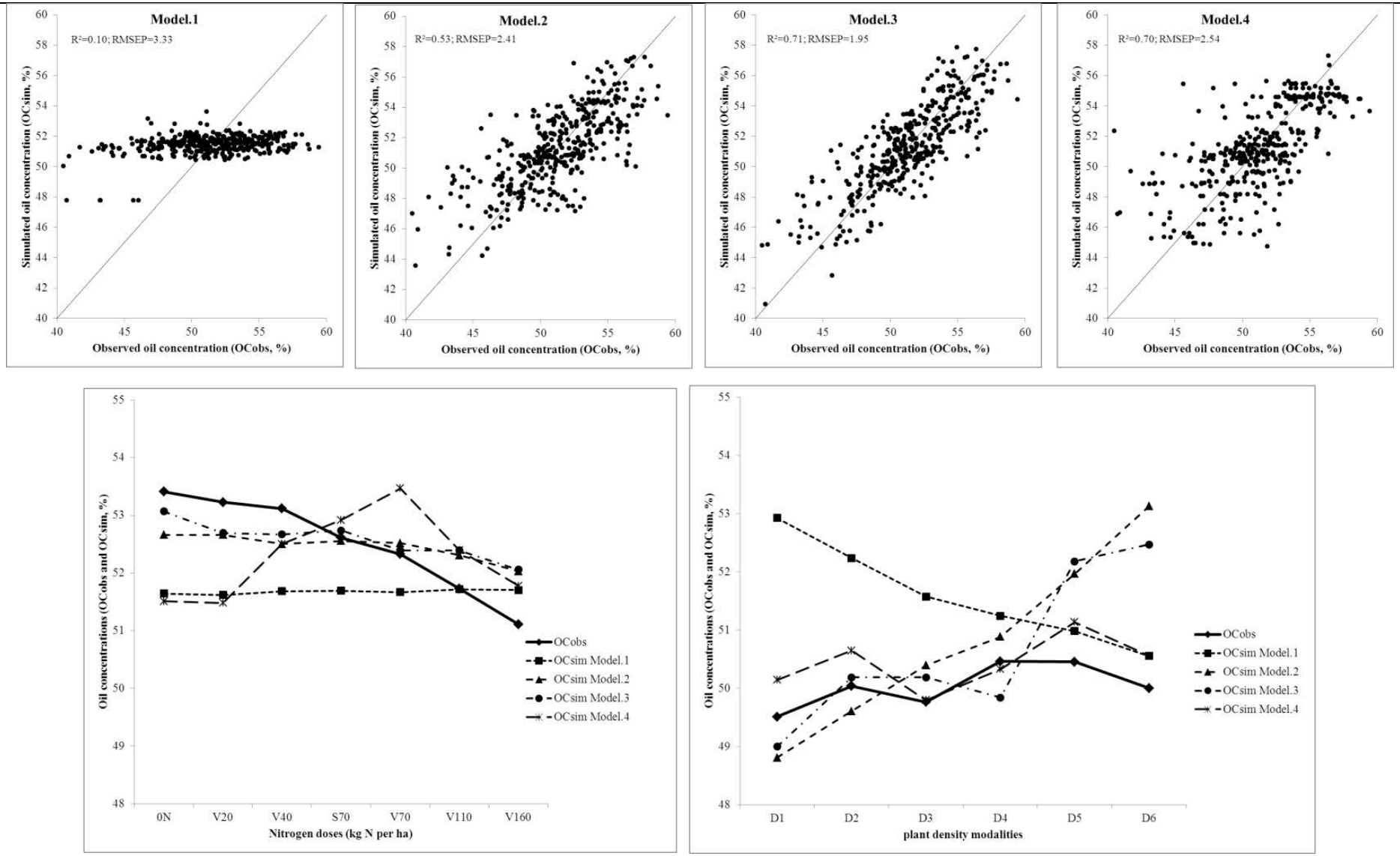


Fig.6. Graphical patterns of simulated oil concentrations (OCsim) plotted against observed ones (OCobs). Upper line displays global scatterplot of each model, while lower line focuses on oil concentrations patterns in some agronomic trials (N. and D.trials from left to right).R^2 and RMSEP are provided as indicators of global performances of each model. Dynamics are plotted such that, for growing amounts of each factor, we could easily visualize oil concentrations patterns.

Table.3. Fit and prediction performances indicators of built sunflower oil concentration statistical models, averaged across all trials. Bias was measured from differences between observed and simulated oil concentrations. Coefficients of determination (R^2) and model efficiency (EF) gave equal values and are expressed on the 0 to 1 scale. RMSEP (Root mean squared error of prediction) was computed using k-fold cross-validation. For error diagnosis, we decomposed mean squared error into SB (Squared bias), SDSD (Squared Difference of standards deviations) and LCS (Lack of correlation between standard deviations). Models are numbered from 1 to 4 and correspond to adjusted Pereyra-Irujo and Aguirrezábal (2007), BIC-stepwise selected, GAM-based and regression tree models respectively.

Model	Bias	EF/R^2	RMSEP	SB	SDSD	LCS
Model.1	-0.16	0.09	3.33	0.03	7.96	3.16
Model.2	0.00	0.53	2.41	0.00	0.80	4.96
Model.3	0.01	0.71	1.95	0.10	0.32	3.48
Model.4	0.06	0.70	2.54	0.20	0.28	6.61

3.4.2. Variable importance comparison

We computed relative variable importance for each statistical model and compared rankings (Table.4). In all models, the most contributing variable to observed oil concentration was the potential one (OC): from 25 to 56% in Model.4 to 2, except in Model.1 where SIRPFW ranked first (88%). In this model, density and potential oil concentration had similar weights (7 and 5% respectively).

In the other models, ranking differed from second place. MRUE2 ranked second (~12%) in Model.2 and 3, while it was SNNIE (15%) in regression tree (Model.4). LAD2 had similar relative importance as MRUE2 (~11%) in the models 2 and 3; SNNIE indicator was followed by SGR1 (10%) in Model.4.

4th most important variable was found to be water stress (SFTSW1~5%) in Model.2, while it was density (8%) in Model.3 and post-flowering global radiation in Model.4 (SGRPFW=8%). Thermal stress was also accounted for in the models 2 and 3 (4 and 7%). There were lower contributions of other predictors in the models 2 and 3 (from 2 to 5%) and in Model.4 (from 4 to 7%).

Table.4. Relative variables contribution (in % R^2 of total oil concentration variation) of retained predictors in each of the statistical models. Models are numbered from 1 to 4 and correspond to adjusted Pereyra-Irujo and Aguirrezábal (2007), BIC-stepwise selected, GAM-based and regression tree models respectively.

Categories of predictors	Predictors	Model.1	Model.2	Model.3	Model.4
Environmental resources	SGR1	-	-	-	10.0
	SGRPFW	-	-	-	8.0
Water stress	SFTSW1	-	5.0	3.5	7.0
	SFTSW2	-	2.0	4.5	7.0
Nitrogen stress	NAB2	-	2.0	3.0	6.0
	SNNIE	-	3.5	2.0	15.0
Thermal stress	NT18	-	-	-	4.0
	NT34	-	4.0	7.0	1.0
Canopy functioning	LAD2	-	11.5	12.0	12.0
	SIRPFW	88.0	-	-	-
	MRUEPFW	-	-	-	-
	MRUE2	-	12.0	13.0	6.0
management	density	7.0	4.0	8.0	4.0
genotype	OC	5.0	56.0	47.0	25.0

3.5. Patterns in response to management practices

We proposed to compare patterns of simulated (OCsim) and observed (OCobs) oil concentrations in some agronomic trials, *e.g* D. and N.trials. We computed mean values of observed and simulated oil concentrations per modality of each agronomic factor (Fig.6), and plotted dynamics of OCobs and OCsim against growing levels (amounts) for each model.

3.5.1. N.trials oil concentration patterns

There were three phases in observed oil concentrations patterns in response to growing nitrogen fertilization doses; OCobs stagnated between 0N to V40 (~53.5%), then slightly decreased between V40 to V70 (by 1 oil point) and sharply decreased thereafter (from 52.5 to 51%). Model.1 showed no response to growing nitrogen doses (stagnating 51.5% value). OCsim by Model.2 and 3 displayed very close patterns; those models were able to simulate only a slight oil concentration decrease at highest dose (less than 0.5 oil points); their global behavior in response to nitrogen was a stagnating oil concentration. Model.4 described a sharper decrease of OCsim starting from V70 (from 53.5 to 51.5%) compared to OCobs, but the other nitrogen modalities were badly simulated (sharp increase of 2 oil points between V20 and V70 and stagnation from 0 to V20).

3.5.2. D.trials oil concentration patterns

Mean OCobs increased slightly from D1 to D2 (from 3 to 4 plants per m², it varied by 0.5 oil points) and reached up to 50.5% at D4 modality. Its mean value stagnated between D4 and D5 modalities, then decreased by 0.5 oil points at highest plant density (D6). There was a sharp decrease of OCsim from 53% to 51% with increasing plant density in Model.1, while it was the opposite trend in Model.2 (from 49 to 53%). Model.3 had similar patterns as OCobs but only between D1 and D2 modalities. Model.4 displayed the closest pattern to observed data, though values differed up to 0.5 oil points.

4. Discussion

Models building and methods of evaluation

This study aimed at building and comparing statistical models for predicting sunflower oil concentration in contrasting French conditions. While the statistical approaches we proposed are common in literature (Landau et *al*., 2000), we took particular care of avoiding statistical modeling pitfalls, especially when working with linear models (Dormann et *al*., 2013).

Stepwise methods for variables selection in linear modeling are widely used in science, but highly criticized for their instability, uncertainty and biased parameters (Whittingham et *al*., 2006). Prost et *al*. (2008) suggested using instead Bayesian model averaging (BMA) for selecting variables. The latter authors also evidenced that stepwise selection by BIC led to a reliable selection of predictors when the ratio between number of situations to the number of putative predictors was high, which is the case here (374 situations for 25 variables). Plus, we checked the most probable variables to be included in the linear model by the use of BMA (data not shown), from which we confirmed that the 9 predictors selected by BIC had the highest probabilities of being selected in BMA procedures as well.

We dealt with multi-collinearity by computing stepwise VIF-based indices till a threshold value of 2. The remaining predictors were considered as non-redundant; we assumed that deleted ones did not contribute essentially to oil concentration. However, it is worth noting that other VIF thresholds have been established in the literature: 5 as a common rule of thumb, or even 10 (Kutner et *al*., 2004). Though lower than usual approaches, chosen VIF threshold seemed adapted to the highly correlated predictors that we used in this study.

Cross-validation method was found to be a reasonable way of evaluating our models given the relatively low number of units of simulation; this technique is recommended when dataset is small in order to avoid model over-fitting (Utz et *al*., 2000; Hawkins et *al*., 2003). We used comparable method to compute models performance indicators, though the value of K differed between linear model and the non-parametric and non-linear ones (K= number of USM, K= 10 respectively). However, a 10-fold cross-validation is considered to be the minimum reliable number of sampling for minimizing bias and variance (Fushiki, 2011). Regression tree gave very good fits but very bad predictions, similarly to the study of Borra and Di Ciaccio (2010) where regression tree model over-fitted data. Rao et *al*. (2008) stated that the probability of under-estimating model error of prediction increased with increasing complexity of functions and decreasing number of situations; RMSEP is then probably under-estimated in the case of GAM-based model. Repeated bootstrapping methods should be used in order to obtain reliable predictive error (Efron and Tibshirani, 1997; Jiang and Simon, 2007). However, we could still compare models performances relative to each other.

Models performances

The best model was the one obtained from GAM curves and further formulized into parametric equations. This is not surprising since GAM fits closer to real data so that we can deduce simple to complex relationships depending on the structure of the data (Shatar and McBratney, 1999; Wullschleger et *al*., 2010). We decided to parameterize our GAM in order to obtain quantifiable indicators and compare it to other models. Its performance was equal to that of the non-parametric version (R^2=0.71; RMSEP=1.95). We did not perform any model selection with GAM since we did not have *a priori* known forms of non-linear relationships (Marra and Wood, 2011).

Regression tree fitted well to data (R^2=0.70) and performed as well as the GAM-based one, but predicted badly (RMSEP=2.54). Multiple linear regression with 9 predictors displayed intermediate performance (RMSEP=2.41, EF=0.53). Compared to existing RMSEP value in literature (1.4 oil points error for Pereyra-Irujo and Aguirrezábal (2007)), we obtained higher prediction error values. This could be explained by the wider range of cropping situations and varieties that were used to calibrate and evaluate models; also, method of validation differed (made on an independent dataset in the case of Pereyra-Irujo and Aguirrezábal (2007), cross-validated in our study). This makes the use of RMSEP as the only method of comparing model performances questionable; we however proposed complementary indicators to evaluate our models.

Despite our willing to adapt Pereyra-Irujo and Aguirrezábal (2007) model to our dataset, the model poorly performed in our situations (RMSEP=3.33%; EF=0.09). Re-parameterization was justified by the fact that initial model value of potential oil concentration was set to 50%, whereas our dataset displayed a wider and higher range of OC (47.7 to 60.8%) as well as OCobs (~40.7 to 59.4%). The parameter "a "(intercept) differed by 12 oil points with the non-adjusted model, and "b" (slope) was lower in the new model (0.5 and 0.17% oil accumulation rate per MJ per m² respectively). Our oil concentrations were less responsive to SIRPFW/density ratios than in the Pereyra-Irujo and Aguirrezábal (2007) model. Uncertainties about SIRPFW, as part of simulated predictors, are discussed in next section. For most USM (359 out of 374), simulated oil concentration did not reach their corresponding potential value. Knowing that this potential is defined by genotype in our adjustment, it is not surprising that the new adjusted model was not able to reproduce different varietal behavior. In most cases, oil concentration was governed by factors other than potential, *i.e* environmental factors. Only slight variations of oil concentrations were obtained in response to intercepted radiation and density effects, but Model.1 did not take into account nitrogen or water stress factors. We could add, though, that the concept of "potential" differed slightly in Model.1 compared to other models. For all cases, potential is a maximum value to be reached. In the models 2 to 4, genotypic potential was included in the calculation of a "real" oil concentration just as other factors, whereas for Model.1, genotypic potential played only when it was equal to the maximum, otherwise oil concentration was determined by intercepted radiation and density. For adding "power" to genotypic determinism, varietal diversity should be included in the linear equation part.

Anyhow, this means that oil concentrations variability in our dataset could not be explained only by sum of intercepted radiation, density and potential oil concentration, so that other factors should be included in the model. Plus, the initial model was constructed such that radiation had higher importance than potential oil concentration, which contrasted greatly to other models obtained from this study.

Most of our predictors values were simulated by SUNFLO, therefore tainted with uncertainty though we took particular care of selecting USM that were acceptably predicted for their grain yields. Excluded USM (~10% of total database) displayed a mean prediction error of 40% (RMSEP=13.8 quintals per ha). Prediction error higher than 5 quintals per ha was partly linked to soil characteristics for some trials (Middle-West region of France) where soil stoniness and shallowness limited input data accuracy and reliability. SUNFLO could not correctly simulate some

extreme situations (very intense water stress, and/or over N fertilization); these might deserve a deeper physiological analysis of water x nitrogen interactions, producing specific effects probably not well reproduced by SUNFLO yet. These limitations have already been mentioned in Debaeke et *al*. (2010) and deserve further attention. When compared to these authors yield evaluation, we obtained similar or better mean performances, suggesting that threshold of 10 quintals per ha was comfortably acceptable. For the 374 USM remaining, mean prediction error of a given variety in a given environment was 3.88 quintals per ha (against 5 for cited authors), while mean prediction error of a given variety over all its environments equaled 4.46 quintals per ha (against 3.5 for cited authors). We could neither detect any genotype nor climatic year effects that could be linked to poorer performances of the model.

Predictors' hierarchy and contribution to oil concentration

In the models we built, potential oil concentration was considered to be the main determinant of final oil concentration (from 25 to 56% depending on the model used). This is in line with Borredon et *al*. (2011) and Andrianasolo et *al*. (2012) conclusions, who observed that genotype effect on sunflower oil concentration led to three times more oil variability than other factors (nitrogen, density). Regression tree suggested differences of functioning depending on a given threshold of potential oil concentration (54.4%); older varieties would depend on less factors than newer ones (particularly less environmental factors), thus confirming the higher sensitivity of kernel oil concentration towards environment (Aguirrezábal et *al*., 2009). The genetic determinism of potential oil content is complex; though QTLs for this trait have been identified in sunflower, the phenotypic variance explained by these QTLs remains relatively low (Ebrahimi et *al*., 2008). Leaf area duration and mean radiation use efficiency during post-flowering period were found to have similar contributions to final oil concentration and rank second after potential oil concentration (~12%), corroborating their places as the main source of photosynthetic carbon after flowering (Merrien, 1992). Differences between hierarchies given by our models were inner linked to each model own method of variance partitioning. It is reinsuring though, to obtain similar hierarchies for the top-determinant factors –OC, MRUE2/LAD2, SGR1/density; contributions of temperature and water stress deserve to be further investigated. Nitrogen was found to be as important as radiation until flowering period (regression tree); this goes in line with the observation that leaf nitrogen profile is determined by light profile in the canopy, at least until flowering and under non-limiting water conditions (Archontoulis et *al*., 2011). Neither radiation (SGR) nor intercepted radiations (SIR) were retained in the models 2 and 3. SIR and SGR were in fact dropped from potential

predictors in the BIC stepwise procedure. We believe that radiation effects, especially intercepted radiation ones on oil concentration, have been mitigated by the higher contributions of genotypic and stress factors effects to oil concentration variability. LAD2 (green leaf area duration after flowering) and MRUE2 (mean radiation use efficiency after flowering) behaved as better indicators of canopy functioning diversity than sum of radiation/intercepted radiation in this study, maybe because of the narrower range of variation of SGRPFW and SIRPFW indicators (coefficients of variations: SGRPFW:10%; SIRPFW:13%; LAD2:25%; MRUE2:21%, respectively). SGR indicators were though retained in Model.4 and contributed up to 10% of oil concentration variations. This can be explained by the fact that tree model helped to unravel meaningful interactions and identify important variables for contrasting situations, typically limiting/non limiting conditions. Radiation effects might be mitigated a bit less in situations where nitrogen and water were not limiting. Anyhow, this reinforces the necessity to include radiation/intercepted radiation effects in oil concentration mechanistic modeling processes. Density had a similar contribution to oil concentration in GAM-based model (8%); we could suppose that density took into account part of radiation effects though not explicitly expressed in the model. Model.2 and 4 also highlighted the importance of water availability before (5 to 7%) and after flowering (2 and 7%), but NT34 contribution was as high in Model.3 (7%).

Toward a better understanding of sunflower oil concentration elaboration

On a physiological point of view, genotypic effect could play whether through hull content for older varieties (López-Pereira et *al*., 2000) or through oil concentration in kernel for more recent ones (Izquierdo et *al*., 2008; Aguirrezábal et *al*., 2009). Mantese et *al*. (2006) demonstrated that contrasting oil-potential cultivars differed in initial pericarp and embryo weights and dynamics, as well as oil deposition duration.

Canopy functioning indicators ranked second: as for other yield components (grain number and grain weight), sunflower oil concentration elaboration was largely source-dependent (Andrade and Ferreiro, 1996). Source could be modulated by genotype, as illustrated for stay-green varieties able to maintain longer functioning leaves (de la Vega et *al*., 2011) or by varieties more efficient to remobilize pre-flowering assimilates after flowering (Sadras et *al*., 1993).

Considering intercepted radiation/density effects, lower radiation/higher plant density effects could result in lower pericarp weights as observed in Lindström et *al*. (2006), but could also play at source level through the relationship between radiation use efficiency and SLN (specific leaf nitrogen) for

maintaining photosynthesis capability (Steer et *al*., 1984; Massignam et al., 2009). All things being equal, higher nitrogen doses favor higher duration of green leaf area since the onset of senescence is linked to the achievement of a minimal value of SLN in leaves (de la Vega et *al.*, 2011); at sink level, nitrogen would enhance protein and other seed components accumulation relative to oil, leading to what Connor and Sadras (1992) call "dilution" of oil concentration.

High temperature and water stress effects deserver further investigation, especially since they could be confounded; drying could be triggered by temperature and/or water deficit, which would lead to shorter grain filling duration at sink level (Chimenti et *al*., 2001) and/or sooner leaf senescence at source level (Aguirrezábal et *al*., 2009). Higher hull weights were measured in water-stress conditions (Denis and Vear, 1994); some authors demonstrated that remobilization of pre-flowering assimilates was triggered when water was limiting (Blanchet et *al*., 1988; Hall et *al*., 1990). Others evidenced specific genotype behavior of source regulation in response to water stress (Maury et *al.*, 2000; Casadebaig *et al*., 2008).

A step further in oil physiology understanding would be the calculation of source-sink indicators (Ruiz and Maddonni, 2006; Izquierdo et *al*., 2008) that could help to decorrelate effects of factors (genotype and environment) specifically impacting sink, source, or both.

Models error diagnosis

Diagnosis per trial type helped to highlight problems of lack of correlation (LCS) –*i.e*. faithfulness to patterns – of simulated oil concentration for all statistical models; problems of differences in magnitude (SDSD) were found in Pereyra-Irujo and Aguirrezábal (2007) adjusted-model only, but this could be explained by the fact that it could not reproduce nitrogen and water stress effects. For comparison, model error for oil concentration prediction also originated mainly from lack of correlation component (82%) in Pereyra-Irujo and Aguirrezábal (2007) paper.

We observed an average decreasing pattern of oil concentration in response to growing nitrogen fertilization amounts. Merrien (1992) stated that such depressive effect of nitrogen highly depended on water availability and water x nitrogen interaction. However, models displayed differential patterns, and none of the models could closely describe negative effect of growing nitrogen amounts. Density effect also highly depended on the model considered; each one of them revealed different thresholds at which density effect was positive or negative. For Pereyra-Irujo and Aguirrezábal (2007) adjusted-model, effect of density was observed to systematically be negative on oil concentration, though it was sometimes stagnating (from 3 to 4 plants per m²), positive

(between 4 and 6 plants per m²) or negative (highest density) in observed oil concentration values. There were actually contrasted patterns depending on the variety x site interaction (data not shown); Vellox variety displayed decreasing OCobs values in En Crambade, while they stagnated in Montmaur. The OC values of LG5450_HO variety also stagnated in Montmaur, while increasing in En Crambade.

Assuming that each model establishes mean threshold effect of a given factor, it is not surprising that they displayed differential OCsim patterns and could not take into account individual specific pattern. This is thus the limitation of statistical models: generic relationships (or patterns) are computed, and specific genotype behavior that deviates from this generic relationship could not be correctly predicted (Shatar and McBratney, 1999; Ferraro et *al*., 2009). The use of more process-based models could help to unravel such specific genotype x environment x management interactions, and greatly reduce lack of correlation model error component. Before moving to more complex process-based models, correct hypotheses about oil concentration elaboration should be validated by field experiments, otherwise only the choice of process-based indicators in statistical models should be preferred (Landau et *al*., 2000).

With our best minimum adequate model, we were able to explain up to 70% of sunflower oil concentration variability. The general performances of our models can be considered as satisfactory when compared to other existing statistical models involving a wide range of varieties/cropping conditions (R^2: 45 to 61% for GAM in Tulbure et *al*., 2012; 51 to 56% for regression tree in Ferraro et *al*., 2009; 36 to 43% for multiple linear regression in Khamis et *al*., 2006). The remaining unknown 30% might be linked to several causes. Though we considered simulated predictors as reliable, we could not ignore SUNFLO model uncertainties; if predictors were measured/measurable, this could have generated a much wider range of variability for some predictors, which in turn could potentially increase their contribution to oil concentration variability while reducing final prediction error (RMSEP). Also, the explicit inclusion of interacting terms might improve R^2, although regression tree highlighted simple to complex interactions but performed equally to the best minimum adequate model (GAM-based). It is not excluded that some predictors we dropped by stepwise VIF procedure could have added some explanatory power, suggesting that a less "severe" threshold could have been chosen for dealing with multi-collinearity.

We have established a comprehensive list of putative predictors by the help of our conceptual framework, but we may have missed other possible variables that could have been relevant if expressed in a different way. For instance, identifying periods of thermal time-based sensitivity to

stress factors (water stress, high/low amounts of nitrogen), as done for intercepted radiation (Aguirrezábal et *al*., 2003), could lighten the weight of complex interactions and establish a strong common physiological basis for oil concentration response to water or nitrogen factors, regardless of genotype or other environmental conditions.

In this study, we aimed at building the most parsimonious minimum adequate model and particularly focused on the trade-off between low number of variables, predictive and explanatory power. Model.1 was not enough explanatory nor predictive, though it was totally the contrary in Argentine experiments (Aguirrezabal et *al*., 2003; Pereyra-Irujo and Aguirrezabal, 2007). Indeed, this model was initially calibrated on mainly one variety and in non-limited conditions. Our attempt to re-parameterize the model did not give satisfactory results; there is rather a need to include more than 3 variables for describing oil concentration variability in response to contrasting environmental and management effects.

Model.2 (MLR) better fitted to our data; we gained in satisfactory predictive power with 6 more variables, and those that were selected have a legitimate physiologically-sound basis, assuming a linear relationship between each predictor and oil concentration which might be a too simple way to model reality. Though, most important contributors have been identified and confirmed in more complex equations (Model.3). Then, Model.2 could be used by agronomists if this is about identifying determining factors and bringing more information about sunflower grain oil physiology.

Model.3 (GAM) added more predictive power to Model.2 with the same number of variables, but the transformed relationships deserve to be assessed on other datasets. The aim would be to dissociate relationships artificially generated by the structure of our dataset and “real” relationships having sound physiological explanation. Anyhow, Model.3 could be used by both physiologists and crop modellers, for understanding and predicting sunflower oil concentration.

Finally, Model.4 (regression tree) retained more or less the same number of variables as Model.2 and 3; most contributing variables were identified but types of relationships remained unknown. This model could be more useful to an agronomist or a crop modeller, who wants to be routed for identifying main trends or possibly for decision support tool.

Decomposition of processes by source and sink and effects of determining factor on respective components appear to be essential for better understanding final oil concentration elaboration

regarding genotype x environment x management interactions and leading towards a more mechanistic model.

5. Conclusions

This study aimed at building and comparing statistical models for sunflower oil concentration prediction. The GAM-based model performed best whereas the Pereyra-Irujo and Aguirrezábal (2007) adjusted one was not adapted to our data. Though displaying differential patterns in response to agronomic practices, the models helped to establish an hierarchy among determining factors of observed oil concentration; varietal potential oil concentration ranked first, and depending on oil percent amounts, interacted differently with environmental (radiation, nitrogen, water, temperature) and management practices (density) factors. This helped us to better understand source and sink relationships and order of priority for oil elaboration, which could be of valuable interest when moving to more mechanistic models.

Bilan du chapitre I

- La comparaison des 3 nouveaux modèles statistiques (GAM, linéaire multiple, et arbre hiérarchique (RT)) avec le modèle de Pereyra-Irujo and Aguirrezábal (2007) sur la base d'un même jeu de données montre que le meilleur modèle (plus faible erreur de prédiction) est le modèle GAM (9 variables, RMSEP = 1.9), tandis que le moins bon est le modèle de Pereyra-Irujo and Aguirrezábal (2007) (3 variables ; RMSEP= 3.3). Ce dernier prend en compte les effets liés au génotype, à la densité et au rayonnement mais pas ceux liés à la disponibilité hydrique et azotée.
- Les facteurs qui expliquent le plus la variabilité de la teneur en huile finale sont le génotype (50%), puis, à des pourcentages variables selon les modèles, le fonctionnement photosynthétique en post-floraison, les facteurs azote, eau, densité et température.
- Comparée au module « huile » du modèle SUNFLO (Casadebaig et al., 2011), l'erreur moyenne de prédiction est divisée par deux dans nos modèles (RMSEP ~4 points d'huile dans le module initial ; RMSEP~2 points d'huile dans les nouveaux modèles statistiques).

La modélisation statistique a permis de faire le tri des variables importantes dans la détermination de la teneur en huile, mais elle ne peut expliciter le mode d'action des différents facteurs ni expliquer leurs interactions.

Sachant par ailleurs que les facteurs agissent soient sur les « sources », soit sur les « puits », nous proposons dans les deux chapitres suivants d'expliciter ces modes d'action via des suivis d'expérimentations au champ et en serre sur les « sources » et/ou sur les « puits » en réponse aux facteurs qui contribuent le plus à expliquer la teneur en huile.

Chapitre II: Analyzing oil and proteins accumulation through source and sink framework establishment in sunflower achenes

Article en préparation pour Field Crops Research.

Analyzing oil and proteins accumulation through source and sink framework establishment in sunflower achenes

Fety Nambinina Andrianasolo [a,b,c], Pierre Maury [b,c,1], Philippe Debaeke [b,c,1]

[a] CETIOM, Centre INRA de Toulouse, CS 52627, 31326 Castanet-Tolosan Cedex, France

[b] INRA, UMR AGIR, CS 52627, 31326 Castanet-Tolosan Cedex, France

[c] Université de Toulouse, INP, ENSAT, CS 52627, 31326 Castanet-Tolosan Cedex, France

[1] Co-advisors of the PhD thesis

Authors e-mail addresses:

fandrian@toulouse.inra.fr

maury@ensat.fr

debaeke@toulouse.inra.fr

Corresponding author: debaeke@toulouse.inra.fr

Abstract

Background and Aims. Most recent mechanistic crop models are based on source-sink relationships. To create a conceptual framework for a model of achene quality traits in sunflower (*Helianthus annuus* L.), source-sink relationships were investigated in contrasting environments.

Methods. Two experiments were carried out under field conditions in 2011 and 2012 at Auzeville research station. Oil, proteins, hulls (sinks), and leaves, receptacles and stems (sources) were measured weekly under various contrasting nitrogen (N-: no fertilization; N+: 150 kg N per ha), plant density (3 and 4.5 plants per m²) and genotype (cv. Kerbel and cv. Olledy) treatments. In particular, photosynthetic activity and nitrogen content in source organs were measured in 2012 in *cv*. Kerbel. Bi-linear models were fitted to source and sink organs weights (DW) per m² dynamics.

Key results. Nitrogen and density effects mainly acted through rates of processes (all sink and source components for N; hulls, green leaves and receptacles for density) while genotype effects occurred through differences in initial values, timings and durations of processes (oil, leaves, receptacles and stems). Increase of dry weight and N uptake after flowering contributed 100% and 72% respectively to grain DW and N content in the N+ treatment; though, N remobilization was greater in N-. A significant correlation ($R^2 = 0.87$) was observed between oil weight during grain filling and photosynthetic activity duration in both N treatments. A bi-linear positive relationship was found between achene oil and protein and with photosynthesis activity duration for all situations.

Conclusions. Modulations of sink dynamics (oil, proteins) by source (leaf area duration, N remobilization per grain) were found in nitrogen-deficient situations. Such results could be used as a starting framework for a source-sink-based mechanistic model in sunflower.

Keywords: *Helianthus annuus* L.; leaf area duration; nitrogen uptake; photosynthetic activity; remobilization; source-sink; senescence; sunflower oil

1. Introduction

Grain oil concentration is a major economic criterion for oilseeds production. In sunflower (*Helianthus annuus* L.) particularly, French farmers are rewarded or penalized up to 1.5% per oil point relative to oil concentrations higher or lower than 44% (commercial standards, CETIOM). However, last studies demonstrated that grain oil concentration could vary greatly depending on genotype and environment (Borredon et al., 2011); Champolivier et al. (2011) observed that for two oil-potential contrasting genotypes, variability related to cropping conditions was much higher than that related to genotypic difference. For maintaining high oil concentrations, one should control its determining factors by understanding how they act and/or interact.

Genotype, radiation, nitrogen, plant density, water stress and high temperatures are often reported as determining factors of oil concentration in the literature (Connor et Hall, 1997). While water stress, temperature, density and genotype effects have been extensively studied (Rizzardi et al., 1992; Aguirrezábal et al., 2003; Rondanini et al., 2003; Mantese et al., 2006; Anastasi et al., 2010), few references exist for nitrogen effect on oil concentration, especially in interaction with other factors. Some authors reported that nitrogen effect is negative on oil concentration and reversely positive on proteins concentration. This apparent antagonism is rather linked to a "dilution effect" since oil and proteins neither accumulate at the same time, nor at similar rates and durations (Connor and Sadras, 1992). Moreover, Merrien (1992) stated that depending on cropping conditions, there existed variable degrees of antagonisms between oil and protein concentrations. This suggests that a given factor does not always have proportional effects on oil and proteins. For better understanding the effects of agronomic factors, Connor and Hall (1997) proposed to analyze temporal differences in the patterns of oil and proteins accumulation.

Besides, other studies demonstrated that those determining factors could also have effects on organs that provide assimilates for elaborating oil and proteins – *i.e.* stems, leaves and receptacles during post-flowering period. Changes in assimilates partitioning were reported in response to nitrogen (Massignam et al., 2009), plant density (Villalobos et al., 1994) or water stress (Blanchet et al., 1988), for instance. Under later situations, Hall et al. (1990) demonstrated that contribution of carbon remobilization from vegetative organs increased in comparison with non-limiting water stress conditions. Evaluation of the relative contributions of remobilization vs assimilation to grain filling was carried out in many other species such as maize (Pommel et al., 2006), barley (Dordas, 2012), sorghum (van Oosterom et al., 2010), wheat and soybean (Borras et al., 2004), but their relationship with oil or proteins accumulation was not investigated.

Modulations by agronomic factors of rates and durations of remobilization processes across time at source organs level could thus explain observed "antagonism" between oil and protein concentrations. Such modulations could also affect the order in which assimilates are remobilized, though priority order of assimilates elaboration at sink level appears to be stable in sunflower: hulls first, then kernel components: proteins and finally oil (Connor and Hall, 1997; Mantese et al., 2006, Izquierdo et al., 2008). This order could be partly linked to chronological assimilates demand during grain filling- N first, C after (Goffner et al., 1988; Triki et al., 1997). Neither the order in which source organs are mobilized to provide assimilates to grain components, nor the order of N and C remobilization from sources were extensively studied in sunflower. Understanding those priority orders could help to draw a conceptual framework of source-sink relationships in sunflower, as established in other species such as soybean (Sinclair et al., 2003) or sorghum (van Oosterom et al., 2010).

In this study, we propose to evaluate effects of most important agronomic factors –nitrogen, plant density and genotype- on dynamics (rates and durations) observed at sink and source organs levels, as well as on contributions of remobilization and assimilation to final quantities of oil and proteins at harvest.

Assuming that nitrogen effect was the least studied factor in literature, we proposed to first evaluate nitrogen main effect in one given genotype and one plant density for two experimental years. Then a particular focus will be done in one experimental year where collected data were more complete to evaluate dynamics of source and sink organs. After nitrogen, plant density effect was analyzed in contrasted nitrogen conditions (for one genotype) and genotypic differences were finally evaluated in contrasting plant density and nitrogen conditions. Indicators of source and sink compartments were computed at harvest to assess the relationships between sources and sinks.

2. Materials and methods

2.1. Site and experimental design

2.1.1 Site characteristics

Two field experiments were conducted at INRA station in Auzeville, France (43°31'41.8" N, 1°29'58.6" E) in 2011 and 2012. Sunflower was grown on a deep loamy soil (100 cm depth) with a potential available water reserve of 180 mm, with little or no stoniness. The crop was preceded by maize in 2010 and sorghum in 2011; residual N before sowing amounted 48 and 33 kg N per ha in 2011 and 2012 respectively. Two commercial hybrids were used: *cv*. Kerbel (for both years) and *cv*. Olledy (2012). Cultivars contrasted for their achene oil concentration (*cv*.Kerbel: high-oil; *cv*.Olledy: low-oil) and hull/kernel ratio (higher for *cv*.Kerbel). Daily weather data -*i.e.* global radiation, rainfall, minimum and maximum temperature and potential evapotranspiration were collected locally at 2 m height.

2.1.2. Experimental design and crop management

The crops were sown on 08 and 06 April and harvested on 30 August and 11 September in 2011 and 2012 respectively. For both years, the experiment consisted in a split-plot design with nitrogen factor as main plot level (N+: non-limiting; N-: deficiency) and genotype and plant density (D1: 3 plants per m²; D2: 4.5 plants per m²) as subplots; these were randomly distributed with 3 replicates. The size of each unit plot was 30 m² (6 rows) and row width was 50 cm.

In N+ modality, 150 kg N per ha were split and brought into two doses during vegetative growth; 75 kg N per ha on 19 May and the remaining half on 06 June in 2011; dates of fertilization in N+ were 15 May and 01 June in 2012.

Plots were over-sown at 9 plants per m² then thinned to desired densities at 8-leaves stage. In order to avoid severe water stress, the crop was irrigated on 13 April (28 mm) and 26 May (43 mm) in 2011 and 29 June (25 mm) and 17 July (29 mm) in 2012. Weeds and pests were adequately controlled.

2.2. Simulation of water balance with SUNFLO crop model

We proposed to monitor soil water availability during grain filling period by using the SUNFLO crop model (Casadebaig et al., 2011). SUNFLO is a dynamic crop growth model simulating sunflower growth and development submitted to genotypic, environmental and crop management factors. It was evaluated on a wide range of pedo-climatic situations in France (Debaeke et al., 2010; Casadebaig et al., 2011). Soil water availability was described as fraction of transpirable soil water (FTSW, Casadebaig et al., 2008); it corresponds to the ratio between available and maximal transpirable soil water. We assumed that FTSW values lower than 0.6 indicated water stress. Only mean FTSW dynamics per year between start of flowering and physiological maturity were illustrated.

2.3. Dynamic monitoring and model adjustment

2.3.1. Phenology

Sunflower phenological stages were recorded weekly at the whole plot scale from bud visible stage till late after physiological maturity (which was noted as R9, when capitulum bracts turned from yellow to brown, Schneiter and Miller, 1981). Flowering was scored when 95% of plants in a plot displayed stamens in all florets from the outer ring of capitulum (R5.1 stage, Schneiter and Miller, 1981). Half of total number of plots were dedicated to non-destructive measurements throughout the grain filling period (source); destructive measurements were operated in the other half (sink).

2.3.2. Sink dynamics

Starting from flowering, 40 random plants in the four middle rows of each plot were tagged for weekly destructive measurements. At each harvest date, three heads per plot were randomly cut; receptacles and grains were separated and dried at 80°C for 48 hours. Capitula and grain parts were weighted. Empty grains were excluded, and kernel and hull parts were separated for 2012 experiment. There were 8 dates of harvest from flowering to maturity for both years. Grain weight and grain number per m² were determined on a 1000 achenes sample. Only filled grains were taken into account. Grains and kernels were then analysed for their oil and N concentrations by NMR (Minispec NMS 110, Bruker, Analytische Mestechnik, Rhinsteten, Germany) and Dumas method (Flash 2000, Thermo Scientific, inc., US) respectively. N concentrations were multiplied by 6.25 to

obtain protein concentrations. Since hulls weights dynamics were not available for both genotypes (only at harvest for *cv*. Olledy), hulls weights were computed as the differences between achene and oil and protein weights for both genotypes and years. Observed and computed hulls weights were highly correlated for *cv*. Kerbel ($r = 0.97$; $p < 2.2$ e-16).

2.3.3. Source dynamics

2.3.3.1. Green leaves

Five plants in the two middle rows of each plot were tagged for monitoring leaf area dynamics after flowering. Plant leaf area (PLA) at flowering was measured from length and width of each green leaf multiplied by 0.74 (Casadebaig et al., 2008). Leaf area index (LAI, m^2. m^{-2}) at flowering was then obtained by multiplying PLA by plant density which was extensively determined on central rows. LAI was computed weekly by discounting areas of senesced leaves –*i.e*. leaves having 50% or more yellow or desiccated area. Leaves weight per m² was determined by dividing LAI by specific leaf area (SLA, $m^2.g^{-1}$) that was measured in some *cv*. Kerbel modalities (see section 2.3.3.4). When SLA was missing, mean value of SLA computed over years and N modalities was used. Leaf area duration (LAD) corresponded to the integral of all LAI values across thermal time after flowering (base temperature of 6°C; CETIOM). In 2012, total nitrogen percentages were measured in leaves from cv.Kerbel x D2 x (N+,N-) modalities, following Dumas method (EURO EA-3000, Eurovector Spa, Milan).

2.3.3.2. Receptacles

Receptacles dry weights per m² were available from flowering to maturity at each sampling date. In 2012, total nitrogen percentages were measured in receptacles from *cv*. Kerbel x D2 x (N+ and N-) modalities, following Dumas method (EURO EA-3000, Eurovector Spa, Milan).

2.3.3.3. Stems

Stems dry weights per m² were available only in 2012. At flowering, 3 plants were harvested, separated into stems, receptacles and achenes (López-Pereira et al., 2008) and oven-dried. Stems parts were again collected at R6 stage and physiological maturity (R9 stage). Total nitrogen percentages were measured in stems from *cv*. Kerbel x D2 x (N+ and N-) modalities, following Dumas method (EURO EA-3000, Eurovector Spa, Milan).

2.3.3.4. Photosynthesis activity and SLA measurements

In 2012, 2 plants per plot were tagged at flowering for photosynthesis monitoring in *cv*. Kerbel x D2 x (N+ and N-) modalities. For 3 occasions during grain filling (293, 577 and 672 °C days after flowering), photosynthetic activity was measured with a portable gas exchange system (LI-COR 6400, Lincoln, Nebraska, USA) at different leaves nodes: 1, 5, 10, 15, 20 and 25 from capitulum. Measurements of net photosynthetic rate were made at light saturation of 1000 µmol m^{-2} s^{-1} photosynthetic photon flux density and ambient CO_2 concentration (approximately 400 µmol mol^{-1}). Meanwhile, proportion of PAR intercepted was computed from incident PAR (0.48 x global radiation) and measurements of radiation at each leaf level with a line quantum sensor (AccuPAR, Decagon, Washington, USA) following the method of Dosio et al. (2000). Photosynthetic activities were then corrected by considering individual leaf area and intercepted PAR at each leaf level (Maury, 1997; Hammadeh et al., 2005). A photosynthetic index was computed as the sum of each leaf level photosynthetic activity multiplied by plant density. Photosynthetic activity duration (PAD) corresponded to the integral of photosynthetic index across thermal time after flowering. Small discs (2 cm diameter) of leaves were harvested in order to measure SLA.

2.3.3.5. Nitrogen Nutrition Index at flowering

Aboveground biomass at flowering and relative total nitrogen content were used to compute Nitrogen Nutrition Index (NNI, Debaeke et al., 2012) for characterizing plant nitrogen status at flowering.

2.3.4. Model adjustement

In order to obtain a single skeleton for comparing various crop management practices, a bi-linear model was fitted to source and sink dynamics across thermal time after flowering:

$$y = a1 \times x + b1, \text{for } 0 \leq x \leq t1,$$

$$y = a1 \times t1 + b1 + a2 \times (x - t1), \text{for } x \geq t1,$$

where y is whether oil, proteins, hulls, leaves, receptacles or stems weights per m² and x is thermal time after flowering. a1, a2, b1 and t1 are model parameters corresponding to first phase evolution rate (accumulation for sink, stagnation or decrease for source), second phase evolution rate

(stagnation or slow decrease for sink; sharp decrease for source), intercept and thermal time threshold value at the change from first to second phase, respectively.

Fits were performed with the *nls* function of R (R 3.0.2 version).

We also computed ft1 (value of y at t1), tmax (thermal time at which source weights are minimum, corresponding to physiological maturity) and ty0 (thermal time at which y = 0 for sink dynamics).

Duration of each process could be deduced by subtracting t_1 from tmax for source organs and ty0 from t1 for sink organs. We assumed herein that a1 nearly equals zero for source and a2 nearly equals zero for sink. The bi-linear model was also used to establish relationships between sink and source organs. The t1 parameter then corresponds to the value of source at which sink dynamics change. Base temperature was considered to be 6°C.

2.4. Sink and source indicators at harvest

We proposed some indicators to describe "sink" and "source" compartments at harvest. Those were computed in all nitrogen, plant density and genotype conditions.

2.4.1. "Sink" indicators: percentage and quantities

Variables describing "sink" organs were computed at harvest: percentage and quantities (in g m^{-2}) of grain oil, proteins and hulls. It is assumed that indicated oil and proteins quantities correspond to quantities measured in achenes.

2.4.2. "Source" indicators

2.4.2.1. Indicators of assimilation

For evaluating carbon and nitrogen assimilation in post-flowering period, we proposed to compute frequently used indicators in literature: nitrogen uptake (g m-2) and leaf area duration (g m-2 °Cd^{-1}) from flowering to maturity.

2.4.2.2. Indicators of remobilization

Nitrogen and dry matter remobilization contributions were evaluated as the differences between their quantities in vegetative organs at flowering and their quantities at physiological maturity.

2.5. Statistical analysis

The quality of model performance was evaluated by computing R^2 and root mean squared error of prediction (RMSE). Significant effects of agronomic factors were analyzed by using ANOVA; unless otherwise stated, p-value threshold was set to 0.05. When significant, the test of least significant difference (LSD) was performed to identify similar groups. Pearson correlation test was carried out for evaluating the correlations between source and sink indicators. Student t test was performed to compare t1 parameters values between processes and genotypes.All analyses were performed with R software (R Core Team, 2014).

3. Results

3.1. Weather conditions, crop phenology and status

Climatic data and phenological stages are displayed in Table 1. The 2011 growing season was globally dryer than 2012, except in July where rainfall was particularly high in 2011 (87 mm). There was more rainfall from April to June in 2012 (pre-flowering period). Temperatures were higher in 2011 in the pre-flowering period compared to those of 2012, but similar starting from June to August period. During the latter, global radiation was slightly higher in 2012 compared to previous year.

Flowering occurred on 29 June in 2011 and on 27 June and 02 July for *cv*. Kerbel and *cv*. Olledy in 2012, respectively. Grain filling period of *cv*.Kerbel lasted 57 days in 2011. In 2012, grain filling lasted 71 and 75 days in N+ and N- modalities respectively (*cv*.Kerbel). Physiological maturity happened after 63 and 65 days after flowering in N+ and N- modalities in *cv*. Olledy.

Table.1. Monthly meteorological data (global radiation, total rainfall and mean temperatures) in 2011 and 2012 experiments. Main phenological stages (emergence, flowering and physiological maturity) dates were noted. In 2012, physiological maturity dates varied with genotype and nitrogen modalities.

		April	May	June	July	August	September
2011							
climatic data	global radiation ($MJ\ m^{-2}$)	20.5	22	19.8	20.9	20.1	
	rainfall (mm)	26	38.5	41	86.5	21.5	
	mean temperature (T°C)	15.2	18.2	18.7	20.2	22.5	
phenological data	emergence	24/04/2011					
	flowering			29//06/2011			
	physiological maturity					25/08/2011	
2012							
climatic data	global radiation ($MJ\ m^{-2}$)	13.8	21.2	22.4	22.9	20.9	14.7
	rainfall (mm)	69	75.5	53.5	58	48.5	26
	mean temperature (T°C)	11.4	16.7	20.5	21.1	23.5	18.9
phenological data	emergence	21/04/2012					
	flowering			from 27/06 to 02/07/2012			
	physiological maturity						from 03/09 to 10/09/2012

NNI varied from 0.38 to 1.31 in 2011 and from 0.29 to 0.96 in 2012; in particular, N- situation led to higher plant nitrogen deficit in 2012 than in 2011, while nitrogen status was better in N+ x 2011 (0.97) compared to N+ x 2012 (0.70) (data not shown).

According to the simulations of water balance with SUNFLO crop model (Fig.1), we could consider that field experiments in 2011 and 2012 were both conducted under water deficit since FTSW values did not exceed 0.55 throughout the grain filling period. However, soil water availability was higher in 2012, as FTSW values ranged from 0.55 to 0.25 (0.4 to 0.08 for 2011) between Julian days 180^{th} and 200^{th} (from end of June to mid-August).

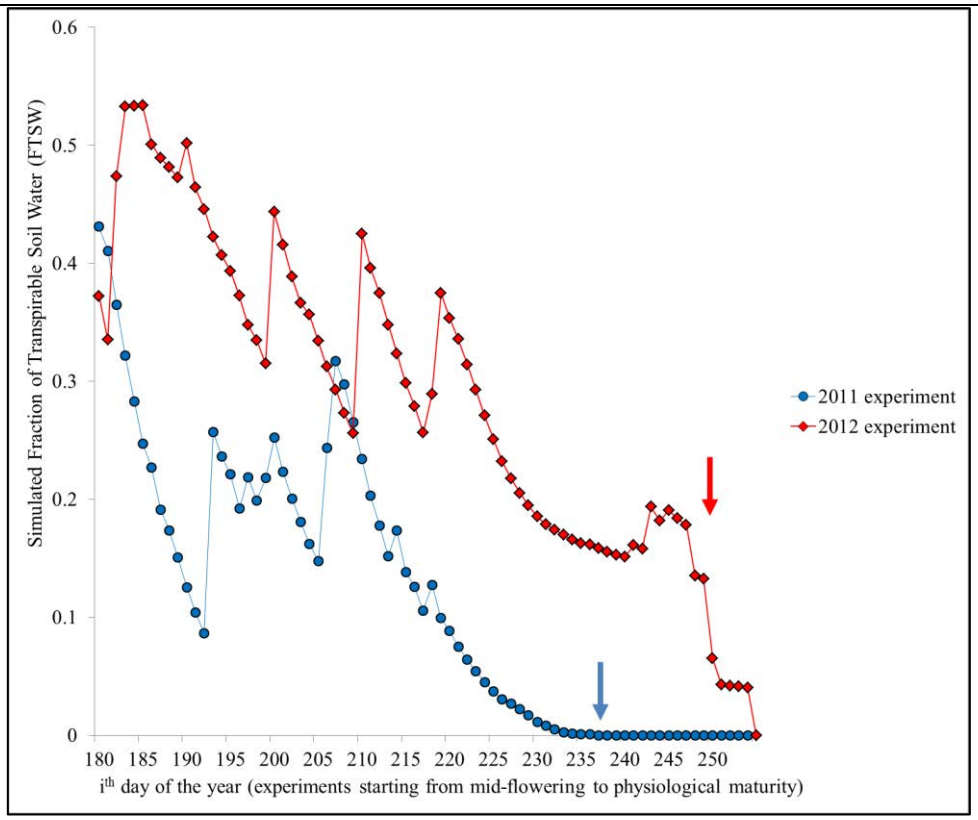


Fig.1. Evolution of fraction of transpirable soil water (FTSW) during grain filling period as simulated by SUNFLO model for 2011 and 2012 field experiments. Arrows indicate physiological maturity averaged per year.

3.2. Year and nitrogen effects on source and sink indicators at harvest

We first proposed a global comparison of performances by computing agronomic indicators at harvest (oil, proteins, hulls concentrations and weights; leaf area duration, nitrogen uptake after flowering, biomass and nitrogen remobilization). Year and nitrogen effects were analyzed and similar groups were established based on LSD test, all things being equal (1 genotype: cv. Kerbel and 1 plant density: D2). Only effects with p-values lower or equal to 5% were considered as significant.

No year by nitrogen (Y x N) interaction effect was found (Table.2). Nitrogen effect was significant on all indicators at harvest, except hulls concentration: situations in N+ displayed higher values than in N-, except for oil concentration where it was reversed.

There was significant year effect on oil concentration and biomass remobilization (with mean values lower in 2011 than in 2012), and proteins concentrations, hulls weight and nitrogen uptake after flowering for which values were significantly higher in 2012 compared to 2011.

We established a ranking of all 4 year by nitrogen combination situations (Y x N) and observed that 2012 x N- displayed highest oil concentrations but not highest oil quantities (which were found in 2012 x N+). Contrarily, 2011 x N+ displayed highest proteins concentration but not highest proteins quantities (2012 x N+). Most productive situation was 2012 x N+ with highest values for most sink and source quantities and durations, while 2012 x N- was the least productive one.

Table.2. Summary table of agronomic performance of *cv*. Kerbel in 2011 and 2012 (plant density in D2) after performing ANOVA and LSD test. Agronomic indicators at harvest were subdivised into sink (achene component weights and percentages) and source (leaf area duration, nitrogen uptake after flowering, biomass and nitrogen remobilization). Otherwise stated, variables correspond to mean values. Year, nitrogen and year by nitrogen effects were tested and level of test of significance is indicated by "***" (p-value < 0.1%); "**" (p-value < 1%); "*" (p-value < 5%). p-values higher than 5% were considered non-significant ("-" symbol). Similar groups established from LSD test share at least one letter.

sink and source categories	variables at harvest	ANOVA table				
		significant effects	2011		2012	
			N+	N-	N+	N-
sink	oil concentration (%)	Y** N*	45.8 c	50.4 b	53.9 ab	56.1 a
	proteins concentration (%)	Y** N**	19.7 a	15.7 b	15.8 b	11.9 c
	hulls concentration (%)	-	24.7 a	26.4 a	26.3 a	27.3 a
	oil weight ($g\ m^{-2}$)	N***	220.9 a	138.6 b	252.7 a	140.6 b
	proteins weight ($g\ m^{-2}$)	N**	68.9	43.8 b	74.3 a	29.8 b
	hulls weight ($g\ m^{-2}$)	Y* N**	186.4 a	93.9 c	142.3 b	80.4 c
source	leaf area duration ($g\ m^{-2}\ °Cd^{-1}$)	N***	1481.6 a	843.8 b	1710.7 a	762.1 b
	nitrogen uptake after flowering ($g\ m^{-2}$)	Y* N**	8.9 a	4.6 b	9.5 a	2.9 c
	biomass remobilization ($g\ m^{-2}$)	Y**	-197.3 b	-85.1 ab	-133.3 a	102.3 a
	nitrogen remobilization ($g\ m^{-2}$)	N*	3.4 a	1.35 a	3.7 a	2.6 a

3.3. Nitrogen, density and genotype effects on source and sink indicators at harvest

Given the wider availability of data in 2012 concerning source and sink dynamics, we decide to further investigate nitrogen, plant density and genotype effects on dynamics by focusing on 2012 experiment. We proposed to first analyze indicators at harvest, then along grain filling period (dynamics). Concerning variables at harvest, we found 3 types of interaction effects on the different source and sink indicators (Table.3); nitrogen by genotype interaction effect (N x G) was significant on oil weight, leaf area duration, nitrogen uptake and biomass remobilization. For the latter, *cv*.Kerbel displayed lower values than *cv*.Olledy in N+ while it was reversed in N- situation. For the first three variables, *cv*.Kerbel had higher values in N+ but lower in N-, compared to *cv*.Olledy. There was also significant nitrogen by density effects (N x D) on oil concentration and weight, proteins, hulls weights and leaf area duration. Globally, density effect was stronger in N+ situation than in N-. Finally, genotype by density effects were detected (G x D) on oil and protein concentration; density effect was stronger in *cv*.Kerbel (positive effect for oil and negative for proteins).

Table.3. Summary table of agronomic performance of *cv*. Kerbel and *cv*.Olledy in 2012 experiment (2 plant densities: D1 and D2 and 2 nitrogen modalities: N+ and N-) after performing ANOVA and LSD test. Agronomic indicators at harvest were subdivised into sink (achene component weights and percentages) and source (leaf area duration, nitrogen uptake after flowering, biomass and nitrogen remobilization). Otherwise stated, variables correspond to mean values. Genotype, nitrogen, density, and their potential interaction effects were tested and level of test of significance is indicated by "***" (p-value < 0.1%); "**" (p-value < 1%); "*" (p-value < 5%). p-values higher than 5% were considered non-significant ("-" symbol). Similar groups established from LSD test share at least one letter.

sink and source categories	variables at harvest	ANOVA table	*cv*.Kerbel				*cv*.Olledy			
			N+		N-		N+		N-	
		significant effects	D1	D2	D1	D2	D1	D2	D1	D2
sink	oil concentration (%)	GxD** NxD*	50.4 c	53.9 b	54.8 ab	56.1 a	51.9 c	51.9 c	55.1 ab	54.6 b
	proteins concentration (%)	N*** GxD*	17.2 a	15.8 a	13.3 b	11.9 b	16.7 a	16.4 a	12.2 b	12.9 b
	hulls concentration (%)	-	26.6 a	26.3 a	27.1 a	27.3 a	24.8 a	24.4 a	25.7 a	27.5 a
	oil weight ($g\ m^{-2}$)	NxD*** NxG*	154.4 cde	252.7 a	132.3 e	140.6 de	173.6 c	213.8 b	149.2 cde	168.1 cd
	proteins weight ($g\ m^{-2}$)	NxD*	52.4 c	74.3 a	32.2 d	29.8 d	56.3 bc	67.7 ab	32.9 d	39.7 d
	hulls weight ($g\ m^{-2}$)	NxD*	99.01 bc	142.3 a	77.4 d	80.4 cd	105.0 b	130.7 a	88.8 bcd	100.1 bc
source	leaf area duration ($g\ m^{-2}\ °Cd^{-1}$)	NxG* NxD*	1122.3 bc	1710.7 a	506.0 e	762 de	1098 cd	1507 ab	892 cd	849 cde
	nitrogen uptake after flowering ($g\ m^{-2}$)	NxG*	5.05 bc	9.5 a	1.8 d	2.9 cd	3.4 bcd	6.2 ab	2.3 d	5.9 b
	biomass remobilization ($g\ m^{-2}$)	NxG**	-81.5 cd	-133.3 cd	154.0 a	102.3 ab	69.9 abc	-88.4 cd	-26.1 bcd	-177.5 d
	nitrogen remobilization ($g\ m^{-2}$)	-	3.4 ab	3.7 ab	3.3 ab	2.6 ab	4.4 a	2.9 ab	2.7 ab	0.9 b

3.4. Analysis of dynamics of source and sink organs

Source and sink variables were oil, proteins and hulls (sink) and green leaves, receptacles, and stems weights per m² (source) respectively. Data were fitted with a bi-linear model and significant effects on model parameters were analyzed. For clarification purposes, we proposed to first illustrate nitrogen effect (all other things being equal: 1 genotype: *cv.* Kerbel and 1 plant density: D2), then plant density (1 genotype: *cv.*Kerbel) and genotype effects (*cv.*Kerbel and *cv*. Olledy).

3.4.1. Effect of crop N status on source and sink dynamics

3.4.1.1.Biomass, protein and oil dynamics after flowering

Nitrogen effect was analyzed on source and sink dry weights dynamics in *cv.*Kerbel (plant density D2). All dynamics were affected by nitrogen factor for one or more parameters (Fig.2): parameters of rates were significantly affected for oil and receptacles weights (a1) and green leaves, receptacles and stems (a2) dynamics. ft1 differed between N situations for proteins, hulls, receptacles and stems weights. Duration parameter was also significantly affected by N treatment for stems weights dynamics. For all situations, values of parameters were higher in N+ than in N- (Table.6), except for duration. Quality of fit was very satisfactory (mean $R^2 = 0.92$).

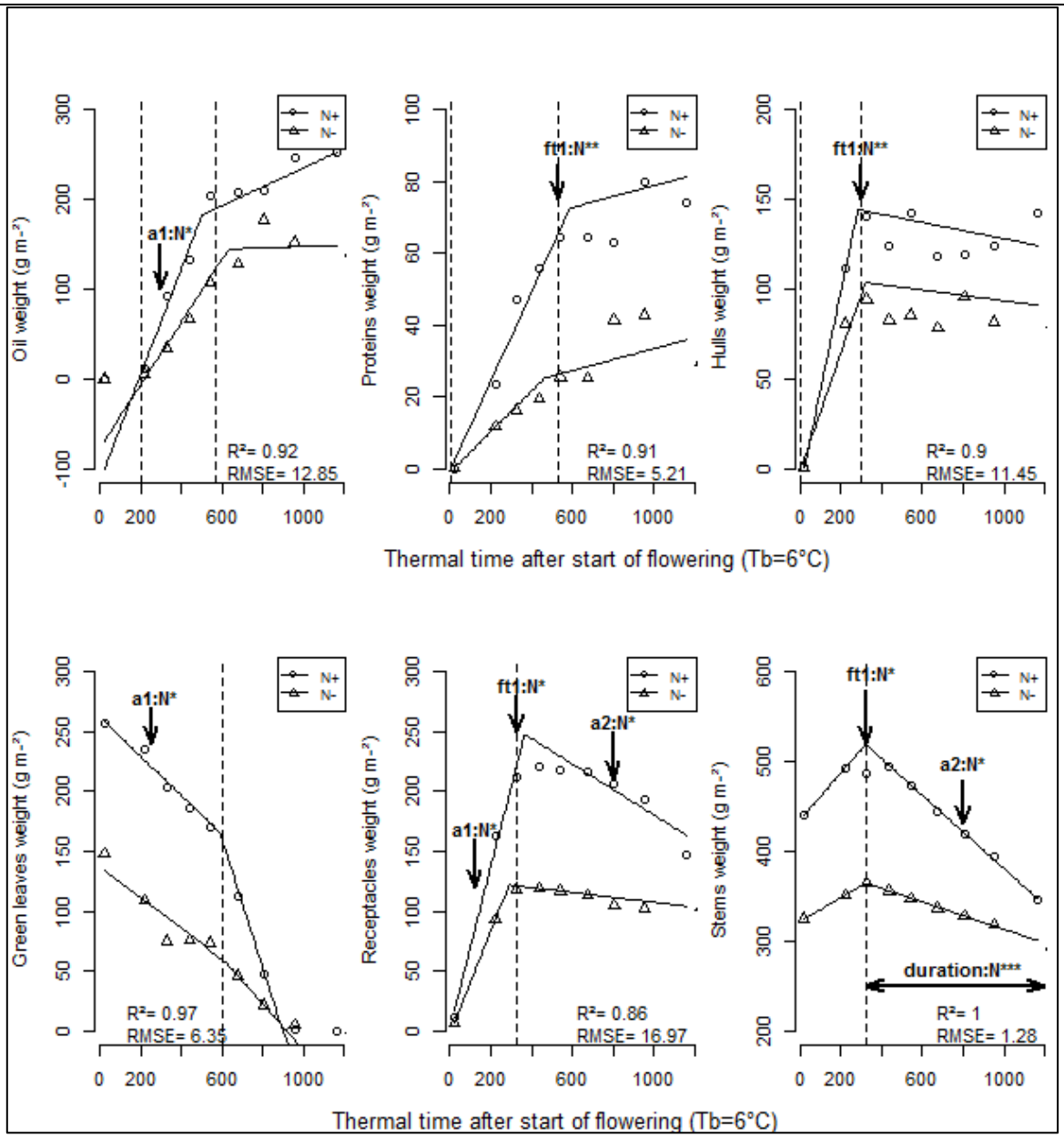


Fig.2. Representation of sinks (oil, proteins and hulls) and sources (green leaves, receptacles and stems) weights dynamics during grain filling in cv.Kerbel (plant density in D2, 2012 experiment). Symbols correspond to nitrogen modalities (N+, N-) and solid lines to result of the fit by a bi-linear model. Only sink and source indicators for which nitrogen effect was significant for at least one parameter are represented. Adjustment for green leaves dynamics was carried out by excluding values close to zero. Model parameters were rates of evolution in the first and second phase of dynamics (a1, a2), intercept (b1) and thermal time after start of flowering at which the second phase starts (t1). t1 corresponds to the start or triggering of source property for source organs. Complementary variables (ty0, ft1, duration) were computed as well as quality of model fit (R^2, RMSE). Dashed lines illustrate ty0 (sink) and t1 (sink and source) parameters. Only significant parameters are indicated by arrows on the graphs with their level of significance: "***" (p-value < 0.1%); "**" (p-value < 1%); "*" (p-value < 5%). p-values higher than 5% were considered non-significant. For parameter values, please refer to Table.6.

3.4.1.2.Plant N dynamics for two contrasted N regimes

Nitrogen effect was also analyzed on nitrogen weights dynamics in source organs (Fig.3): green leaves, receptacles and stems (*cv*.Kerbel, plant density: D2). Nitrogen effect was significant only on

nitrogen dynamics in receptacles (b1 and duration parameters) and stems (b1 parameter). p-values were 0.05, 0.10 and 0.08 respectively (data not shown). Patterns of nitrogen dynamics in green leaves were similar in both N conditions: sharp decrease first and less steep decline in a second phase. For nitrogen in receptacles and stems, patterns differed depending on N status; in receptacles, nitrogen quantities increased in a first phase in both situations then decreased in N+ and slightly increased in N- (Table.4). In stems, nitrogen quantities decreased in a first phase in both situations and still kept on decreasing in N+ while they were stable in N-. Regarding parameters values, it was observed that source organs were depleted faster in N+ than in N-. Initial values (ft1) were higher in N+ though.

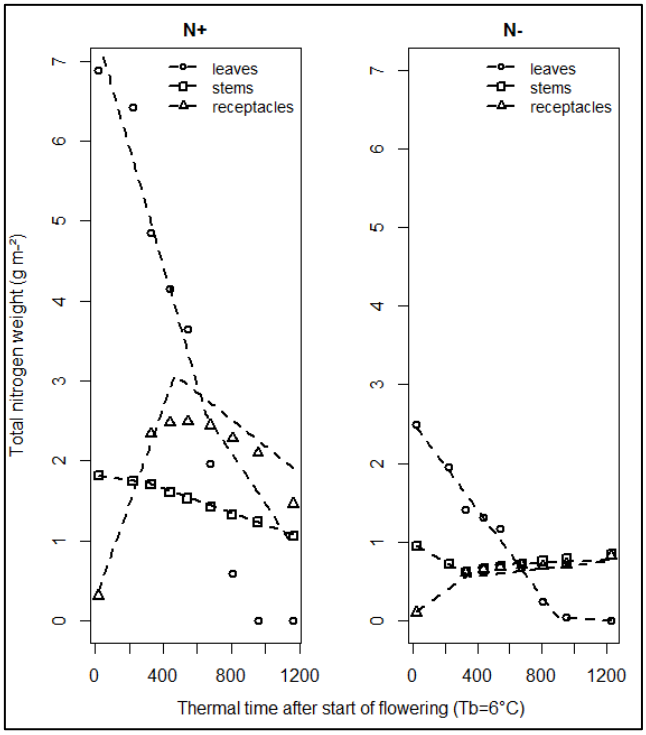


Fig.3. Representation of nitrogen weights dynamics in source organs (green leaves, receptacles and stems) during grain filling in cv.Kerbel (plant density in D2, 2012 experiment). N+ and N- situations were separately illustrated. Dynamics were fitted with a bi-linear model (dashed line). Model parameters were rates of evolution in the first and second phase of dynamics (a1, a2), intercept (b1) and thermal time after start of flowering at which the second phase starts (t1). t1 corresponds to the start or triggering of source property for source organs. Complementary variables (ty0, ft1, duration) were computed as well as quality of model fit (R^2, RMSE). For parameters values, please refer to Table.4.

3.4.1.3. Photosynthetic activity

Relationship between photosynthetic activity duration (PAD) and oil and protein weight accumulation was analyzed in *cv*.Kerbel (plant density in D2). The aim was to compare oil and proteins dependency upon photosynthesis according to nitrogen status. Model parameters applied to those relationships were newly defined (Fig.4). PAD varied from 600 to 4400 µmol CO_2 m^{-2} s^{-1} °C days in N- situations while it ranged from 1500 to 11000 µmol CO_2 m^{-2} s^{-1} °C days in N+.

We found both good relationships between sink (oil, proteins) elaboration and photosynthetic activity duration, though relationship with oil was better (R^2= 0.87 vs 0.55 for proteins). Oil weight per m² increased with increasing PAD till a PAD threshold of 6500 µmol CO_2 m^{-2} s^{-1} °C days. PAD threshold was around 7900 µmol CO_2 m^{-2} s^{-1} °C days for proteins. Pattern for both relationships was similar; first increasing, and then decreasing.

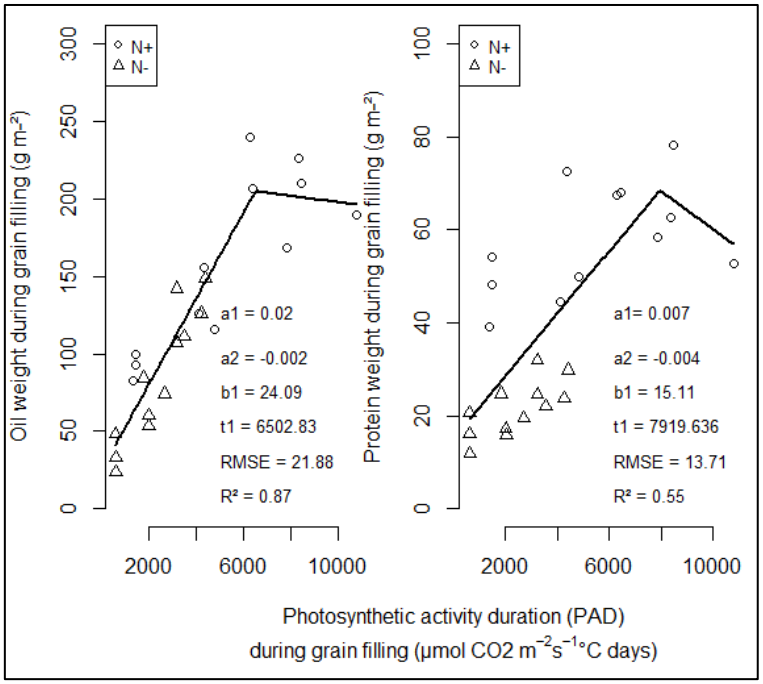


Fig.4.Relationship between achene oil and proteins weights per m² with photosynthetic activity duration (PAD) during grain filling, in cv.Kerbel and plant density D2 (2012 experiment) as a result of bi-linear model fitting. Model parameters were: a1: rate of oil or protein accumulation if any first phase; a2: rate of oil or protein accumulation if any second phase; b1: intercept value of oil or protein when PAD is null; t1: PAD threshold value at which oil or proteins accumulation rate decreases or increases. RMSE and R² values were computed.

Table.4. Summary table of parameters values as a result of model fitting of nitrogen weights dynamics in source organs (green leaves, receptacles and stems in *cv*.Kerbel, plant density: D2). Parameter values were averaged by nitrogen condition (N+: non-limiting; N-: limiting). Model parameters were rates of evolution in the first and second phase of dynamics (a1, a2), intercept (b1) and thermal time after start of flowering at which the second phase starts (t1). t1 corresponds to the start or triggering of source property for source organs.

	leaves N		receptacles N		stems N	
	N+	N-	N+	N-	N+	N-
a1	-0.0075	-0.0028	0.006	0.0015	-0.0003	-0.0012
a2	-0.0032	-0.0001	-0.0017	0.0001	-0.0008	0.0002
b1	7.43	2.51	0.24	0.09	1.82	0.98
t1	649.66	896.59	468.47	421.28	318.28	358.58
ft1	1.91	0.04	2.81	0.68	1.71	0.64
ty0	-	-	-52.3	-66.70	-	-
duration	513.99	336.91	695.18	89.96	845.37	874.92

3.4.2. Effect of plant density on source and sink dynamics

We analyzed plant density effects on source and sink dry weights dynamics in *cv*.Kerbel (D1 vs D2). We observed that density effect was significant only on hulls, receptacles and green leaves dynamics. 5 parameters were affected depending on the organ: a1 (hulls, receptacles), b1 (green leaves), t1 (receptacles), ft1 (hulls) and duration (green leaves, receptacles). Parameters of rates, quantities, and duration were globally higher in D2 than in D1. t1 parameter was lower in D2 compared to D1 (Table.6). Quality of fit was very satisfactory (mean R^2 =0.92). Interaction with nitrogen was found on a2 parameter in green leaves dynamics (D x N**).

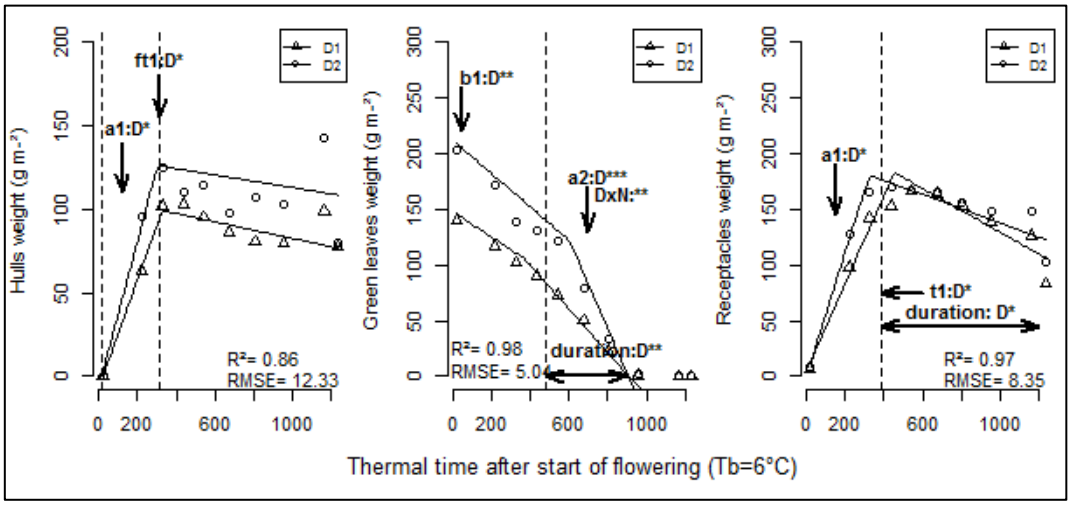


Fig.5. Representation of sinks (hulls) and sources (green leaves and receptacles) weights dynamics during grain filling in *cv*.Kerbel (2012 experiment). Symbols correspond to plant density modalities (D1, D2) and solid lines to result of the fit by a bi-linear model. Only sink and source indicators for which plant density effect was significant for at least one parameter are represented. Adjustment for green leaves dynamics was carried out by excluding values close to zero .Model parameters were rates of evolution in the first and second phase of dynamics (a1, a2), intercept (b1) and thermal time after start of flowering at which the second phase starts (t1). t1 corresponds to the start or triggering of source property for source organs. Complementary variables (ty0, ft1, duration) were computed as well as quality of model fit (R^2, RMSE). Dashed lines illustrate ty0 (sink) and t1 (sink and source) parameters. Only significant parameters are indicated by arrows on the graphs with their level of significance: "***" (p-value < 0.1%); "**" (p-value < 1%); "*" (p-value < 5%). p-values higher than 5% were considered non-significant. For parameter values, please refer to Table.6.

3.4.3. Effect of genotype on source and sink dynamics

Finally, genotype effect was analyzed on model parameters (*cv*.Kerbel and *cv*.Olledy). Significant genotype effect was detected on oil, hulls, green leaves, receptacles and stems weight dynamics (Fig.6). Affected parameters were a1 (hulls, green leaves), a2 (green leaves), b1 (receptacles, stems), t1 (green leaves, stems), ft1 (green leaves), ty0 (oil, receptacles) and duration (receptacles).Parameters of rates, quantities and duration were higher in *cv*. Olledy than in *cv*.Kerbel (Table.4). t1 parameter was significantly lower in *cv*.Olledy. Genotype interactions were found for green leaves (a1, a2, t1, duration) and for stems weight dynamics, whether with nitrogen, density, or both.

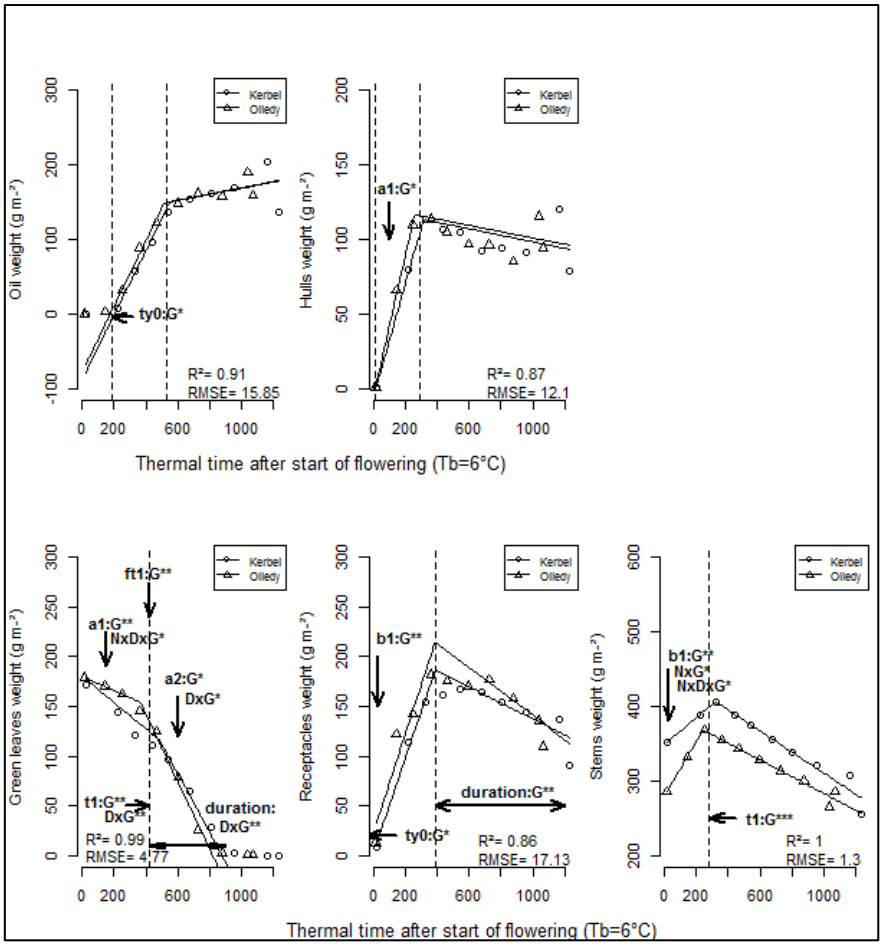


Fig.6. Representation of sinks (oil, hulls) and sources (green leaves, receptacles and stems) weights dynamics during grain filling in *cv*.Kerbel and *cv*. Olledy (2012 experiment). Symbols correspond to genotype modalities (*cv*.Kerbel, *cv*.Olledy) and solid lines to result of the fit by a bi-linear model. Only sink and source indicators for which genotype effect was significant for at least one parameter are represented. Adjustment for green leaves dynamics was carried out by excluding values close to zero. Model parameters were rates of evolution in the first and second phase of dynamics (a1, a2), intercept (b1) and thermal time after start of flowering at which the second phase starts (t1). t1 corresponds to the start or triggering of source property for source organs. Complementary variables (ty0, ft1, duration) were computed as well as quality of model fit (R^2, RMSE). Dashed lines illustrate ty0 (sink) and t1 (sink and source) parameters. Only significant parameters are indicated by arrows on the graphs with their level of significance: "***" (p-value < 0.1%); "**" (p-value < 1%); "*" (p-value < 5%). p-values higher than 5% were considered non-significant. For parameter values, please refer to Table.6.

3.5. Dynamic framework of source-sink relationships and orders of priority

We computed parameters of timing in source (t1) and sink (ty0) organs dynamics in order to (1) establish chronological source availability and sink demand along grain filling period (2) and check if there was a stable framework of source and sink relationships in sunflower. Mean values were computed by genotype and grouped by source and sink categories. Receptacles were included as sink organs since they accumulate dry matter in their first phase of weight evolution. Parameters of timings were compared by the means of Student t test (Table.5).

We observed that chronology of sink demand was similar in both genotypes: receptacles first, then hulls, proteins and oil. Concerning the chronology of source availability, stems were the first being mobilized as sources in both genotypes, followed by leaves then receptacles in cv.Olledy, while it was reversed in *cv*.Kerbel, though not significant.

Comparing the two genotypes, *cv*.Olledy significantly differed from *cv*.Kerbel concerning ty0 for oil and receptacles ($p = 0.02$ and 0.04 respectively), and t1 for green leaves ($p = 0.003$), receptacles ($p = 0.04$) and stems ($p < 2.2e\text{-}16$). For those parameters, *cv*.Olledy displayed lower values than *cv*.Kerbel.

Within a same genotype, there were no significant differences in all nitrogen and plant density conditions, neither for ty0 nor for t1 values, which indicates a relative stability of those parameters across cropping conditions (Table.6). Though, some exceptions were found, such as in *cv*.Kerbel where t1 for receptacles significantly differed in N+ x D1 situation compared to N- x D2, and in *cv*.Olledy where ty0 for oil was different between N+ x D1 and N- x D1, and for stems concerning N- x D1 and N-x D2.

Given that orders of priority were stable across genotypes, we proposed a conceptual source-sink framework for sunflower during grain filing (Fig.7). The framework was based on the principle that an available amount of mobile C or N in source –*i.e*. leaves, receptacles and stems- is usable for C and N grain filling at a time t. We used model parameters obtained in this study to compute mean values of start of accumulation (ty0) for sink and t1 for source property, plus durations of processes. For graphics readability, crop cycle duration was limited to end of oil accumulation.

Table.5. Relative chronology of parameters of timing (ty0, t1 and duration) obtained from a bi-linear model fitting. Parameters were averaged by sink and source groups of organs and by genotype. Mean values are provided as well as p-values from Student test. "<" symbol illustrates chronology between sink and source organs. ty0 corresponds to the value of thermal time at which sink or source weight is null. t1 refers to the start or triggering of source property for source organs. Duration equals the difference between t1 and ty0 for sink organs and difference between physiological maturity thermal time and t1 for source organs. p-values are indicated by their level of significance: "***" (p-value < 0.1%); "**" (p-value < 1%); "*" (p-value < 5%). P-values higher than 5% were considered non-significant (ns).

category	genotype	parameters of timing							
sink	*cv*.Kerbel	order of ty0	receptacles	<	hulls	<	proteins	<	oil
		ty0 p-value		*		ns		***	
		ty0 value (°Cd)	-7.72		12.9		20.5		204.2
		duration (°Cd)	397.9		302.5		494.2		468.8
	cv.Olledy	order of ty0	receptacles	<	hulls	<	proteins	<	oil
		ty0 p-value		.		ns		***	
		ty0 value (°Cd)	-69.6		11.2		22.3		173.6
		duration (°Cd)	456.9		248.5		416.4		513.5
source	*cv*.Kerbel	order of t1	stems	<	receptacles	<	green leaves		
		t1 p-value		*		ns			
		t1 value (°Cd)	322.4		390.2		484.2		
		duration (°Cd)	876.2		799.4		321.7		
	cv.Olledy	order of t1	stems	<	green leaves	<	receptacles		
		t1 p-value		ns		ns			
		t1 value (°Cd)	244.3		363.2		387.3		
		duration (°Cd)	736.5		365.4		639		

Table.6. Table of model parameters obtained from a bi-linear model fitting in 2012 field experiment. Data originate from observations of sink (oil, proteins, hulls) and source (green leaves, receptacles and stems) weights dynamics.Values were computed by genotype (*cv.*Kerbel and *cv.*Olledy), nitrogen condition (N+: non limiting, N-: limiting), and plant density (D1: 3 plants per m²; D2: 4.5 plants per m²). Model parameters were rates of evolution in the first and second phase of dynamics (a1, a2), intercept (b1) and thermal time after start of flowering at which the second phase starts (t1). t1 corresponds to the start or triggering of source property for source organs. Complementary variables (ty0, ft1, duration) were computed as well as quality of model fit (R^2, RMSE).

category	components	parameters	cv.Kerbel				cv.Olledy			
			N+		N-		N+		N-	
			D1	D2	D1	D2	D1	D2	D1	D2
sink	oil	a1	0.48	0.59	0.31	0.35	0.42	0.59	0.38	0.42
		a2	0.02	0.11	0.03	0.01	0.02	0.05	0.02	0.08
		b1	-95.67	-113.23	-65.49	-77.15	-60.28	-104.46	-75.2	-79.23
		t1	501.1	500.67	563.6	632.86	545.89	494.04	536.48	471.64
		ft1	144.68	175.93	110.56	141.47	169	187.32	127.39	112.91
		ty0	199.19	188.11	210.01	219.44	141.74	176.93	193.18	183.65
		duration	455.55	662.98	484.15	272.36	389.21	544.26	532.92	597.76
	proteins	a1	0.11	0.12	0.06	0.06	0.13	0.14	0.06	0.07
		a2	-0.002	0.02	0.01	0.02	-8.30E-05	0.02	0.003	0.02
		b1	-4.08	-0.47	-2.24	-1.21	-4.6	-4.57	-1.4	-0.4
		t1	524.15	588.18	479.72	466.6	495.56	461.74	505.26	322.29
		ft1	55.97	68.33	25.28	25.26	56.34	59.57	27.94	23.09
		ty0	35.33	-13.64	39.27	21.05	36.54	31.97	22.15	1.81
		duration	488.82	601.83	440.45	445.55	459.02	429.77	483.11	320.48
	hulls	a1	0.38	0.54	0.27	0.33	0.48	0.58	0.3	0.52
		a2	-0.04	-0.02	-0.01	-0.01	-0.02	-0.01	-0.03	-0.02
		b1	-7.89	-10.27	-5.92	-1.31	-6.42	-7.92	-1.96	-7.41
		t1	340.54	284.39	315.29	321.55	263.81	229.53	319.93	223.45
		ft1	120.72	142.02	79.38	93.67	120.2	123.1	91.84	106.99
		ty0	20.89	18.78	21.85	-9.77	11.67	13.67	4.9	14.42
		duration	319.64	265.62	293.44	331.32	252.14	215.87	315.03	209.04
source	green leaves	a1	-0.06	-0.16	-0.19	-0.13	-0.06	-0.05	-0.05	-0.12
		a2	-0.27	-0.53	-0.08	-0.19	-0.38	-0.47	-0.30	-0.26
		b1	170.44	262.31	115.13	137.34	183.94	249.39	156.38	154.24
		t1	398.10	587.23	324.65	618.36	375.02	320.16	345.76	397.73
		ft1	143.50	162.41	52.67	58.01	159.90	232.52	141.13	108.00
		duration	407.80	218.67	481.25	187.54	353.68	408.54	382.94	330.97
	receptacles	a1	0.52	0.67	0.27	0.43	0.67	0.65	0.28	0.44
		a2	-0.15	-0.11	-0.04	-0.02	-0.14	-0.12	-0.15	-0.07
		b1	2.44	2.82	5.89	-2.69	8.78	18.01	31.43	27.32
		t1	473.46	364.65	432.57	290.15	350.85	345.3	488.19	350.9
		ft1	239.31	243.13	122.92	119.79	229.54	239.45	167.3	163.78
		ty0	-10.44	-5.56	-21.16	6.27	-21.29	-29.09	-124.37	-90.15
		duration	690.19	799	800.93	961.3	657.29	722.26	581.21	665.85
	stems	a1	0.2	0.26	0.12	0.13	0.39	0.37	0.23	0.46
		a2	-0.16	-0.2	-0.13	-0.08	-0.16	-0.07	-0.09	-0.12
		b1	339.13	434.92	298.41	321.84	249.67	299.45	223.53	351.03
		t1	322.04	321.96	322.55	322.92	242	245.12	245.18	245.29
		ft1	404.2	519.4	337.21	364.43	345.78	389.75	280.96	464.64
		duration	841.61	841.69	910.95	910.58	796.3	793.18	824.22	551.26

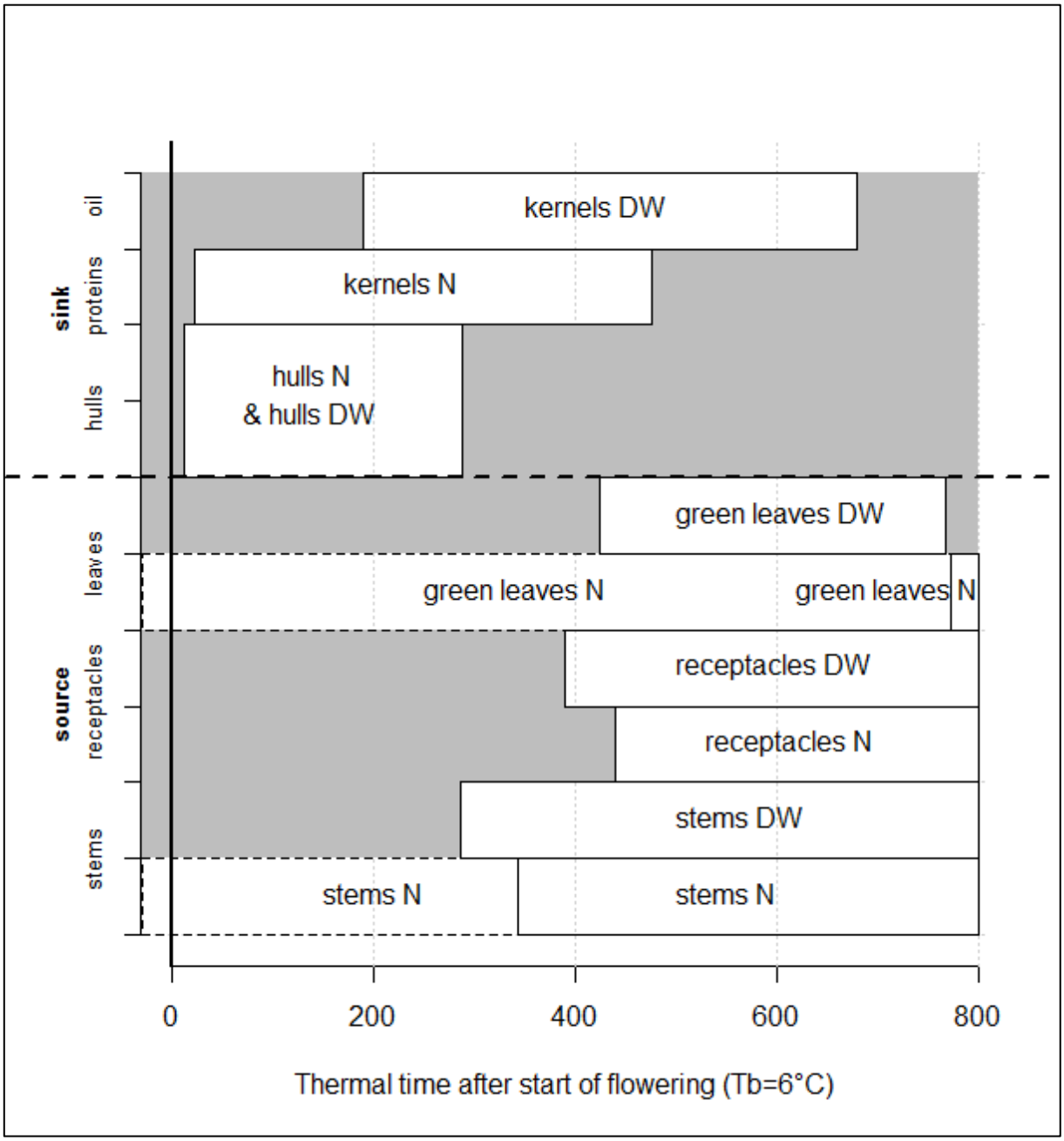


Fig.7. Established source-sink framework in sunflower across thermal time after flowering as averaged in two contrasting genotypes (*cv*.Kerbel and *cv*.Olledy, 2012 experiment). Main sink components were considered to be hulls, proteins, and oil. Source components were leaves, receptacles and stems. Boxes were limited by a left-border representing start of sink (ty0) and source (t1) property, and a right border as the sum of ty0 or t1 and duration of each process. Some literature-based hypotheses were made on N and C dynamics in source (stems) sink (hulls) when not available. They were materialized by dashed-border boxes. The framework relied on the hypothesis that all available sources at a thermal time t are usable for C and N grain filling. C originating from photosynthesis activity and N from roots absorption after flowering were not represented.

4. Discussion

In this paper, we aimed at bringing insights into nitrogen, plant density, genotype and their potential interactions effects on sunflower agronomic performances. For this purpose, we conducted a separate study of source (leaves, receptacles, stems) and sink (oil, proteins, hulls) components assuming that source dynamics modulate sink ones (Merrien, 1992). The final aim was to conceptualize a dynamic framework of source-sink relationships in sunflower.

Unravelling oil, proteins and hulls percentages values by corresponding weights to understand crop management and year effects

We first compared 2011 and 2012 experiments global agronomic performances under both N (limiting, non-limiting) treatments in cv. Kerbel. N+ x 2012 was globally more productive than N+ x 2011, except for protein concentration. N- x 2011 was more productive than N- x 2012 considering its protein variables and related source indicators. N- x 2012 very unfavorable post-flowering conditions (with quite dry July and August months) led to lowest N uptake after flowering, lowest oil, hulls and proteins weights per m² , and lowest source variables values except for biomass remobilization where it was highest (13.2 g m^{-2}). Globally, there tended to be more biomass remobilization in N- than in N+ in both years; for comparison, LAD was highest in N+ and lowest in N-. This reinforces the idea that remobilization contribution to yield increases while photosynthetic activity capacity diminishes in stressing environments (Blanchet et al., 1988; Sadras et al., 1993).

Differences in rankings between achene component percentages and weights occurred because components did not vary in the same proportions in response to nitrogen effect. Protein weights met the highest increase from N- to N+ in 2011 and 2012 (167 % and 150%, respectively). Hull weights increased by 67% in 2011 and 78 % in 2012; oil increased in both years, but more in 2012 than 2011 (80% against 59% respectively). Achene weights increased by 89% and 87% in 2011 and 2012. 2011 experiment was more favorable to proteins increase compared to that of oil; this could be due to lower rainfall during the pre-flowering and second part of grain filling periods, leading to the higher accumulation of lower-energetic compounds (Merrien, 1992; Aguirrezábal et al., 2009). Limited soil water availability could have led to observed higher hull thus achene weights per m², as suggested by Denis and Vear (1994). Note that high hull percent did not correspond to high hull weights per m².

Is there any equilibrium between DW remobilization and LAD and/or N remobilization and N uptake after flowering with experimental conditions?

We highlighted several adaptative behavior of genotypes depending on nitrogen and/or plant density conditions in 2012 experiment, such as higher nitrogen effect in *cv*.Kerbel on nitrogen uptake and dry weight remobilization, or varying effect of plant density (negative in *cv*.Kerbel, no effect in *cv*.Olledy) on protein concentration. N uptake after flowering and oil and proteins weights were positively correlated in both genotypes. LAD and N uptake after flowering were higher in N+; higher NNI and N remobilization were also found in those situations, but higher N remobilizations did not necessarily lead to higher post-flowering LAD and N uptake. The latter were positively correlated ($r = 0.63$; $p = 0.0004$). N remobilization appeared to rely only on available nitrogen at flowering (Ercoli et al., 2008), while post-flowering N uptake was conditioned by available nitrogen at flowering, sink size and water conditions during grain filling (Dordas, 2012). Both processes were not opposed.

Highest oil and hull weights per m² were obtained with highest N uptakes (N+ x 2012), while highest N remobilization led to highest protein weights. In those situations, oil accumulation was not systematically injured if post-flowering conditions were favorable. Oil accumulation proved to be highly related to photosynthetic activity duration ($R^2 = 0.87$) particularly during its highest rate accumulation period (200 to 500 °Cd after start of flowering). This is consistent with Aguirrezábal et al. (2003) who identified similar thermal time window for which oil is mostly sensitive to intercepted radiation. The lack of opposition between N uptake and N remobilization is in contrast with Pommel et al. (2006). It seems likely that sunflower can rely on both N uptake and N remobilization in contrast to maize.

Such finding highlights the importance of N uptake after flowering in sunflower. Borrell et al. (2001) integrated scheme for explaining stay green and non-stay green hybrids differences could apply to our finding ; in N+, demand is higher, so that higher N uptake occurs and helps to maintain nitrogen level in leaves thus delaying senescence, which in turn permits to increase N uptake capacity. However, senescence was not delayed but rather accelerated in N+ situations, even if LAD was higher. This could be explained by (i) difficulty to visually notate start of senescence in N- conditions (based on more than 50% yellow colour notation) or (ii) the fact that LAD was higher in N+ only because initial LAI was higher, but senescence was actually sooner. We lean toward the latter hypothesis since photosynthetic measurements tended to confirm the earlier senescence of leaves in non-limiting N conditions. We computed aboveground biomass acquired and N uptake after flowering contributions vs remobilized biomass and N contributions to respective grain DW and N weights per m², expressed in

percentages. In N+, aboveground DW acquired after flowering contribution to grain DW was 100% (i.e., no DW remobilization process) while it equaled 72 and 28% for N uptake after flowering and N remobilization respectively. In N-, those contributions were 63 and 37% for aboveground DW acquired and DW remobilization, and 53 and 47% for N uptake after flowering and N remobilization respectively. It is noteworthy that those contributions varied with genotype: N uptake after flowering tended to be higher in *cv*.Olledy while DW remobilization was clearly lower; *cv*.Olledy might have a lower remobilization capacity than *cv*.Kerbel but possibly higher photosynthetic capacity. Such differences in remobilization capacities in genotypes were also highlighted in Sadras et al. (1993).

About the relevancy of bi-linear models for analysis source and sink dynamics

For describing and comparing source and sink weights evolution with thermal time, we chose to use a simple bi-linear model with 4 parameters. Such model was found to be adapted for all sources and sink dynamics, so that their parameters could be mutually compared. Bi-linear or piecewise models are often used in agronomical analysis (Grimm et al., 1994; Dosio et al., 2000; Aguirrezábal et al., 2003; Izquierdo et al., 2008; Tanaka and Maddonni, 2009). These models have the advantage to be simple, lighter to compute and with physiologically interpretable parameters (Nickerson et al., 1989; Toms and Lesperance, 2003). Though, they are limited by their simplicity; let us consider the case where processes have more than 2 phases for instance or the case where growth or decrease is not linear but exponential. Yin et al (2009) proposed source and sink models through improved sigmoid curves; those could be tested for our sink and source (leaves) dynamics, especially in situations where quality of fit was unsatisfactory (N-). In any case, we were able to provide additional information from our model dynamics when comparing source and sink.

Did dynamics analysis help to understand nitrogen effects on source and sink?

Leaves senescence dynamics was similar to that of de la Vega et al. (2011), with a first period of low senescence rate followed by an increase of leaves deaths. Considering nitrogen effect on leaves senescence, leaves in N+ senesced faster than in N- starting from mid-grain filling. Senescence processes could likely be different depending on N situation (Pommel et al., 2006); in our study, senescence happened obviously once at a time in N+ (senescence of uppermost leaves corresponding to second phase of senescence dynamics), while it was more difficult to assess and probably more progressive in N-. For confirming start of senescence, one could measure SLN values in leaves and identify threshold of leaf senescence which could be different depending on N status.

All plant organs (receptacles and stems weights) were positively affected by nitrogen application (Ercoli et al., 2008). As for leaves, stems and receptacles weights decreased faster in N+ once they moved to source organs. As stated above, all achene components benefited from nitrogen effect, particularly through higher rates of accumulations.

When comparing source N timings, it is likely that there was a hierarchy in source organs mobilization; since N from stems was mobilized first and that the latter is highly linked to N uptake after flowering, it appears that N accumulation in grains relies on N uptake first; if necessary, N from receptacles (start of senescence) then green leaves take over. This is similar to sorghum (van Oosterom et al., 2010) or wheat (Jamieson and Semenov, 2000) N dynamics.

However, we want to point out that since green leaves weight displayed a decreasing pattern starting from the start of flowering (which is not the case of stems and receptacles), it suggests that they are the first organs playing as source. This confirms the limitation of our model as stated above: more than 2 phases should be considered in order to better estimate this first "source" phase start and duration.

Did dynamics analysis help to understand density nitrogen and genotype effects on source and sink?

Density displayed effects on rates of hulls and receptacles biomass accumulation, as well as quantities (b1: green leaves) and timings parameters (t1 and duration).

cv.Olledy showed higher values than *cv*.Kerbel for its timing of start of accumulation of oil and receptacles (ty0), rates of processes (a1 in hulls and a1 and a2 in leaves) and quantities (ft1 in leaves and b1 in receptacles). *cv*.Kerbel displayed higher t1 values than *cv*.Olledy (leaves and stems) and b1 value for stems. Meanwhile, we found that *cv*. Olledy displayed lower kernel oil percent than *cv*. Kerbel and its kernel protein percentage was higher (data not shown). In parallel, *cv*. Kerbel and *cv*. Olledy mainly differed in their source dynamics, where *cv*. Olledy seemed to lengthen source organs remobilization durations (receptacles, stems) probably in response to more marked leaves senescence.

From a global point of view, it appeared that nitrogen and density effects played through components accumulation (sink) or decrease (source) rates, while genotype effects played through initial values of source (ft1 for leaves, b1 for receptacles and stems) timings (ty0 for oil and receptacles, t1 for leaves and stems) and durations (stems and receptacles) of processes.

Is there a unique relationship between source and sink in sunflower?

In the nitrogen conditions tested in this study we found one positive bi-linear relationship between achene oil, proteins weights and photosynthetic activity duration. Distinction between N- and N+ situations helped to confirm that oil and protein weights were both enhanced in N+. Ruiz and Maddonni (2006) found a linear-plateau relationship between oil weight per achene and leaf area duration; we obtained better R^2 (R^2=0.41 vs 0.87 and 0.55) for our oil and protein weights per m^2 respectively.

It has never been attempted to link protein weights to green leaf area/ photosynthetic activity duration, although it is known that proteins accumulation is highest before grain filling but can be modulated by nitrogen factors thereafter (Picq and Abramovksy, 1989; Merrien, 1992). Changes in N nutrition modify proteins by oil ratio since both are influenced by nitrogen in different proportions.

We demonstrated that oil and proteins dynamics were highly correlated to leaves activity duration during grain filling. When compared to leaf area duration (LAD), photosynthetic activity duration (PAD) appeared to be a better source indicator when related to oil dynamics (R^2 =0.76 for LAD), but not to proteins where LAD was better (R^2 = 0.77). It is likely that PAD reflects intrinsic C instantaneous availability to grain oil which is the main source of C, while it is insufficient to represent N direct availability to proteins. By the fact that LAD is not as narrowly targeted as PAD, it might allow to take into account N dynamics (remobilizations, storage) inside plant. This would imply that the evolution of LAD reflects the evolution of N in leaves. This study suggested that N remobilization originated from receptacles, though its relationships with oil and proteins are not satisfactory (R^2=0.43 and 0.52 respectively). It is likely that receptacles could behave as transient compartments for N mobilized from leaves; excess N in receptacles could be used only when N remobilization from leaves and N uptake from roots are both limiting. Though we did not have sufficient date to confirm our hypothesis, this is consistent with the literature stating that the main source of N for grain proteins is leaves (Steer and Hocking, 1984; Bauchot and Merrien, 1988; Merrien, 1992).

Towards a source-sink framework for sunflower

This study enabled to better understand source, sink, and source-sink relationships in sunflower under varying crop management practices. It provides tracks for establishing a conceptual framework of source-sink mechanistic model in sunflower.

We confirmed chronology of sink demand along grain filling period, as stated in literature (Goffner et al., 1988; Connor and Hall, 1997): receptacles first, then grain components: hulls, proteins and oil. It was stable in both genotypes and all cropping conditions. From source DW and N timings computations and considering the limitations of our model as previously stated, we could hypothesize that leaves were the first organs playing source remobilization role (probably due to lower leaves death), followed by stems and receptacles. This framework appeared to be quite stable in all nitrogen, plant density and genotype conditions.

The first draft of sunflower source-sink relationships that we proposed seems justified since much of our results confirm what had been already long-established in other non-modelling literature experiments. We had to make some hypotheses since N and C in hulls were not available. We proposed that due to early senescence of basal leaves in sunflower after flowering (Merrien, 1992), it is probable that N in leaves decreases early after flowering (short remobilization). It is likely that N in stems could be used if N in leaves is not in excess (Steer and Hocking, 1984). Both serve as N source for N hull and kernel (proteins) accumulation from -12.1 to 477 °C days for hull N and proteins respectively. It is now about to put into test in the context of crop modelling. This source-sink framework, in combination with our bi-linear model could be used to analyze oil and/or proteins accumulation in other oilseed and protein crops.

5. Conclusions

This study aimed at analyzing nitrogen, plant density and genotype effects on dynamics of sink –oil and proteins- and source –leaves, receptacles and stems- in sunflower by the use of a bi-linear model. Comparison of model parameters showed that nitrogen and plant density effects played through rates of processes, while genotype affected timings and durations. N uptake mostly contributed to proteins weights while photosynthetic activity duration highly correlated to oil and proteins dynamics. N remobilization contribution to proteins increased in nitrogen deficient conditions. This study resulted in a source-sink based framework of oil and proteins accumulation in sunflower.

6. Acknowledgments

This work was supported by French National Research Agency (SUNRISE project 2012-19) and Association Nationale de la Recherche et de la Technologie (CIFRE N° 2010/1467).We greatly thank Laure Lagarrigue, Zakia El Mastouri and Yannick Dakaud for their help in field data collection, as well as Experimental Unit team (INRA Auzeville) and Ardon Analysis laboratory (CETIOM).

Bilan du chapitre II

- Nous nous sommes intéressés à l'effet du génotype, de l'azote et de la densité de peuplement sur les dynamiques des « sources » et des « puits » chez le tournesol à partir de la floraison. Nous avons également comparé les contributions relatives des processus de remobilisation et d'assimilation post floraison sous l'effet de ces différents facteurs.
- Nous avons pu établir une relation linéaire-plateau entre la durée d'activité photosynthétique d'une part et le poids d'huile et de protéines (g m^{-2}) d'autre part (R^2=0.87 et 0.55 respectivement).
- En conditions d'azote limitant, la contribution des processus de remobilisation du carbone augmente alors que la part de l'assimilation photosynthétique diminue. Nous n'avons pas constaté d'antagonisme dans les processus d'assimilation et de remobilisation de l'azote.
- Le facteur génotype influence les périodes (timings) et les durées des processus à l'échelle des sources (feuilles, réceptacles, tiges) et des teneurs finales en huile et protéines, tandis que les facteurs azote et densité affectent les vitesses d'accumulation.
- A partir de la comparaison des différentes périodes et durées, nous avons pu établir un cadre conceptuel (framework) des relations source-puits chez le tournesol.

Cette expérimentation a été réalisée dans des conditions non-limitantes en eau. Or, la disponibilité hydrique figure parmi les facteurs qui pénalisent à la fois le fonctionnement de la source et le bon remplissage du puits. Nous allons nous intéresser dans le chapitre suivant à l'effet du stress hydrique sur le fonctionnement de la « source » (assimilation carbonée) en phase post-floraison via l'analyse de la régulation des flux transpiratoires et photosynthétiques (à l'échelle de la feuille et de la plante) pour plusieurs génotypes.

Chapitre III: Effects of growth stage and leaf ageing on transpiration and photosynthesis response to water stress in sunflower

Article en préparation. Soumission prévue en Avril 2015 (Journal of Agronomy and Crop Science).

Effects of growth stage and leaf ageing on transpiration and photosynthesis response to water stress in sunflower

Fety Nambinina Andrianasolo [a,b,e], Pierre Casadebaig [b,e], Nicolas Langlade[c,d], Philippe Debaeke [b,e,1], Pierre Maury [b,e,1]

[a] CETIOM, Centre INRA de Toulouse, CS 52627, 31326 Castanet-Tolosan Cedex, France

[b] INRA, UMR AGIR, CS 52627, 31326 Castanet-Tolosan Cedex, France

[c] INRA, Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR441, F-31326 Castanet-Tolosan, France

[d] CNRS, Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR2594, F-31326 Castanet-Tolosan, France

[e] Université de Toulouse, INP, ENSAT, CS 52627, 31326 Castanet-Tolosan Cedex, France

[1] Co-advisors of the PhD thesis

Authors e-mail addresses:

fandrian@toulouse.inra.fr

pierre.casadebaig@toulouse.inra.fr

nicolas.langlade@toulouse.inra.fr

maury@ensat.fr

debaeke@toulouse.inra.fr

Corresponding author: maury@ensat.fr

Phone number: +33534323897

Abstract

Water stress regulates leaf transpiration rate and photosynthetic activity. The genotype-dependent response of the latter was not assessed in sunflower (*Helianthus annuus* L.), particularly during reproductive period where grain filling and lipogenesis highly depend on photosynthates availability. For evaluating genotypic responses to water deficit before and after flowering, two experiments were conducted in greenhouse (Exp.I and Exp.II) where inbred lines (PSC8, XRQ) and cultivars (Inedi, Melody) were submitted to progressive water deficit. Non-linear regression was used to compute the threshold (FTSWt) at which processes (transpiration and photosynthetic activity) began to be affected by water deficit. In vegetative growth stage, photosynthetic activity was affected at a lower soil water deficit threshold (FTSWt = 0.39) than transpiration (FTSWt = 0.53). However, in reproductive stage, photosynthetic activity was more sensitive to soil water deficit. We found significant plant growth stage effect ($p = 0.02$) on the difference between photosynthesis and transpiration rates thresholds, as well as leaf age effect on transpiration process ($p = 0.03$; Exp.I). Fully developed leaves could be considered as representatives of whole-plant transpiration pattern. Such results will help to lighten phenotyping methods and provide tracks for integrating genotypic variability to crop models.

Keywords: genotype, Net CO2 assimilation rate, senescence, transpiration, water stress*, Helianthus annuus* L.

1. Introduction

Water stress is one major limiting factor for growing sunflower crop (Helianthus annuus L.) in Southern Europe. Indeed, sunflower is cultivated during summer period where evaporative demand is high, particularly during grain filling. In addition, the crop is mostly confined to shallow soils where water scarcity often occurs (CETIOM, 2006; Casadebaig, 2008). Despite sunflower is deemed to tolerate water stress, the crop has to be managed properly to optimize grain production and quality. Convenient management especially implies the right choice of genotypes, sowing dates, planting density and eventually the timing of irrigation regarding pedo-climatic conditions (CETIOM, 2014).

Patterns of water stress response in sunflower are similar to most cultivated species (Hsiao, 1973; Chaves et al., 2002). They consist in early and progressive leaf stomatal closure due to high evaporative demand in atmosphere and/or soil dryness and loss of leaf turgor. Regulation of stomata is likely governed by ABA (Chaves et al., 2002; Pantin et al., 2012). Reduction of carbon assimilation –photosynthesis- can later occur because of lower possibility of carbon uptake and consequent down-regulation of biochemical demand for carbon dioxide (Chaves et al., 2002). Besides, Connor and Hall (1997) reported that mechanisms involved in water stress response varied with growth stage; before flowering period, leaf conductance is considered to be little sensitive to water stress, while after, the regulation of stomatal aperture highly control plant water status, along with interactions between senescence and reduced light interception. In this case, the progressive closure of stomata in vegetative period allowed photosynthesis to be maintained. Since elaboration of grain and oil highly depend on available carbohydrates mostly originating from photosynthetic activity after flowering (Merrien, 1992), any impairment of photosynthesis due to water stress could likely reduce grain production and oil content. Sensitivity of photosynthesis to water stress and its relationship to stomatal or non-stomatal limitations after flowering still needs to be investigated in sunflower.

Water stress revealed high genetic variability for photosynthesis and water status regulation when compared with well-watered sunflower genotypes (Maury et al. 1996; Maury et al., 2000; Kiani et al., 2007a; Kiani et al., 2007b). Genotypic differences were also evidenced for transpiration response to water stress (Casadebaig et al., 2008) in experiments where plant water status was expressed as a fraction of transpirable soil water (FTSW). From the comparison of 25 genotypes, early-regulating (conservative) and late-regulating (productive) strategies could be distinguished during vegetative stage. In the SUNFLO model, Casadebaig et al. (2011) assumed that photosynthetic activity was regulated after transpiration rate; they

used an offset parameter value for distinguishing both processes. We propose to check whether this delay between transpiration and photosynthetic activity rates varies with genotype and/or the plant growth stage.

Indeed, leaf types – or ages - involved in vegetative and reproductive periods differ in their functioning, growing history, micro-climatic environment, carbon metabolism (Danuso et al., 1988) and senescence (Agüera et al., 2012). For example, young leaves could still undergo different stages of cell expansion and/or division while fully expanded leaves reach their highest photosynthetic rates (Pantin et al., 2012; Nooden et al., 2012); in mature senescing leaves, ageing progressively leads to accumulation of soluble sugars, decrease of photosynthesis, degradation of chlorophyll and photosynthetic system triggered by oxidative stress (Aguëra et al., 2012). Sensitivity to water stress in these leaves types could also differ; it was demonstrated that young leaves accumulated more proline content compared to mature leaves when exposed to water stress; impairment of stomatal conductance and photosynthetic rate were more pronounced in mature leaves (Cechin et al., 2006; Cechin et al., 2010). Yegappan et al. (1982) argued that impact of water stress depended on the time of leaf life at which the stress occurred, and on the intensity of stress: mild stress affected unfolding leaves while still unfolded and expanding ones were only sensitive to severe stress. Neither differences in transpiration rate regulation between leaf ages nor differences between genotypes were assessed in those experiments. Could differential plant growth stage effect on transpiration be explained by differences in leaf types differences for their regulation of stomatal conductance?

Fully developed leaf type is usually taken for high throughput varietal assessment of water stress response (Cechin et al., 2006; Casadebaig et al., 2008; Cechin et al., 2010). Though, proportions of types of leaves vary across growth cycle and different types and ages co-exist at a given period (Pantin et al., 2012). It is thus questionable to use only one leaf type for representing a whole-plant functioning, particularly when considering differences before and after flowering; legitimacy for taking a type or another should be validated.

The aims of this study were to (i) analyze transpiration and photosynthetic activity response to water deficit in vegetative and reproductive stages in contrasting sunflower genotypes (ii) evaluate leaf age effect on stomatal regulation process (iii) and thereby identify the type of leaf best representing whole plant transpiration rate regulation under water deficit.

2. Materials and methods

2.1. Experiments design

Two greenhouse experiments were carried out at INRA station in Auzeville, France (43°31'41.8" N, 1°29'58.6" E) in 2009 and 2012. Two sunflower inbred lines (PSC8 in 2009, XRQ in 2009 and 2012) chosen because Rengel et al. (2012) previously characterized their contrasted behaviours under water deficit, and two contrasting commercial F1 hybrids (Inedi and Melody in 2012) were germinated on Petri discs. Plantlets were rapidly transferred to large 15-L individual pots filled with a mixture of 50% clay loam, 40% P.A.M.2 potting soil (Proveen distributed by Soprimex, Chateaurenard, Bouches-du-Rhône, France) and 10% sand. Seeds were sown on 1st April in 2009; two dates of sowing were set in 2012 experiments in order to obtain similar environmental conditions for monitoring simultaneously vegetative and reproductive growth stages: 24th April and 16th March, for vegetative and reproductive stages respectively. Pots were randomly distributed in the greenhouse but replicates were gathered in different blocks. There were 5 (2009) and 6 blocks (2012) and a total of 20 and 72 pots in 2009 and 2012, respectively. Plants were adequately irrigated and fertilized before dry-out in both experiments. Relative humidity of air and temperature inside and outside the greenhouse were recorded.

In next sections, 2009 and 2012 experiments will be referred Exp.I and Exp.II, respectively.

2.1.1. Water deficit pattern

Right after reaching the 8-leaf stage in Exp.I and Exp.II "vegetative growth stage" experiment and at full flowering stage (R5.5, Schneiter and Miller, 1981) in Exp.II "reproductive growth stage" experiment, we paired pots by stress/control modality. All pots were re-watered to full soil water saturation capacity the day before the water stress experiment; then, no more water was brought to water stressed plants till the end of the experiment. Irrigated pots were daily re-watered to full soil water capacity. It is noteworthy that daily extraction of soil water was very high in reproductive period experiment due to higher plant leaf area. To manage comparable water content dynamics in both phenological stages, stressed plants in reproductive stage were re-watered when necessary.

2.1.2. Measurements

2.1.2.1. Leaf transpiration

Leaf transpiration rate (TL (mmol H2O m-2 s-1) was measured with a porometer (LI-1600, LI-COR inc., Lincoln, NE, USA) in Exp.I and a portable gas exchange system (LI-COR 6400, Lincoln, Nebraska, NE, USA) in Exp.II.

Transpiration rate was monitored on a reference fully expanded leaf (number 9 to 11 from the bottom of the plant) every day at 10 a.m in Exp.I and II vegetative growth stage. This leaf type will be further noted "developed" and corresponds to a dark green leaf, supposedly displaying its highest photosynthetic rate and having recently reached its maximum size; a leaf was considered as "developed" at700°C day from plant emergence on average (Dosio et al,. 2003). In reproductive growth stage (Exp.II), the "developed" leaf was chosen at two-third of leaf number from the bottom of the plant (i.e. nodes between 18 and 22), assuming that this upper part of the canopy mostly contributes to total plant carbon assimilation (Alkio et al., 2003): its age was 900°C days on average at the start of the experiment. Two other leaf nodes/ages were considered in Exp.I: a mature fully expanded leaf (noted "mature") which corresponded to "developed" one at the start of experiment and will be in post-expansion phase during the experiment. The third leaf type was noted "young" and corresponded to a green expanding leaf. It is noteworthy that "developed" leaf node varied with plant growth in order to obtain similar thermal ages, and that "young" leaf was systematically chosen 3 nodes above "developed" ones. "Mature" and "young" leaves were 550 and 430 °C days old on average, respectively. For further comparison of transpiration regulations between growth stages and genotypes, we computed a normalized indicator of leaf transpiration (NTL), corresponding to the ratio between stressed and irrigated values.

2.1.2.2. Leaf net photosynthesis

Leaf net photosynthesis rate (A, µmol CO_2 m^{-2} s^{-1}) of the mature fully expanded leaf was measured with the LI-COR 6400 device (Exp.II). Measures were performed under satured radiation value of 1500 $\mu mol.m^{-2}.s^{-1}$. Leaf carbon dioxid concentration was 400 $\mu mol.m^{-2}.s^{-1}$ As for leaf transpiration, leaf net photosynthesis was converted into a normalized ratio between net photosynthetic rate of stressed and control plants (NA).

2.1.2.3. Daily plant transpiration

Plant transpiration was obtained from the daily weighing of pots and daily measurements of leaf area following Casadebaig et al. (2008) method. When a leaf displayed more than 50% senescence (yellowing or browning), its area was discarded from leaf area daily measurement. Surfacic plant transpiration (TP, kg m-2 day-1) was obtained by dividing daily water loss by plant leaf area. Values of plant transpiration were normalized (NTP) for further water use dynamics comparison.

2.1.2.4. Plant water status

FTSW (fraction of transpirable soil water) was used as indicator of daily plant water status. It was computed from daily stressed pot weight ($pot\ weight\ j$), pot weight at saturation water capacity ($pot\ weight\ sc$) and pot weight when leaf transpiration of stressed pot is lower than 10% of its corresponding control pair (NTL < 10%; $pot\ weight\ 10\%$), such that:

$$FTSW = \frac{(pot\ weight\ sc - pot\ weight\ j)}{(pot\ weight\ sc - pot\ weight\ 10\%)} \quad eq(1)$$

2.2. Modelling transpiration and photosynthesis response to soil water deficit and statistical analysis

Dynamics of transpiration and photosynthesis regulation in relation to FTSW were adjusted with a modified version of Casadebaig et al. (2008) model:

$$y = \frac{1}{1 + 4.5 \times exp(a \times FTSW)} \quad eq(2)$$

where y corresponds to the physiological process (*i.e.* NTL, NA or NTP) and a to the model parameter of transpiration or photosynthetic rate regulation. For higher a value, the considered process is affected at a higher FTSW. Fits were performed with R software (R version 3.0.2) using *nls* regression. Quality of fit (RMSE and R^2) was assessed. Values of a were compared between genotypes, leaf ages and plant growth stages using analysis of variance. Fisher's Least Significant Difference (LSD) test was used for establishing groups when genotype, plant growth stage or leaf age effects were significant. Corresponding FTSWt values, *i.e.* values of FTSW at which transpiration or photosynthesis starts to decrease, are computed from eq (2) assuming that FTSWt was achieved when maximal normalized variables were reduced by 0.25%; this threshold was chosen in order to assess the right timing

at which processes start to be regulated under water deficit. Higher FTSWt values could be interpreted as higher "sensitivity" of a given process to water deficit.

Correlation between plant and leaf transpiration rate (NTP and NTL) was assessed using correlation and Student test in R (R Core Team, 2014).

3. Results

Results about differences in normalized leaf transpiration (NTL) and photosynthetic activity (NA) thresholds before and after flowering (Exp.II) will be considered first; then, an analysis of leaf types transpiration regulation will be performed (Exp.I) before trying to link leaf transpiration (NTL) to plant transpiration (NTP) in both experiments.

3.1. Vapor pressure deficit (VPD), transpiration and photosynthesis ranges

Vapor pressure deficit (VPD, kPa) was computed from daily internal and external air relative humidity and temperature greenhouse data in Exp.I and Exp.II (Fig.1). VPD in Exp.I was particularly high compared to Exp.II: it varied from 0.8 to 1.6 kPa in the former, while it was 0.5 to 0.9 kPa in the latter. VPD differed by 0.45 kPa at the start of each experiment but highest gaps were observed in the first half of experiments period, with differences reaching up to 1.0 kPa. To reach FTSW values of 0.1, water stress duration was 14 and 20 days in Exp.I (2009) and Exp.II (2012) experiments respectively, because of the VPD differences.

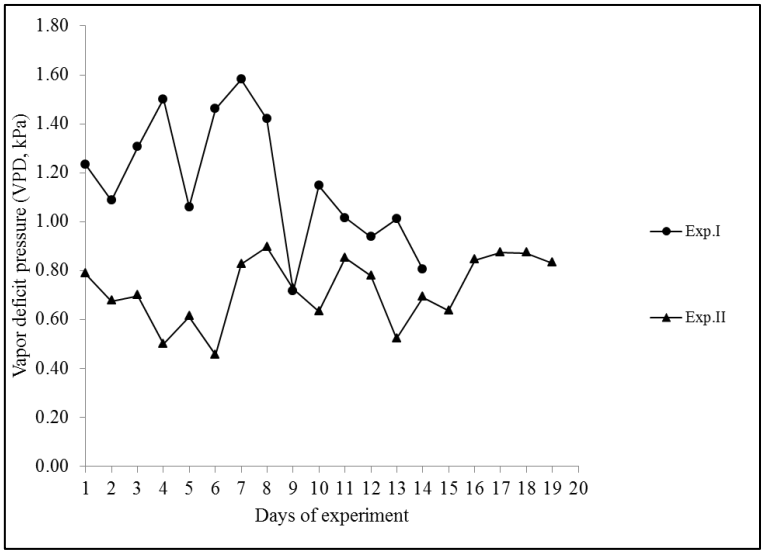


Fig.1. Evolution of vapor pressure deficit (VPD, kPa) in experiments I and II.

Leaf transpiration values were higher in Exp.I experiment (Table.1) in both stressed or control situations (mean 6 mmol m^{-2} s^{-1} in Exp.II vs mean 8 mmol m^{-2} s^{-1} in Exp.I). Plant transpiration was also about twice higher in Exp.I: values ranged from 1.5 to 3.4 kg m-2 day-1. Net photosynthetic rate ranged from 13.7 to 26.1 µmol CO_2 m^{-2} s^{-1} in Exp.II. Values were lower in stressed plants.

	A (μmol CO2 m^{-2} s^{-1})		TL (mmol H_2O m^{-2} s^{-1})		TP (kg m^{-2} day^{-1})	
	stressed	control	stressed	control	stressed	control
Exp.I						
"mature" leaves	n/a		7.1 ± 3.8	9.0 ± 2.8	3.2 ± 3.4	3.6 ± 3.3
"developed" leaves			8.2 ± 3.9	10.1 ± 2.6		
"young" leaves			6.1 ± 3.3	7.7 ± 2.7		
Exp.II						
Vegetative growth stage	17.7 ± 11.5	26.1 ± 6.0	4.6 ± 3.1	7.8 ± 1.5	1.1 ± 0.9	1.9 ± 0.7
Reproductive growth stage	13.7 ± 9.2	18.3 ± 5.8	4.9 ± 3.3	7.0 ± 2.0	1.3 ± 0.7	1.6 ± 0.6

Table.1. Net photosynthetic rate (A), leaf transpiration (TL) and plant transpiration (TP) values in Exp.I and Exp.II. Means values ± standard deviations are presented. Stressed and control situations were distinguished as well as leaf age in Exp.I experiment and growth stage in Exp.II. A was not measured in Exp.I.

3.2. Regulation of leaf transpiration and photosynthesis in vegetative and reproductive growth stages

The Inedi, Melody and XRQ genotypes were compared for their normalized leaf transpiration (NTL) and photosynthesis (NA) regulation in vegetative and reproductive periods. Values of a for NTL tended to be lower in reproductive stage (Table.2) compared to vegetative one. Growth stage effect was significant (-13.7 for vegetative and -17.9 for reproductive growth stages, $p < 0.01$). Neither genotype nor growth stage effects were detected on a.NA (mean value of -18.7). However, the effect of the interaction between genotype and growth stage was significant ($p < 0.01$). When computing the difference between values of a for NTL and NA, we detected a significant growth stage effect ($p<0.01$) and growth stage by genotype effect ($p < 0.05$) (Table.3). In vegetative period, leaf transpiration rate was regulated earlier than

photosynthesis and there was a delay between both processes (FTSWt = 0.4 and 0.6 for NA and NTL, respectively). In reproductive period, only Melody regulated leaf transpiration before leaf photosynthetic activity (a.NTL- a.NA > 0). For the two other genotypes, photosynthetic activity was regulated at higher FTSW threshold than transpiration rate (mean FTSWt was 0.46 and 0.53 for NTL and NA respectively; Table.2). We could note the higher variability of mean a.NTL- a.NA values in reproductive period. (Table.3 and Table.4).

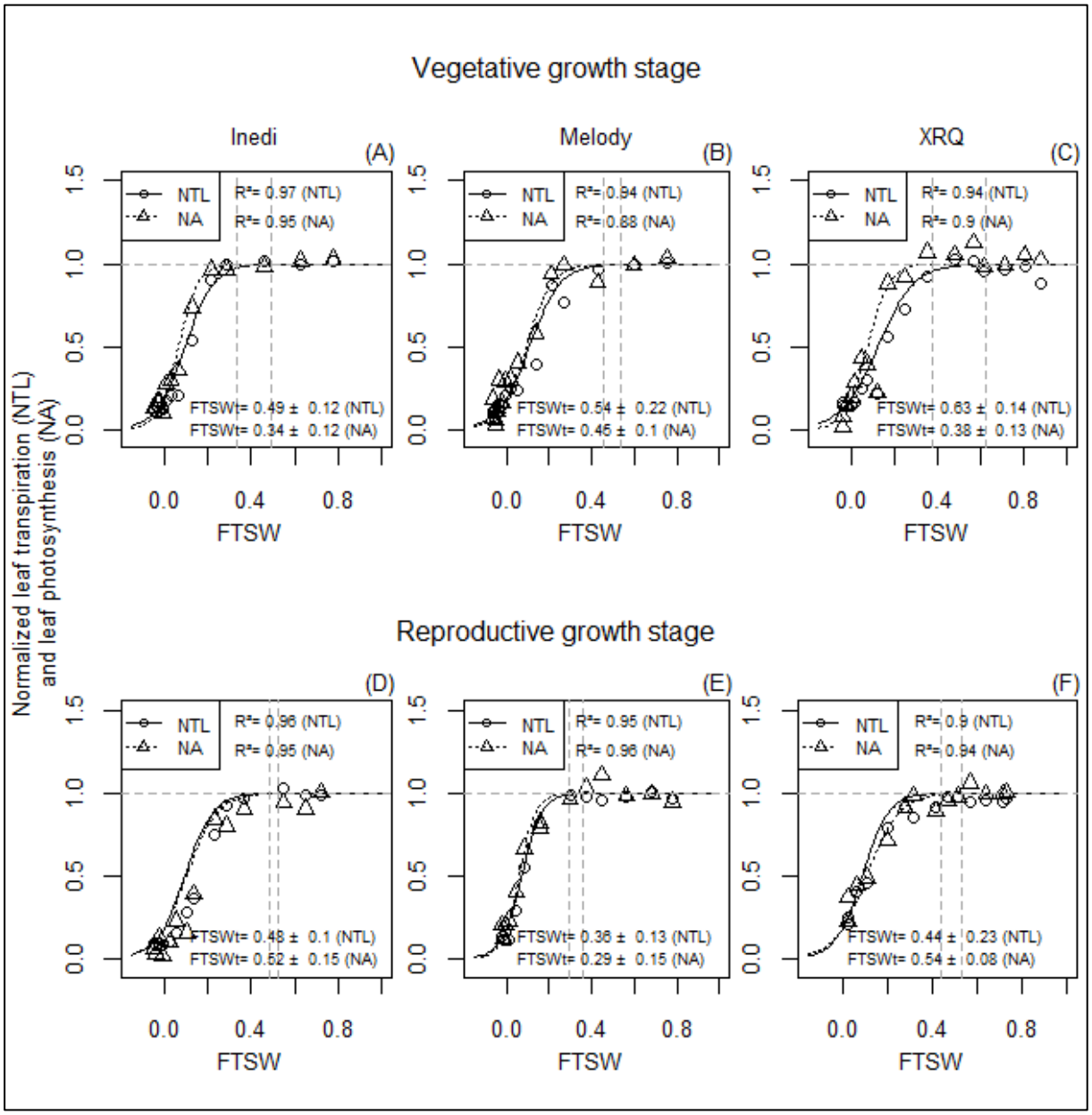


Fig.2. Responses of normalized leaf transpiration rate (NTL) and photosynthetic activity (NA) to available plant water content (FTSW) in two plant growth stages: vegetative (noted A, B, C) and reproductive (noted D, E, F) periods and 3 genotypes (Inedi, Melody, XRQ) during Exp.II. Quality of fit (R^2) and corresponding FTSWt threshold values and standard deviations are provided for each process.

Table.2. Summary table of response parameters to water stress for normalized leaf photosynthesis (NA), leaf transpiration (NTL) and plant transpiration (NTP) indicators. Fitted values of a (FTSWt) and indicators of quality of fit (root mean squared error RMSE and R^2) are presented per genotype and growth stage for Exp.II.

plant growth stage	genotype	parameters	NA	NTL	NTP
Vegetative growth stage	Inedi	*a*	-22.31	-15.31	-12.58
		FTSWt	0.34	0.49	0.60
		RMSE	0.11	0.09	0.11
		R^2	0.95	0.97	0.95
	Melody	*a*	-16.48	-13.93	-15.0
		FTSWt	0.45	0.54	0.50
		RMSE	0.16	0.10	0.11
		R^2	0.88	0.94	0.95
	XRQ	*a*	-19.82	-11.97	-19.02
		FTSWt	0.38	0.66	0.39
		RMSE	0.14	0.10	0.13
		R^2	0.90	0.94	0.89
Reproductive growth stage	Inedi	*a*	-14.35	-15.51	-17.98
		FTSWt	0.52	0.48	0.42
		RMSE	0.11	0.10	0.18
		R^2	0.95	0.96	0.91
	Melody	*a*	-25.60	-21.08	-25.96
		FTSWt	0.29	0.36	0.29
		RMSE	0.10	0.11	0.12
		R^2	0.96	0.95	0.88
	XRQ	*a*	-14.00	-17.14	-29.09
		FTSWt	0.54	0.44	0.26
		RMSE	0.11	0.09	0.12
		R^2	0.94	0.90	0.78

Table.3. ANOVA table on the difference between response parameters of normalized leaf transpiration and photosynthesis (a.NTL- a.NA) in Exp.II. Significant effects are indicated by: «*» (p-value < 5%).

a.NTL-*a*.NA source of variation	Df	Sum Sq	Mean Sq	F	p
growth stage	1	264.6	264.65	6.61	0.02*
genotype	2	8.4	4.21	0.11	0.90
block	2	22	11	0.28	0.76
genotype x growth stage	2	277.5	138.73	3.47	0.05*
Residuals	25	1000.6	40.02		

Table.4. Summary table of mean values ± standard deviations of the differences in response parameters between normalized leaf transpiration and leaf photosynthesis (a.NTL- a.NA) in Exp.I. Significant effects of genotype and growth stage were tested with analysis of variance. Groups were established from LSD test and similar ones share at least one letter.

a.NTL-a.NA	Growth stage	
Genotype	Vegetative	Reproductive
Inedi	7.0 ± 7.91 ab	-2.17 ± 7.20 bc
Melody	2.55 ± 2.5 abc	4.52 ± 5.99 ab
XRQ	7.85 ± 5.62 a	-3.14 ± 6.79 c

3.3. Regulation of leaf transpiration depending on leaf age

In Exp.I, leaf transpiration (NTL) was compared between 3 leaf ages and 2 genotypes (XRQ and PSC8). Leaf age and genotype had both significant effects ($p < 0.01$ and $p < 0.05$ respectively) on a.NTL (Fig.3 and Table.5). In both genotypes, "mature" and "young" leaves regulated leaf transpiration rate at similar FTSWt threshold value (mean FTSWt = 0.78). "Developed" leaves were significantly less sensitive for leaf transpiration rate than the two other leaf ages (FTSWt for transpiration in "mature" and "young" leaves equaled 0.8 while it was 0.2 and 0.5 for "developed" leaves in PSC8 and XRQ, respectively; Table.6). The lower sensitivity of developed leaves transpiration regulation in PSC8 contributed to significant genotypic difference in a.NTL parameter. FTSW thresholds for transpiration rates were generally lower at plant than at leaf level, because transpiration rates at leaf level were measured at conditions where vapor pressure deficit (VPD) were higher.

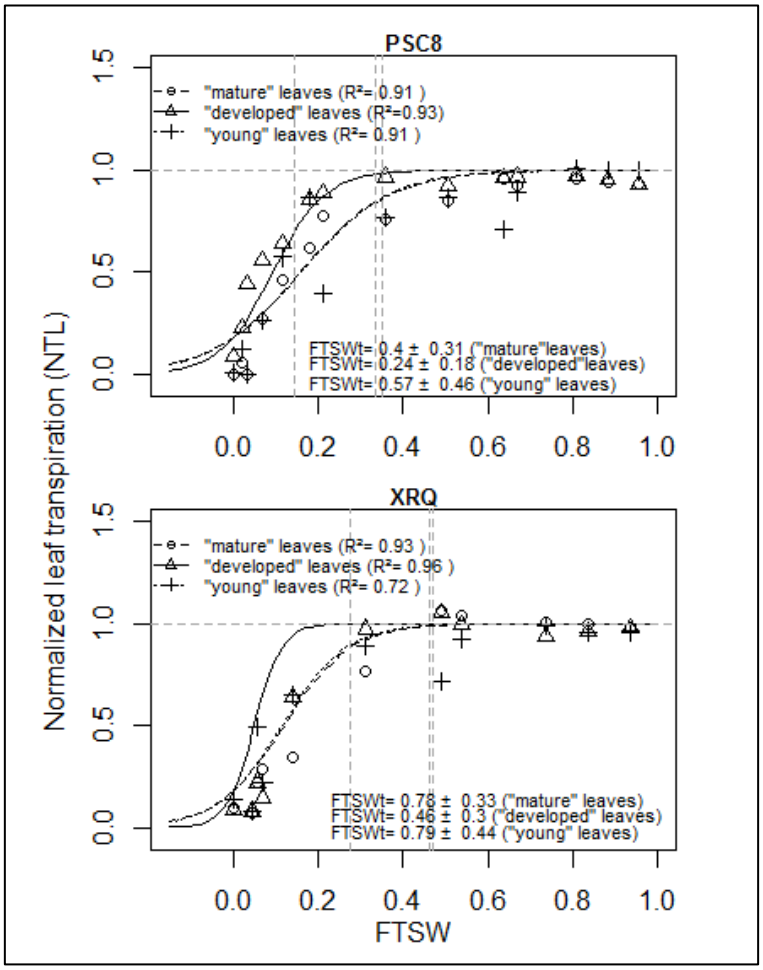


Fig.3. Responses of normalized leaf transpiration (NTL) to plant water stress (FTSW) in 2 genotypes (PSC8, XRQ) and 3 leaves ages in Exp.I. Leaf ages were categorized into mature expanded leaves ("mature"), young expanded leaves ("developed") and young expanding leaves ("young"). Quality of fit (R^2) and corresponding FTSWt threshold values ± standard deviations are indicated. Note that fitted curves for "mature" and "young" leaves overlap.

Table.5. ANOVA table on the values of a parameter for normalized leaf transpiration ratio (NTL) in Exp.I. Level of significance is represented by: «*» (p-value < 5%).

a .NTL source of variation	Df	Sum Sq	Mean Sq	F	p
genotype	1	625.80	625.80	6.41	0.02*
leaf age	2	830.70	415.30	4.26	0.03*
Block	1	65.40	65.40	0.67	0.42
genotype x leaf age	2	151.00	75.50	0.77	0.47
Residuals	23	2244.30	97.60		

Table.6. Summary table of response parameters to water stress for normalized leaf and plant transpiration ratio (a.NTL and a.NTP respectively) during Exp.I. Fitted values of a (FTSWt) and indicators of quality of fit (root mean squared error RMSE and R^2) are presented, per genotype and leaf age.

plant growth stage	genotype	parameters	NTL			NTP
			"mature" leaves	"developed" leaves	"young" leaves	
Vegetative growth stage	PSC8	*a*	-9.79	-31.01	-9.75	-24.65
		FTSWt	0.77	0.24	0.77	0.30
		RMSE	0.12	0.09	0.11	0.13
		R^2	0.91	0.93	0.91	0.86
	XRQ	*a*	-9.63	-16.26	-9.48	-28.48
		FTSWt	0.78	0.46	0.79	0.26
		RMSE	0.12	0.10	0.19	0.14
		R^2	0.93	0.96	0.72	0.81

3.4. Upscaling from individual leaf to whole plant transpiration

We checked the relationship between leaf (NTL) and plant (NTP) normalized transpiration rates by computing R^2 and its significance for each leaf age. In Exp.I, transpiration for each leaf age significantly correlated to the entire plant transpiration (Fig.4). Best relationships were obtained with either “mature” or “developed” leaves (R^2 = 0.67 and 0.68 respectively). R^2 value averaged 0.65 in Exp.I while it was higher (0.83) in Exp.II, with a slightly better relationship at the vegetative growth stage (Fig.5).

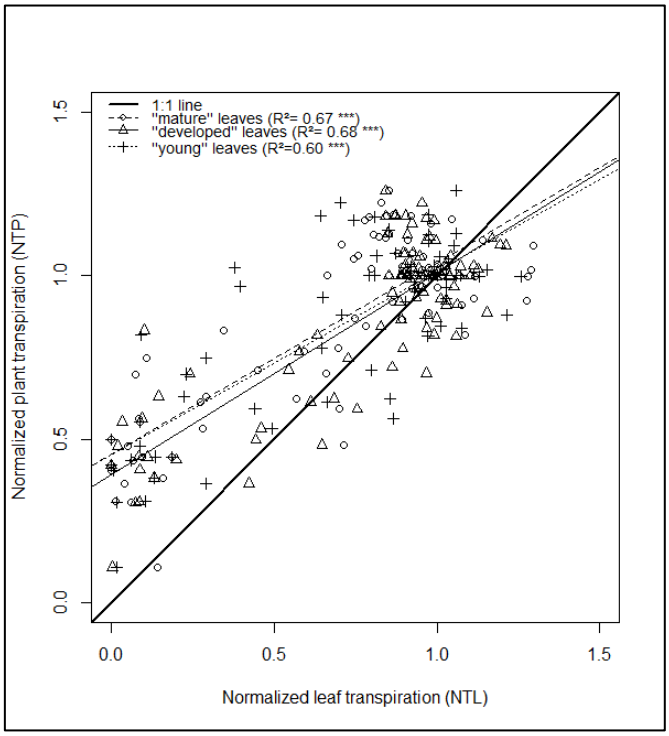


Fig.4. Relationship between normalized leaf (NTL) and plant transpiration (NTP) indicators in Exp.I. Relationships were assessed by leaf age.

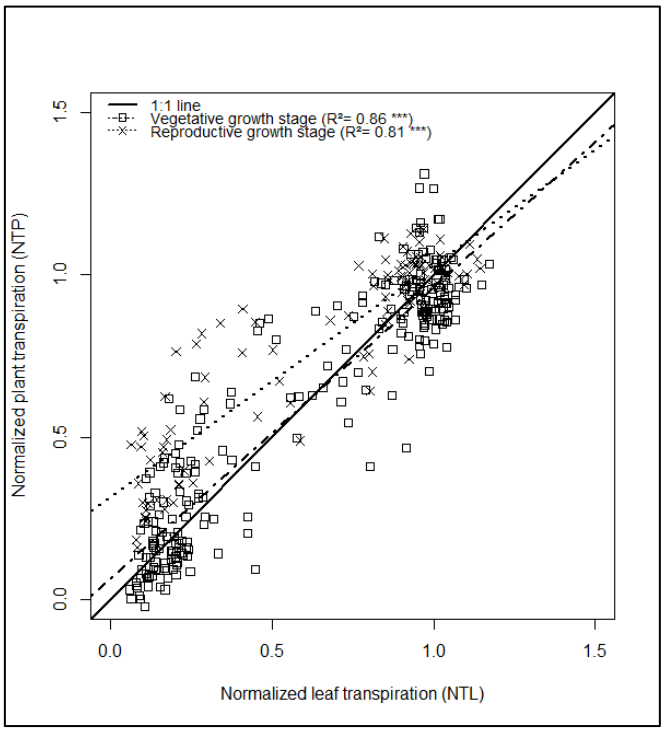


Fig.5. Relationship between normalized leaf (NTL) and plant transpiration (NTP) indicators in Exp.II. Relationships were assessed by plant growth stage.

4. Discussion

The greenhouse experiments we conducted were complementary for better understanding water stress regulation mechanisms with plant ageing in sunflower. Indeed, the experiment in Exp.I helped to provide answers to the differential regulation of leaf transpiration rate in vegetative and reproductive period (Exp.II). Despite contrasted differences in climatic conditions, we were able to find solid relationships.

Changes in vapor pressure deficit (VPD) affect plant transpiration rate regulation

Comparison of vapor pressure deficit of the two years experiment showed that evaporative demand was very high in Exp.I compared to Exp.II; this is mainly explained by higher temperatures in the former experiment (mean 23 ± 0.8°C during whole Exp.I, mean 20 ± 1.0°C in Exp.II). When we checked the regulation of leaf transpiration rate in a genotype that was present in both years (XRQ in vegetative growth stage, "developed" leaf), it appeared that leaf transpiration ratio (NTL) decreased at higher FTSWt in Exp.I (FTSWt = 0.78) than in Exp.II (FTSWt = 0.63). A greater evaporative demand generally increases the thresholds of sensitivity of stomatal conductance to drought (Sadras and Milroy 1996). Though, the objective of this study was not to compare yearly variation of water stress responses but rather to establish differences according to leaf age and growth stage; it was assumed that normalization of processes legitimizes the comparison of both experiments.

At a given plant growth stage, the leaf population showed different sensitivity to water deficit

Whether or not water is limiting, a sunflower plant at a given growth stage displays different leaf types that do not have the same growth history, level of development and metabolism, and biophysical environment. Their functioning is linked to differential needs before and after flowering: before, leaves are demanding assimilates -sink organs- and progressively become source after they reach their full expansion and growth; leaves at the top are the last to expand while bottom leaves have already long acquired their maximal area and undergo progressive senescence processes (Aguëra et al., 2012). Mature leaves could feed upper leaf nodes by translocation process (Merrien, 1992). After flowering, leaves are all source organs and feed mostly reproductive organs (capitula, grains, Connor and Hall, 1997). Still, a gradient of mature leaves exist since uppermost leaves (upper 2/3) significantly contribute to assimilates providing and grain filling (Alkio et al., 2003). Senescence progresses acropetally (from bottom to top of the canopy) during this period. We took special care to choose leaves corresponding to those three types: ages of "developed" leaf, expressed in thermal time after

emergence, were comparable in Exp.I and Exp.II. It may be surprising that the age of a leaf considered as a "developed" one in terms of functioning, significantly differs between vegetative and reproductive stages (700 and 900 °C days respectively). However, Moschen et al. (2014) showed that leaf profile, functioning and senescence varied with developmental stage: higher nodes are initiated later than lower nodes but display longer lifespan, contributing to more assimilates available during grain filling.

In response to water stress, Pantin et al. (2012) argued that stomatal regulation appears progressively as leaves age. This could explain the higher sensitivity of stomatal regulation in post-flowering period stated by Connor and Hall (1997). In Exp.I, we observed that fully expanded leaves ("developed") were the least sensitive to water deficit in terms of transpiration regulation. We suggest that (i) young expanding leaves could be likely more sensitive to water stress as observed by Cechin et al. (2006), (ii) and that mature leaves in vegetative plants corresponded to senescing leaves which were no longer able to control water loss through stomata. If referring to Pantin et al. (2012) work, our results suggest that stomatal closure is best controlled in "developed" leaves. "Young" leaves could display only little stomatal regulation since their stomata are mostly occluded and undeveloped, thus are more sensitive to environmental conditions that do not play through stomatal regulation. It is noteworthy that sensitivity of "youngest" top leaves seems to be kept even during post-flowering period: it is often observed that senescence starts from top to down during the second half of grain filling; leaves in the intermediate part are the last to die. This trend was also observed in maize (Valentinuz and Tollenaar, 2004); if higher sensitivity is confirmed, this could be due to an interaction between lower leaf area and intensity of stress.

Photosynthesis regulation in response to water deficit was not measured in Exp.I; we could however hypothesize that photosynthesis in "young" leaves would be the first and mostly impaired, as observed in Cechin et al. (2010). Assuming that stomatal regulation appears in more grown-up leaves, photosynthesis in "young" leaves could be limited by internal carbon metabolism (Pantin et al., 2012). Senescence could be at an advanced stage in "mature" leaves; photosynthetic activity is progressively impaired because of photosystems destruction (Flexas and Medrano, 2002); that means, photosynthesis regulation is no longer under stomatal control in mature senescing leaves.

Different proportions of leaf types explain differences in pre- and post-flowering regulation of plant transpiration

The differences observed in leaves functioning in vegetative and reproductive periods were linked to the proportions of leaf types, varying with growth stage. At any plant growth stage,

leaf population could be represented by a mixture of "mature", "developed", and "young" leaves: at the start of vegetative growth stage experiments, the proportion was 30/30/30 ("mature, "developed", and "young" respectively) while it was 20/60/20 at the start of reproductive growth stage experiment. There were more "developed" leaves during that stage (leaf maximum expansion is considered to be reached at mid-flowering; Merrien, 1992). The later regulation of transpiration after flowering could be explained by the higher contribution of "developed" leaves that were less sensitive than other leaf types. This finding does not reproduce the conclusions of Connor and Hall (1997) who reported that plants stomatal conductance in reproductive growth stage was more sensitive to water stress than in vegetative period. Indeed, this lower "sensitivity" of plant transpiration rate to soil water stress was observed in our data at a daily scale, but they could be linked to differences in biomass and consequent water uses of post-flowering that could not make the two experiments comparable. When using water stress indicator (FTSW), we demonstrated that post-flowering plants display a lower sensitivity of transpiration rate to soil water deficit.

Photosynthetic regulation in vegetative and reproductive growth stages

In vegetative growth stage, we observed that leaf transpiration was regulated at higher soil water content than leaf photosynthesis. This delay was suppressed or reduced in reproductive period with inversions occurring between regulations of transpiration and photosynthesis. This suggests non-stomatal and/or stomatal limitations of photosynthesis in response to post-flowering soil water deficit. However, cv. Melody appeared to be able to maintain photosynthesis at low FTSW in comparison to the other genotypes. In order to decide whether the type of limitation was stomatal or chloroplastic, it would be interesting to measure the intercellular CO_2 concentrations in leaves. Kiani et al (2007a) showed that down-regulation of fructose-1,6 bisphosphatase could play a role in non-stomatal limitation of photosynthesis, resulting in decrease of photosynthesis under water stress. Key genes associated to leaf transpiration rate and water plant status whose expression regulation differs in sensitive and tolerant genotypes were also identified in sunflower (Rengel et al,. 2012). It is noteworthy that Exp.II permitted to monitor gene expression in leaves during water deficit in order to understand physiological basis of genetic variability of sunflower response to water deficit. The latter work helped to develop a water status plant biomarker in sunflower (Marchand et al., 2012).

The existing delay between regulation of leaf transpiration (NTL) and photosynthesis (NA) to increasing soil water deficit could now be more accurately computed in SUNFLO dynamic crop model (Casadebaig et al., 2011). We established a strong linear relationship between

a.NA and a.NTL (a.NA= 1.7 * a.NTL + 3.3; R^2 = 0.74***) in vegetative growth stage. However, care should be taken for post-flowering photosynthesis regulation; considering previous conclusions, it seems better to compute a.NA in relation to FTSWt (*e.g.* as suggested by our data, FTSWt < 0.50, NA decreases) rather than in relation to a.NTL.

Existing genotypic differences that deserve to be wider investigated

There was a genotype effect on leaf transpiration regulation in Exp.I: PSC8 was able to maintain leaf transpiration at lower soil water content than XRQ, and this was significant for "young" and "developed" leaves. Genotype effects were not significant on individual processes (NTL and NA) in Exp.II because genotypes were not contrasted enough (difference of 0.17 in FTSWt for the 3 genotypes in Exp.II while it was 0.22 in the 2 genotypes in Exp.I). Though, genotype effect played on the delay between transpiration and photosynthesis regulations to increasing soil water deficit, particularly in vegetative period. While other experiments have been carried out for evaluating transpiration to water stress during vegetative growth stage in a wider range of genotypes for SUNFLO crop model (Casadebaig et al. 2008), results obtained in reproductive growth stage are brand new and should be consolidated by further experiments with wider genotype range, though water stress monitoring of post-flowering plants remains cumbersome to implement because of their size and unsuitability to standard pots used in the greenhouse experiments. The maintenance of photosynthesis related to transpiration in post-flowering (which increases water use efficiency) could be an interesting breeding trait for obtaining water-stress tolerant and still long-photosynthetic genotypes, both contributing to increased availability of assimilates for lipogenesis. As suggested by Adiredjo et al. (2014), leaf carbon isotope discrimination should be tested to assess genotypic water use efficiency (WUE) variability. Assessing the delay between transpiration and photosynthesis regulation in high throughput experiments should help to improve WUE in sunflower. Other genotypic responses in post-flowering period should be considered, particularly when stress intensifies. Hall et al. (1989, 1990) demonstrated that contribution of assimilates stored in vegetative organs until flowering significantly increased in water-stressed conditions. In other species such as lupin and wheat (Chaves et al., 2002), differential genotypic ability to mobilize those reserves could be used for selecting tolerant plants. Linking leaf to plant transpiration ratio values confirmed that the choice of a "developed" leaf as a reference (Cechin et al., 2006; Casadebaig et al., 2008; Cechin et al., 2010) was relevant whether in vegetative or reproductive growth stages. Other leaf type (mature one) was also well-related to whole-plant transpiration, but genotypic differences were best marked at that plant leaf level. However, care should be taken since leaf and plant transpiration rates were not measured at the same time step; to confirm our

observations, measurements of leaf transpiration should be carried out over a 24 hours period and then be compared to daily transpiration at plant scale. Correlation between both indicators was lower in Exp.I ($R^2 = 0.65$) compared to that in Exp.II ($R^2 = 0.83$), probably because of the higher variability of transpiration values caused by more intense stress (higher temperatures and VPD in Exp.I).

5. Conclusions

This study aimed at analyzing transpiration and photosynthesis responses of sunflower genotypes to soil water deficit in relation to growth stage (before and after flowering) and leaf age. We demonstrated that plants regulated their transpiration rate before photosynthesis under water stress during vegetative period, while there was no significant delay between both processes in reproductive growth stage. The photosynthesis regulation parameter value could be estimated from transpiration regulation at leaf and plant levels. This could help to lighten phenotyping methods and explore genetic variability in sunflower.

Our results suggest that taking into account differential sensitivity of both processes in response to water stress depending on growth stage in SUNFLO crop model should help to better describe sunflower response for a wider range of soil water deficits.

6. Acknowledgments

This work was supported by French National Research Agency (SUNRISE project 2012-19) and Association Nationale de la Recherche et de la Technologie (CIFRE N° 2010/1467). We greatly thank Patricia NOUVET and the LIPM team for their help in greenhouse data collection.

Bilan du chapitre III

- Ces expérimentations en serre ont permis de confirmer qu'un dessèchement progressif du sol réduit tout d'abord la transpiration puis la photosynthèse, indiquant une plus forte sensibilité de la transpiration (comparativement à la photosynthèse) à la contrainte hydrique. En revanche, cette différence de sensibilité des processus à la contrainte hydrique apparaît moins marquée en phase post-floraison. Les différences de sensibilité au stress hydrique pourraient être expliquées par des différences de fonctionnement liées à l'âge des feuilles ; différents « âges » de feuilles coexistent sur une même plante à un stade donné et leur proportion détermine la réponse globale de la plante avant et après la floraison.
- Ces nouvelles connaissances liées au stade de développement seront à intégrer dans le modèle SUNFLO pour mieux représenter la diversité des réponses transpiratoires et photosynthétiques en conditions de contrainte hydrique, en particulier lorsque celle-ci s'exprime en phase post-floraison. Ces nouveaux éléments de compréhension du fonctionnement des « sources » et des « puits » en phase de post-floraison seront intégrés lors de la construction du modèle dynamique source-puits.

Chapitre IV: A source-sink based dynamic model for simulating oil and proteins accumulation in sunflower grains

Article en préparation. Version provisoire du modèle. Soumission prévue en 2015.

A source-sink based dynamic model for simulating oil and proteins accumulation in sunflower grains

Fety Nambinina Andrianasolo [a,b,d], François Brun [c], Pierre Casadebaig [b,d], Pierre Maury [b,d,1], Philippe Debaeke [b,d,1]

[a] CETIOM, Centre INRA de Toulouse, CS 52627, 31326 Castanet-Tolosan Cedex, France

[b] INRA, UMR AGIR, CS 52627, 31326 Castanet-Tolosan Cedex, France

[c] ACTA, RMT Modélisation, UMR AGIR, CS 52627, 31326 Castanet-Tolosan Cedex, France

[d] Université de Toulouse, INP, ENSAT, CS 52627, 31326 Castanet-Tolosan Cedex, France

[1] Co-advisors of the PhD thesis

Authors e-mail addresses:

fandrian@toulouse.inra.fr

francois.brun@acta.asso.fr

casadebaig@toulouse.inra.fr

maury@ensat.fr

debaeke@toulouse.inra.fr

Corresponding author: debaeke@toulouse.inra.fr

Phone number: +33561285016

Abstract

Oil and meals are the two major outputs of sunflower production. The growing use of oleic oil type and high proteins percent meals increase the competitiveness of sunflower among other oilseed crops. Though, oil and proteins content are under both genotype and environmental control that make them difficult to predict: a comprehensive framework is lacking. We proposed to build a dynamic source-sink model for predicting oil and proteins accumulation in sunflower grains. The model was based on carbon and nitrogen fluxes description between sources (stems, leaves, receptacles) and sinks (receptacles, hulls and kernels). Allocation priority rules based on thermal time windows were established according to previous dedicated field experiments. Soil nitrogen and water availability were simulated by SUNFLO model (Casadebaig et al., 2011) and used as inputs of the source-sink model. Several experiments with different genotypes, contrasted level of density, nitrogen and irrigation were conducted in South-West of France and used to build and to evaluate the model. Patterns of oil, proteins, receptacles weights per m², oil and proteins percent and LAI dynamics were well reproduced by the model, though systematic over-estimation suggested that genotypes and nitrogen differences should be explicitly formalized in the model. Mean prediction error was 6.10 and 3.90 points for oil and proteins percent respectively. Sensitivity analysis performed with the Morris screening method highlighted the parameters that needed to be better estimated: nitrogen uptake, photosynthesis, remobilization of carbon from stems and genotypic differences at sink level.

Keywords: carbon remobilization, crop model, hulls, nitrogen remobilization, nitrogen uptake, photosynthesis

1. Introduction

Oil and meals represent more than 90% of sunflower grains (Helianthus annuus L.) industrial outputs. The increasing production of oleic-types (more than 50% of French dedicated sunflower surfaces, Jouffret et al., 2011) is of interest for health (Aguirrezábal et al., 2009) and environmental matters (Pilorgé, 2010), that makes sunflower a potentially very competitive crop in oilseed market.

Sunflower oil concentration is high (around 50% out of grains dry weight), but it fluctuates according to genotype and environmental growing conditions. In a recent study, variations linked to environment or to interactions between genotype and environment were found to be higher than those related to genotype only (Grieu, 2008; Champolivier et al., 2011). This can lead to unexpected oil concentrations at harvest –*i.e.*, concentration significantly differing from the potential of a given genotype, which can in turn cause economic instability for the producer, rewarded or penalized up to 1.5 % per oil gain or loss referring to commercial standard oil concentration (44% oil, 11% impurities and water, CETIOM).

We assume that the predictability of the oil concentration can be improved by dissecting the physiological processes that lead to the elaboration of oil and grain and the effects of environmental factors on this elaboration.

Concerning meals, they are obtained from grinding seeds and extracting oil. Sunflower meals are by now limited to swine and cattle feed because of their high cellulose and low proteins concentration (29%), compared to other oilseed meals (Borredon et al., 2011). Meals with higher proteins concentration (mid-pro: 33% and high-pro: 36%) are increasingly obtained with the help of improved half-dehulling processes (Peyronnet et al., 2014) that widens its feed usability and competitivity with soybean and rapeseed meals. Hullability and proteins richness depend on genotype and environmental conditions in which the crop is grown (Denis and Vear, 1994; 1996). Both were shown to display variability with year and cropping

environments (CETIOM, 2014) that deserve to be dissected for evaluating genotype, environment and/or both interaction effect contributions to proteins content determination; the ability to predict proteins content on total grain (field scale) and consequent defatted grain (industrial scale) should be of great help for providing management tips (choice of genotype that would best be optimized in a given environment).

Physiology of oil and proteins accumulation has been widely documented in literature (Goffner et al., 1988; Connor and Hall, 1997; Aguirrezábal et al., 2003; Ruiz and Maddonni, 2006), but contradictory results were found about oil and proteins accumulation in response to management and environmental stress factors (Rizzardi et al., 1992; Diepenbrock et al., 2001; Santonoceto et al. (2002); Anastasi et al. (2010). There is still much uncertainty about how assimilates (nitrogen and carbon compounds) behave or interact after flowering.

Crop models as tools for predicting oil and proteins concentrations in sunflower

Main objectives of crops models are (i) predicting agronomic indicators of plants performance (ii) helping to understand complex processes through testing hypotheses of functioning (Boote et al., 1996). Models that simulate oil content in sunflower exist but they all are statistical models (Chapman et al., 1993; Villalobos et al. 1996; Pereyra-Irujo and Aguirrezábal, 2007; Casadebaig, 2008); proteins were not simulated. Though quality of prediction of oil of some of these models were quite satisfactory (1.4 oil points error for Pereyra-Irujo and Aguirrezábal, 2007; 3.8 for Casadebaig, 2008; Andrianasolo et al., 2014), some stressed to very stressed (drought or high fertilization and high plant density) environments could not be well predicted. We consider that given the variability of oil response to environment and genotype, a statistical model is not always adapted. Decomposition through time and space should help to unravel synergistic effects and differential stress factors effects on system components (Chapman et al., 2008). Since the idea is to construct an extrapolable crop model, we assume that quality of prediction could still be improved through a dynamic source-sink based model for oil and proteins (Léchaudel, 2005).

Source-sink studies are frequent in other species (maize: Uharte and Andrade, 1995; Borras et al., 2002; wheat: Schnyder, 1993; Borras et al., 2004; Yin et al., 2009; soybean: Andrade and Ferreiro, 1996, Borras et al., 2004 ; barley: Schnyder, 1993; Dreccer et al., 1997; Dordas, 2012; sorghum : van Oosterom et al., 2010 and sunflower: Andrade and Ferreiro, 1996; Dosio et al., 2000; Alkio et al., 2003; Aguirrezábal et al., 2003; Izquierdo et al., 2008). Consequently, source-sink models based on carbon and/or nitrogen dynamics are increasingly developed (wheat: Martre et al., 2003, Bertheloot et al., 2008; peach: Lescourret et al. (1998); mango: Léchaudel (2005), cotton: Li et al., (2009), grass: Tabourel-Tayot and Gastal (1998a, 1998b)). Those models are considered to be intermediate between empirical and complex models (Lescourret et al., 1998) and mostly display priority rules among sources and/or sinks components.

We propose to build a source-sink model operating at a daily time step with priority rules for oil and proteins elaboration in sunflower. Such model type appears adapted for this crop since (i) elements of oil and proteins physiology and determination are provided by literature (ii) dynamics of oil and proteins elaboration are known but no effort was made on crossing literature knowledge and modeling approach to formalize them through hypotheses in a simulation model yet.

The first part of this study will focus on the description of the model features, based on carbon and nitrogen dynamics (assimilation and remobilization) in the crop at square meter scale. Second part will deal with parameterization and evaluation results, completed by a sensitivity analysis of oil and proteins percent to input data.

2. Materials and methods

2.1. Model overview

The dynamic source-sink model describes nitrogen and carbon assimilations and remobilization processes from sources (stems, leaves, receptacles) to sinks (receptacles, hulls, proteins, oil) components (Hall et al., 1990; Merrien, 1992; López Pereira et al., 2008 for carbon; Steer and Hocking, 1984; Bauchot and Merrien, 1988 for nitrogen. Flows of carbon and nitrogen are governed by priority rules based whether on achievement on thermal time windows (start or end of accumulation of a component), arbitrary order in which the accumulation / remobilization of a compound is determined by the minimum between remaining source (after the use of source by a previous component) and its own "demand" pattern, or achievement of a minimal structural content for a source component that indicates the triggering of another source. Carbon and nitrogen assimilations are modeled following RUE and N absorption (Pan et al,, 2006) approaches respectively. Remobilization (or source depletion) fluxes are considered to follow a sigmoidal negative pattern and and directed exclusively to sink components. Oil and proteins in grains are computed from carbon and nitrogen contents in kernels. Processes are simulated from flowering (R5.1) to physiological maturity (R9, Schneiter and Miller, 1981).

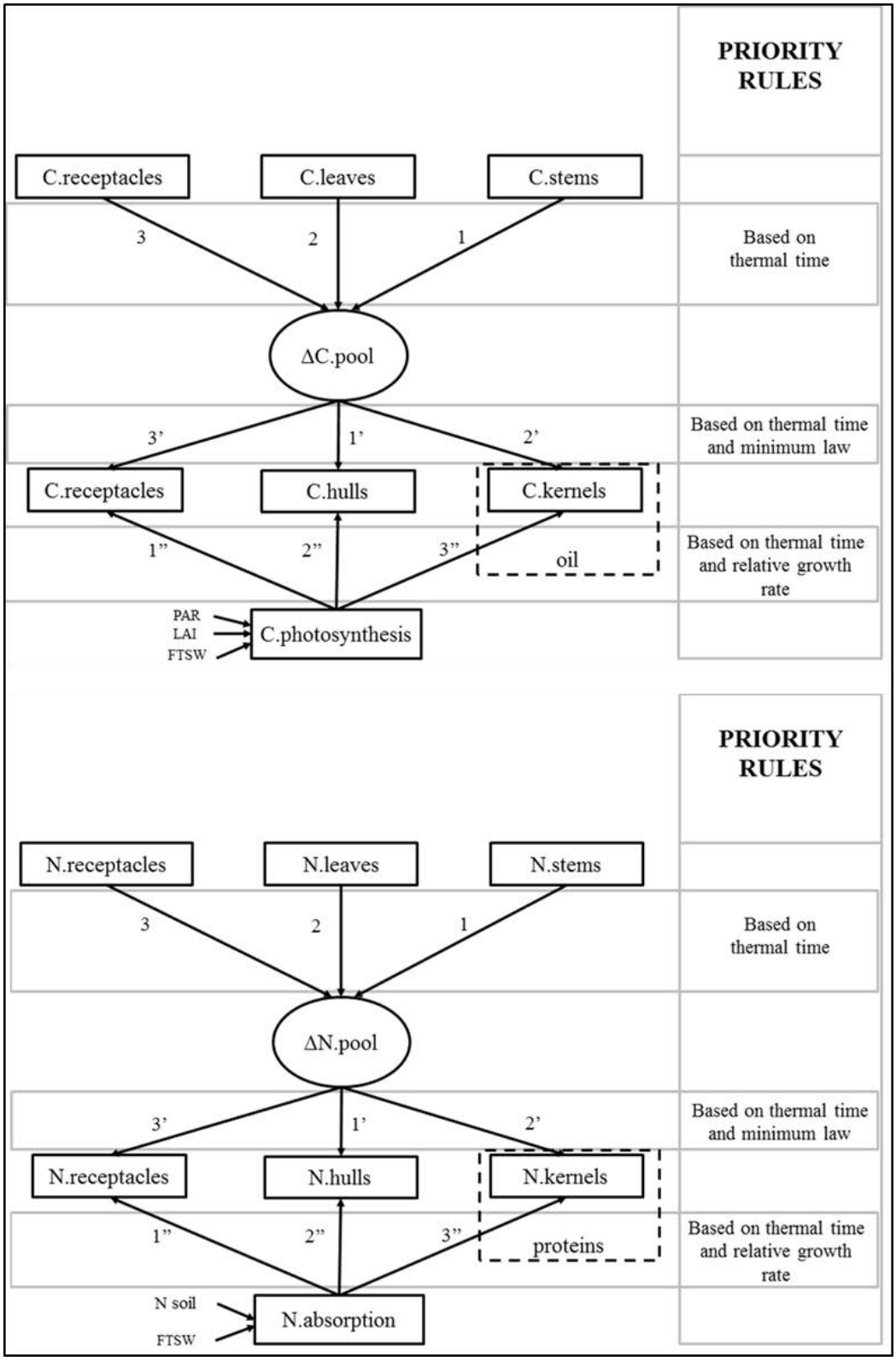


Fig.1. Source-sink carbon (C) and nitrogen (N) framework of the source-sink model. Considered sources are stems, leaves, and receptacles, and sinks are receptacles, hulls, and kernels. ΔC and ΔN.pool represent sum of C or N fluxes from receptacles, leaves and stems. Since they are transitory variables not explicitly formalized in the model, they are represented in circles in the framework. Rectangles correspond to source or sink compartments. Priority rules are indicated by figures and are defined at different levels (priority rules for source depletion (1,2,3), sink replenishing from remobilization (1', 2', 3') and photosynthetic carbon partitioning (1'', 2'', 3''). Oil and proteins weights are deduced from C and N in kernels (dotted rectangles). The model uses SUNFLO nitrogen (N.soil or N2 in the model) and water availability (FTSW) outputs. Details of all formalisms are provided in the text.

2.2. Model formalisms description

2.2.1. Computation of cumulative thermal time

All processes considered in the model are assumed to be depending on cumulative thermal time after flowering (Fig.1), whether directly (carbon assimilation and remobilization, nitrogen remobilization, sink orders of priority for C and N) or indirectly (nitrogen uptake). For the rest of the manuscript, thermal time after flowering will be noted "TT". Since the model is daily, [j] indicates the daily value of one or another variable. The daily variation of TT is noted $\Delta TT[j]$ and calculated as following:

$$\Delta TT[j] = \max\left(\frac{Tmax[j] + Tmin[j]}{2} - Tb, 0\right) \quad eq(1)$$

where $Tmax$ and $Tmin$ are daily maximum and minimum air temperatures respectively. Tb corresponds to temperature base for sunflower (CETIOM). Definitions and values of parameters are provided in Table.2.

2.2.2. Patterns of nitrogen (N) and carbon (C) remobilization

We assumed that all potential reserve organs – stems, leaves and receptacles- contribute to sink carbon and nitrogen supply. All reserve organs are depleted following a decreasing sigmoidal curve such that:

$$X.source[j] = X.source[j-1] + \Delta \mathrm{X.source}[j] \times \Delta TT[j] \quad eq(2)$$

where X corresponds whether to C or N, and source to stems, leaves or receptacles. $\Delta \mathrm{X.source}[j]$ is the daily variation of $X.source$ corresponding to the following derivative:

$$X.source[j] = -\frac{\left(a \times \left(\left(\frac{TT[j]}{c}^{(b-1)}\right) \times b \times \left(\frac{1}{c}\right)\right)\right)}{1 + \left(\frac{TT[j]}{c}^{b}\right)^2} \quad eq(3)$$

Where a = *X.source.max*, b = *X.source.rate* and c = *TT.X.source.half* and correspond to parameters of a sigmoid curve (see Table.2).

We established priority rules at reserve organs level following Hocking and Steer (1983) observation that stems contributed more than leaves to seed dry matter after flowering. We

hypothesized that reserve organs are depleted at different thermal times after flowering or only after a previous source organ reaches its structural content – C or N limit. That is:

$\Delta X.stems[j]$ starts at $TT[j] = 0$, with X corresponding to C and N; $\Delta N.leaves[j]$ starts at $TT[j] = 0$; $\Delta C.leaves[j]$ starts at $TT[j] \geq$ *photosynthesis.TT.threshold* (600 °C days).; $\Delta C.receptacles[j]$ starts when $TT[j] \geq$ *photosynthesis.TT.threshold* or when $C.leaves[j] \leq$ *C.leaves.limit* for carbon; $\Delta N.receptacles[j]$ starts when $TT[j] \geq t1.cap$ (415°C days) or when $N.leaves[j] \leq$ *N.leaves.limit* for nitrogen.

Since receptacles behave first as sink then source, dynamics of C and N in receptacles will be described in the following section.

2.2.3. Assimilation of carbon

We chose to model the acquisition of photosynthetic carbon with Monteith's (1977) radiation use efficiency (RUE, g $MJ^{-1}m^{-2}$ PAR) approach:

$$DM[j] = RUE[j] \times 1 - \exp(-k \times LAI[j]) \times PAR[j] \quad eq(4)$$

where DM is the dry matter produced from photosynthesis (g m^{-2}) and PAR the photosynthetically active radiation (MJ m^{-2}). RUE increases with rate of leaf photosynthesis (Sinclair and Horie, 1989) and decreases with leaf age and ontogeny (second part of grain filling, Hall et al., 1995). We established a plateau-linear relationship between potential RUE (named $RUEpot$) and thermal time from experimental data such that:

$$RUEpot[j] = b.RUEpot + a.RUEpot \times TT[j], for\ TT[j] > t1.RUEpot$$

$$RUEpot[j] = a.RUEpot \times t1.RUEpot + b.RUEpot, for\ TT[j] \leq t1.RUEpot \quad eq(5)$$

$PAR[j]$ is computed from daily meteorological data as $PAR[j] = 0.48 \times RG$, where RG is global radiation (MJ m^{-2}). Calculation of $LAI[j]$ will be provided below (2.2.7).

We define the daily photosynthetic carbon as:

$$C.photo[j] = DM[j] \times C.percent \quad eq(6)$$

All carbon and nitrogen content are expressed as percentage of dry matter and source and sink components are determined at flowering and originate from experimental data.

2.2.4. Remobilized carbon and nitrogen priority rules towards sinks

Remobilized carbon and nitrogen are considered to be dedicated exclusively to grains (hulls and kernels). Similarly to priority rules established at source level, we also established priority rules at sink level. Considering that $X.pool[j]$ is the sum of daily stored carbon or nitrogen available for sink filling:

$$X.pool[j] = X.stems[j] + X.leaves[j] + X.receptacles[j] \quad eq(7)$$

$X.receptacles[j]$ will be included in the $X.pool[j]$ as soon as it starts sourcing, we assume that fluxes of C (or N) toward sink organs equal the fluxes originating from the 3 source organs in a priority order, such that the 2nd prior filled sink is depleted from the flux originating from the first prior filled sink:

$$\Delta X.hulls[j] = \min\left(X.pool[j], -(\Delta X.stems[j] + \Delta X.leaves[j]\Delta X.receptacles[j])\right) eq(8)$$

$$\Delta X.kernels[j] = \min\left(X.pool[j] - \Delta X.hulls[j], -(\Delta X.stems[j] + \Delta X.leaves[j]\Delta X.receptacles[j])\right) eq(9)$$

Knowing that components of grains (hulls, kernels) accumulate at different well established periods (Connor and Hall, 1997; Ruiz and Maddonni, 2006), we defined periods at which $\Delta C.hulls$ and $\Delta C.kernels$ were triggered. We used parameters values established from a previous study (Andrianasolo et al., 2014 for *t0.oil_0*; *t1.oil*; chapter II). We considered an additional parameter for oil accumulation (Aguirrezábal et al., 2003, *t0.oil_1* = 250°C days after flowering) such that oil accumulates slowly between *t0.oil_0* and *t0.oil_1* and exponentially from *t0.oil_1* to *t1.oil*. For $\Delta N.hulls$ and $\Delta N.kernels$, we considered *t0.proteins* = 7.29 and *t1.proteins* = 509 (Andrianasolo et al., 2014, chapter.II).

Dry weights of sinks and sources organs (g m^{-2}) are computed by dividing their carbon content (ex: $C.stems$, $C.hulls$) by carbon percent parameter.

2.2.5. Nitrogen absorption

Little information is available for describing nitrogen absorption dynamics in sunflower, particularly after flowering. Though, considering the sigmoid pattern of kinetics of nitrogen described by Merrien (1992), it is plausible that nitrogen is accumulated proportionally to sink demand (Jamieson and Semenov, 2000; Pan et al., 2006). We proposed to use Pan et al. (2006) formalisms that we slightly modified to adapt to sunflower: we had previously described receptacles, hulls and kernels weights growths (sink demand) so we defined nitrogen absorption per sink demand. N absorption priority rule is the same as C sink organs:

$$N.X.abs[j] = \min(N2[j],\ D \times E \times F) \quad eq(10),\ \text{with}$$

$$D = d \times N.abs2pot \times grainfilling\ duration \quad eq(11)$$

$$E = 1 - \cos(1 - nk \times (N1.F1 - pool.N.threshold)^2 \times \frac{pi}{5} \quad eq.(12)$$

$$F = \exp(N.uptake.sink.growth \times X[j]) \quad eq.(13)$$

$N.X.abs[j]$ corresponds to nitrogen absorption for each sink organ X (receptacles, hulls and kernels) and d is the parameter describing the growth rate of each sink organ (*receptacles.growth.rate, hulls.growth.rate, kernels.growth.rate*)

$grainfilling\ duration$ was expressed as the number of days between start of flowering and physiological maturity.

For $0 \leq TT[j] \leq$ *t0.oil_0*, *receptacles.growth.rate* equalled *cap.growth.1*; *hulls.growth.rate* was *hull.growth.1*. From *t0.oil_0* $\leq TT[j] \leq$ *t0.oil_1*, receptacles, hulls and kernels growth rates were *cap.growth.1* and *hullKernel.growth.2*. Until *N.uptake.TT* (300°C days), those values corresponded to *cap.growth.2*, *hull.growth.3* and *hullKernel.growth.2*.

$N2[j]$ is the available soil nitrogen pool (g m^{-2}) simulated by SUNFLO model (see 2.2.8). *N1* is the vegetative nitrogen weight at flowering. All parameters are similar to Pan et al. (2006) except *pool.N.threshold* which we established at 15 g m^{-2} (experimental data and Massignam

et al., 2009) and *N.abs2.pot* (potential daily N absorption after flowering) set at 0.035 g m^{-2} day^{-1}. We assumed that nitrogen uptake ceased at *N.uptake.TT* (300°C days after flowering).

2.2.6. Priority order among sinks

For each sink carbon organ (including receptacles), we proposed that daily assimilation (of carbon and nitrogen) is partitioned among sink organs. For this, we computed growth rates of each sink from experimental data and computed the proportion of each sink at similar thermal times:

For carbon:

$$C.receptacles[j] = C.receptacles[j-1] + \Delta C.receptacles[j] + d \times C.photo[j]\ eq(14)$$

$$C.hulls[j] = C.hulls[j-1] + \Delta C.hulls[j] + d \times C.photo[j]\ eq(15)$$

$$C.kernels[j] = C.kernels[j-1] + \Delta C.kernels[j] + d \times C.photo[j]\ \ eq(16)$$

For nitrogen:

$$N.receptacles[j] = N.receptacles[j-1] + \Delta N.receptacles[j] + N.receptacles.abs[j]\ eq\ (17)$$

$$N.hulls[j] = N.hulls[j-1] + \Delta N.hulls[j] + N.hulls.abs[j] eq\ (18)$$

$$N.kernels[j] = N.kernels[j-1] + \Delta N.kernels[j] + N.kernels.abs[j]\ eq(19)$$

Nitrogen and carbon accumulation in sinks are regulated with thermal time rules; for carbon, from $0 \leq TT[j] \leq$ *t0.oil_1*, *receptacles.growth.rate* is *cap.growth.1*, while it is equal to *cap.growth.2* from *t0.oil_1* to *t1.hull* and *cap.growth.3* until *t1.cap*. *hulls.growth.rate* corresponds to *hull.growth.1* parameter when $TT[j]$ is lower than *t0.oil_0*, while it is *hullKernel.growth.2* parameter from *t0.oil_0* to *t1.hull* and *hull.growth.3* for $TT[j]$ between *t1.hull* and *t1.cap*. Finally for kernels, *kernels.growth.rate* was *Kernel.growth.1* parameter from *t0.oil_0* $\leq TT[j] \leq$ *t0.oil_1*, while it was *hullKernel.growth2* until $TT[j]$ reaches *t1.hull* and *Kernel.growth.3* until reaching *t1.cap*. After *t1.cap*, we consider that $C.Kernel[j]$ is filled 100% of $C.photo[j]$ until reaching *photosynthesis.TT.threshold* (600°C days).

Receptacles, hulls and kernels weights are computed daily from their carbon percentage. Oil is obtained by multiplying $C.Kernel$ and $Kernel.oil.percent$ (0.63).

2.2.7. Senescence at the crossroad of C and N dynamics

We assumed that LAI dynamics followed a two-step pattern: from flowering, LAI is deduced from N dynamics in leaves such that:

$$LAI[j]pot = \frac{N.leaves[j]}{N.leaves.percent} \times \frac{SLAfv}{10^3} eq\ (20)$$

Where *SLAfv* is specific leaf area of green leaves ($m^2\ kg^{-1}$).When reaching a N% limit (*N.leaves.limit2*) in leaves under which senescence is accelerated (de la Vega et al., 2011) or a physiological thermal time where senescence inevitably occurs (*LAI.TT.threshold*, experimental data), LAI dynamics is governed by thermal time in a linear pattern:

$$LAI[j]pot = \frac{N.leaves[j]}{N.leaves.percent} \times \frac{SLAfv}{10^3} + a2.LAI \times (TT[j] - LAI.TT.threshold)\ eq(21)$$

2.2.8. Taking into account stress factors from SUNFLO simulations

2.2.8.1. Nitrogen availability and stress (and impact on N absorption)

As stated above, soil nitrogen availability was simulated by SUNFLO model (Casadebaig et al., 2008; Casadebaig et al., 2011) due to lack of measurements in our datasets. Available nitrogen was calculated from sowing to harvest by taking into account nitrogen mineralization, fertilization, leaching and denitrification. For further information, please refer to Casadebaig et al. (2011).

Nitrogen stress effect was represented by nitrogen nutrition index (NNI) at flowering, which was generally available in all datasets. NNI at flowering (*NNI.F1*) was used for computing actual $LAI[j]$ ($LAI[j]act$) in the second part of LAI dynamics so that:

$$LAI[j]act = \frac{N.leaves[j]}{N.leaves.percent} \times \frac{SLAfv}{10^3} + a2.LAI \times NNI.F1 \times (TT[j] - LAI.TT.threshold)\ eq(22)$$

Nitrogen percent in plant was also computed from the relationship between nitrogen percent and NNI at flowering (experimental data), following the equation (Fig.2):

$$Nitrogen\ percent = N.percent.NNI\_0 \times \frac{\sin(NNI.F1^2)}{NNI.F1})/100 \quad eq(23) \qquad (R^2 = 0.93)$$

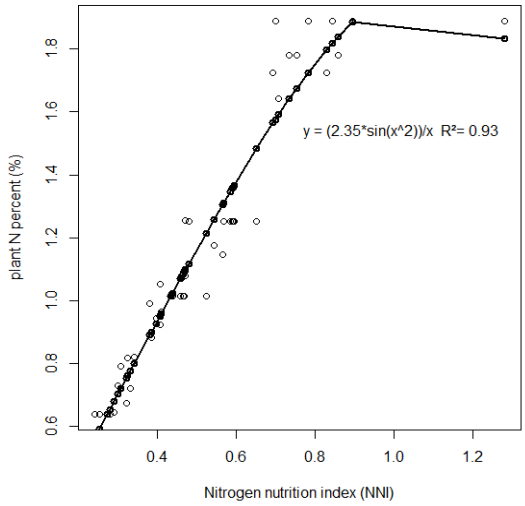


Fig.2. Relationship between nitrogen nutrition index (NNI) and plant nitrogen percent (%) at flowering. Data originate from 2012 Auzeville experiment (parameterization dataset).

Since we had collected information on nitrogen percent on each source compartment (leaves, receptacles and stems), we applied a coefficient corresponding to the nitrogen ratio in the source compartment and the nitrogen ratio in plant:

$$N.X.ratio = N.percent.NNI\_0 \times \frac{sin(NNI.F1^2)}{NNI.F1} * N.percent.NNI\_X \times \frac{1}{100} eq(24)$$

where X corresponds whether to stems, receptacles or leaves, and $N.percent.NNI\_X$ parameter value corresponding to $N.percent.NNI\_stems$, $N.percent.NNI\_cap$ or $N.percent.NNI\_leaves$.

Finally, when not available, we computed stems weight at flowering from vegetative dry weight and the relationship we found between stem percentage and NNI (Fig.3):

$$stems\ ratio = DM \times (stems\_INN\_1 - stems\_INN\_2 \times NNI.F1\ eq(25) \qquad (R^2=0.80)$$

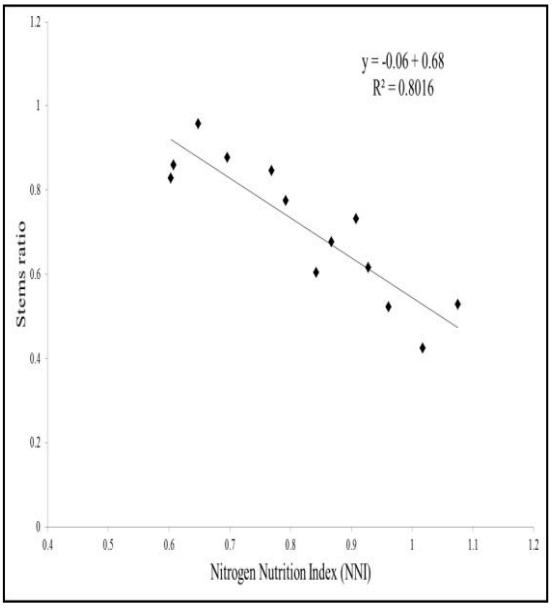


Fig.3. Relationship between nitrogen nutrition index (NNI) and stems percent equivalent at flowering. Data originate from 2012 Auzeville experiment (parameterization dataset).

2.2.8.2. Water availability and stress (and impact on N absorption and RUE)

For taking into account water availability, we also simulated fraction of transpirable soil water (FTSW[j]) with SUNFLO model. We then used a daily water stress factor (W.RUE[j]) on photosynthesis (Casadebaig et al., 2011) and nitrogen absorption (Morot-Gaudry, 1997) such that:

$$W.RUE[j] = -1 + \frac{1}{(1 + 4.5 \times \exp(W.RUE.parameter \times FTSW[j]))} \quad eq(26)$$

W.RUE.parameter was obtained from greenhouse experiments and corresponds to the threshold at which sunflower genotypes regulate their photosynthetic activity in response to water stress.

$$C.photo[j]act = C.photo[j] \times W.RUE[j] \; eq(27)$$

For each *N.X.abs* (*X* being receptacles, hulls and kernels), we multiplied its value by W.RUE[j]:

$$N.X.abs[j]act = N.X.abs[j] \times W.RUE[j] \; eq(28)$$

2.3. Inputs and outputs

2.3.1. Inputs

Model main input variables are SUNFLO simulations (FTSW and N2), climate, and some information on crop state at flowering: vegetative dry matter (DM), NNI, stems weight, LAI, vegetative nitrogen weight.

2.3.2. Outputs

Model final output variables are oil, hull, proteins and seed weights per m² and respective component percentages. Intermediate simulated variables are all sources dynamics (LAI, stems and receptacles weights, carbon and nitrogen assimilations and remobilizations).

2.4. Calibration dataset and parameters

2.4.1. 2012 experiment description

The 2012 experiment at Auzeville was used for calibrating our model (Table.1). The design consisted of a split-split plot with two genotypes (*cv.*Kerbel, *cv.*Olledy) and two plant densities (3 and 4.5 plants per m²) at the plot level, submitted to 2 contrasting nitrogen fertilization treatments (N-: unfertilized and N+: 150 kg ha^{-1}). Plants were irrigated after flowering to avoid water stress. For further experiment description, please refer to Andrianasolo et al., 2014 (chapter II).

2.4.2. Measurements and simulations from SUNFLO

In this experiment, sources (leaves, stems, and receptacles) and sinks (receptacles, hulls and kernels) as well as oil and proteins weights and percentages were monitored from flowering (R5.1, Schneiter and Miller, 1981) till physiological maturity (R9). Leaves and receptacles and all sink values were obtained weekly for 8 weeks while stems were measured only at 3 harvest dates between flowering and physiological maturity. Grains and kernels were analysed for their oil and N concentrations by NMR (Minispec NMS 110, Bruker,

Analytische Mestechnik, Rhinsteten, Germany) and Dumas method (Flash 2000, Thermo Scientific, inc., US) respectively. N concentrations were multiplied by 6.25 to obtain proteins concentrations. In all experiments, carbon and nitrogen percent correspond to total carbon and total nitrogen percent as measured with Dumas method (EURO EA-3000, Eurovector Spa, Milan).

2.4.3. Parameters and origin

Most parameters values were determined in 2012 experiment. Timings parameters resulted from bi-linear regression analysis (Andrianasolo et al., 2014; chapter II). Table.2 provides parameters definitions and origin.

2.5. Evaluation dataset

4 datasets were used for evaluating our model: a set from an experiment in 2011 at Auzeville, the second originating from Auzeville (rainfed) 2012 experiment, field experiments in Agen and Le Magneraud locations in 2013.

2.5.1. Auzeville irrigated experiment

We conducted a field experiment at Auzeville in 2011, that consisted of a split-split plot design with water treatment at largest scale (irrigated vs rainfed), nitrogen treatment at middle scale (N- and N+) and genotype by plant density factors at plot scale (2 genotypes: *cv.*Kerbel and *cv.*LG5451_HO and 2 plant densities: 3 and 4.5 plants per m² respectively) (Table.1). For further information on crop management, please refer to Andrianasolo et al., 2014, chapter II). Similar to 2012 experiment, all sink and source variables were measured weekly during 8 weeks. Stems data were not available.

2.5.2. Auzeville rainfed experiment

We used the non-irrigated part of 2012 Auzeville experimental design for model evaluation. Number of genotypes, nitrogen and density conditions are the same as the parameterization dataset, as well as the availability of source and sink variables and their dynamics.

2.5.3. Agen experiment

A first experiment was conducted in Agen (Department 47) in 2013 by CETIOM institute. 3 genotypes (*cv*.Kerbel, *cv*.LG5451_HO and *cv*.Melody) were organized in randomized complete blocks with 4 replicates. They were sown on April, 19th on a deep clay soil with 182 mm of available soil water content. Nitrogen residual was 96 kg ha^{-1}. Trial was not fertilized but irrigated around flowering (July, 11th). Mean plant density was 7.7 plants per m². LAI and sink data (seed, oil, proteins, kernels, hulls) were monitored at 6 and 3 occasions between flowering and physiological maturity on 2 blocks, respectively. Harvest date was on September, 10th.

2.5.4. Le Magneraud experiment

A second experiment was carried out in Le Magneraud (Department 17) in 2013 by CETIOM institute. It consisted in a split-plot design with 3 genotypes (*cv*.Kerbel, *cv*.LG5451_HO and *cv*.Melody) and 4 replicates. 2 water conditions were distinguished: rainfed and irrigated (3 applications of 34 mm on July, 11th and 24th and August, 8th). Plants were sown on April, 24th. Mean plant density at emergence was 5.4 plants per m². Plots were fertilized with duck manure (5 t ha^{-1}) equivalent to 114 kg ha^{-1} mineral nitrogen. Soil was a shallow silty clay type with 80 mm of available water content. N residuals amounted 74 and 57 kg ha^{-1} in rainfed and irrigated plots respectively. LAI and sink data were monitored on 3 occasions between flowering and maturity on 2 blocks. Harvest date was on September, 24th.

2.6. Parameterization and evaluation tools

For goodness of fit evaluation, we computed root mean squared error (RMSE) and model efficiency (EF) tools (Wallach et al., 2013). They were calculated on whole dynamic as well as on thermal time corresponding to maturity for each situation. We also performed an ANOVA on RMSE values for detecting main effects on RMSE variability for each simulated variable.

For evaluation dataset, we propose to compute RMSEP, EF, and bias (simulated minus observed)

2.7. Sensitivity analysis

We performed a global sensitivity analysis following Morris method (Morris, 1991) improved by Campolongo et al. (2007). It consists of varying one factor at a time and establishing two sensitivity measures: μ^* and σ. μ^* indicates the influence of the parameter on the output while σ (standard deviation) estimates the factor's higher order effects (non-linear effects or interactions with other parameters). We assumed the distributions of all parameters are independent and uniform with). Minimum and maximum values of each parameter defined from literature, expertise of experimental results. For many parameters, ranges are not well documented, thus, we defined ranges as ± 25% of the default values. Then we selected some output variables of interest (oil and proteins weights, oil and proteins percent) at thermal times corresponding to three stages of the dynamics (250, 500 and 700 °C days after flowering). For each situation, Morris method was run with 6 levels by parameter, with a grid jump of 3 and 500 repetitions. (Resulting in 500*(66+1) = 33500 sets of parameters). We compared 6 situations: 2 situations in 2011 (*cv*.Kerbel N+ D1, rainfed and irrigated) and 4 situations in 2013 (Le Magneraud: *cv*.Kerbel and *cv*.LG5451_HO, rainfed and irrigated).

To summarize the information obtained from the different situations and to rank parameters according to their sensitiveness in all situations, we computed means and standard deviations of μ* among sites for each parameter. Parameters were then categorized into 3 categories: 0 (mean 0 for all situations), "low" (situated lower or equal to 25% of max (mean (μ*)), and "high" (higher than 25% of max (mean (μ*)) (Annexe). Only "high" sensitive parameters are represented in following section.

All model construction, parameterization, evaluation and sensitivity analysis were performed with R software (R Core Team, 2014) using sensitivity package (Pujol et al., 2014) and ZeBook package (Brun et al., 2013).version 3.0.2).

Table.1. Summary table of experimental datasets (France) used for model parameterization and evaluation. AWC is the available soil water content as measured or estimated by expertise. Meteorological data were computed on the crop cycle duration. One USM corresponds to a combination of 1 genotype x 1 management x 1 replicate condition in 1 site-year.

year	site	department	sowing date	harvest date	flowering date	maturity date	Genotype number	Fertilization (kg ha^{-1})	Irrigation (mm)	AWC (mm)	mean PAR (MJ m^{-2})	mean temp. (°C)	precipitations (mm)	number of USM
parameterization dataset														
2012	Auzeville	31	6-April	11-September		05-September	2	0	54	225	1016	22.5	120	12
2012	Auzeville	31	6-April	11-September		05-September	2	150	54	225	1016	22.5	120	12
evaluation dataset														
2011	Auzeville	31	8-April	30-August	29-June	29-August	2	0	71	225	986	21.4	108	6
2011	Auzeville	31	8-April	30-August	29-June	29-August	2	150	71	225	986	21.4	108	6
2011	Auzeville	31	8-April	30-August	29-June	29-August	2	0	0	225	986	21.4	108	6
2011	Auzeville	31	8-April	30-August	29-June	29-August	2	150	0	225	986	21.4	108	6
2012	Auzeville	31	6-April	11-September	27-June	05-September	2	0	0	225	1016	22.5	120	4
2012	Auzeville	31	6-April	11-September	27-June	05-September	2	150	0	225	1016	22.5	120	4
2013	Agen	47	19-April	24-September	18-Jluly	10-September	3	0	33	182	1026	22.3	49	6
2013	Le Magneraud	17	24-April	24-September	15-July	06-September	3	114	0	80	1041	21	60	6
2013	Le Magneraud	17	24-April	24-September	15-July	06-September	3	114	102	80	1041	21	60	6

Table.2. Parameters used in the source-sink model. Names, descriptions, units and default values origins are provided.

number	category	name of parameters	defaut value	description	unit	origin
1	initialisation	Tbase	6	Baseline temperature	°C	CETIOM
2		C.percent	0.4	total carbon ratio equivalent in sink and source organs at flowering	-	experimental results
3		N.hull.percent	0.008	total nitrogen ratio equivalent in hull at flowering	-	experimental results
4		N.Kernel.percent	0.03	total nitrogen ratio equivalent in Kernel at flowering	-	experimental results
5		N.percent.INN_0	2.35	constant used for computing N.percent from NNI in all source organs	-	experimental results
6		N.percent.INN_stems	0.3	constant used for computing N.percent from NNI in stems	-	experimental results
7		N.percent.INN_cap	1.2	proportion constant used for computing N.percent from NNI in cap	-	experimental results
8		N.percent.INN_leaves	1.5	proportion constant used for computing N.percent from NNI in leaves	-	experimental results
9		stems_INN_1	0.68	proportion constant used for computing stems weight from NNI	-	experimental results
10		stems_INN_2	0.06	constant used for computing stems weight from NNI	-	experimental results
11	N.uptake	N.uptake.TT	509	thermal time threshold after which N.uptake ends	°C days	experimental results
12		N.abs2pot	0.035	potential N uptake rate after flowering	g m^{-2} day^{-1}	Pan et al.(2006) and experimental results
13		nk	0.0049	constant used in relating nitrogen absorbed before and after flowering	g m^{-2}	Pan et al. (2006)
14		pool.N.threshold	15	constant used as a threshold value limiting nitrogen absorption after flowering depending on absorption before flowering	g m^{-2}	Pan et al.(2006) and experimental results
15		N.uptake.sink.growth	-0.0012	constant used in the relationship between nitrogen absorption and sink (grain) growth after flowering	-	Pan et al. (2006)
16	photosynthesis	k	0.74	coefficient extinction	-	Debaeke et al. (2010)
17		b.RUEpot	3.54	value of intercept in the relationship between RUE and thermal time after flowering	MJ PAR m^{-2}	experimental results
18		a.RUEpot	-0.00323	value of slope in the relationship between RUE and thermal time after flowering	MJ PAR m^{-2} °C day^{-1}	experimental results
19		t1.RUEpot	272.7	value of thermal time threshold in the relationship between RUE and thermal time after flowering	°C days	experimental results
20		photosynthesis.TT.threshold	600	thermal time threshold after which photosynthesis ends	°C days	experimental results
21		W.RUE.parameter	-20	genotypic parameter of photosynthesis regulation under water stress	-	experimental results
22	senescence	N.leaves.limit2.percent	0.02	threshold value of leaves N percent at which leaves senescence depends on thermal time	-	experimental results
23		LAI.TT.threshold	483.41	threshold value of thermal time at which leaves senescence decreases linearly with thermal time	°C days	experimental results
24		a2.LAI	-0.003	slope of leaves senescence after reaching whether N percent limit2 or LAI thermal time threshold	m^2 m^{-2} °C day^{-1}	experimental results
25		SLAfv	14	specific leaf area of green leaves	m^2 kg^{-1}	experimental results
26	sigmoid curve	C.stems.max	145	constant in the negative sigmoide curve of C dynamics in stems	g m^{-2}	experimental results
27		C.cap.max	54	constant in the negative sigmoide curve of C dynamics in receptacles	g m^{-2}	experimental results
28		C.leaves.max	66	constant in the negative sigmoide curve of C dynamics in leaves	g m^{-2}	experimental results
29		N.stems.max	1.45	constant in the negative sigmoide curve of N dynamics in stems	g m^{-2}	experimental results
30		N.cap.max	2.7	constant in the negative sigmoide curve of N dynamics in receptacles	g m^{-2}	experimental results
31		N.leaves.max	4.15	constant in the negative sigmoide curve of N dynamics in leaves	g m^{-2}	experimental results
32		TT.C.stems.half	1200	thermal time constant corresponding to inflexion point in the negative sigmoide curve of C dynamics in stems	°C days	experimental results
33		TT.C.cap.half	1200	thermal time constant corresponding to inflexion point in the negative sigmoide curve of C dynamics in receptacles	°C days	experimental results
34		TT.C.leaves.half	1200	thermal time constant corresponding to inflexion point in the negative sigmoide curve of C dynamics in leaves	°C days	experimental results
35		TT.N.stems.half	500	thermal time constant corresponding to inflexion point in the negative sigmoide curve of N dynamics in stems	°C days	experimental results
36		TT.N.cap.half	500	thermal time constant corresponding to inflexion point in the negative sigmoide curve of N dynamics in receptacles	°C days	experimental results
37		TT.N.leaves.half	500	thermal time constant corresponding to inflexion point in the negative sigmoide curve of N dynamics in leaves	°C days	experimental results
38		C.stems.rate	2.5	rate of decrease at inflexion point in the negative sigmoide curve of C dynamics in stems	g m^{-2} °C day^{-1}	experimental results
39		C.leaves.rate	1.5	rate of decrease at inflexion point in the negative sigmoide curve of C dynamics in leaves	g m^{-2} °C day^{-2}	experimental results
40		C.cap.rate	2.5	rate of decrease at inflexion point in the negative sigmoide curve of C dynamics in receptacles	g m^{-2} °C day^{-3}	experimental results

number	category	name of parameters	defaut value	description	unit	origin
41	sigmoid curve	N.stems.rate	15	rate of decrease at inflexion point in the negative sigmoide curve of N dynamics in stems	g m^{-2} °C day^{-4}	experimental results
42		N.leaves.rate	0.3	rate of decrease at inflexion point in the negative sigmoide curve of N dynamics in leaves	g m^{-2} °C day^{-5}	experimental results
43		N.cap.rate	15	rate of decrease at inflexion point in the negative sigmoide curve of N dynamics in receptacles	g m^{-2} °C day^{-6}	experimental results
44		C.stems.limit.percent	0.39	minimum value of C percent equivalent in stems at harvest	-	experimental results
45		C.leaves.limit.percent	0.3	minimum value of C percent equivalent in leaves at harvest	-	experimental results
46		C.cap.limit.percent	0.32	minimum value of C percent equivalent in receptacles at harvest	-	experimental results
47		N.leaves.limit.percent	0.0036	structural value of N percent equivalent in leaves at physiological maturity	-	experimental results
48		N.stems.limit.percent	0.0014	structural value of N percent equivalent in stems at physiological maturity	-	experimental results
49		N.cap.limit.percent	0.0046	structural value of N percent equivalent in receptacles at physiological maturity	-	experimental results
50	sink	coef.proteins	6.25	coefficient to be multiplied to nitrogen concentration to obtain protein concentration	-	CETIOM
51		t0.oil_0	85.34	thermal time at which oil starts to slowly increase in grains	°C days	Andrianasolo et al. (2014)
52		t0.oil_1	250	thermal time at which oil exponentially increases in grains	°C days	Aguirrezabal et al. (2003)
53		t1.oil	743.16	thermal time at which oil stops accumulating in grains	°C days	Andrianasolo et al. (2014)
54		t1.hull	405	thermal time at which hulls stop growing	°C days	Andrianasolo et al. (2014)
55		t1.cap	415	thermal time at which receptacles stop growing	°C days	Andrianasolo et al. (2014)
56		cap.growth.1	0.67	percentage equivalent of photosynthetic carbon receptacles use from flowering to 250 degree days after flowernig	°C days	experimental results
57		cap.growth.2	0.3	percentage equivalent of photosynthetic carbon receptacles use from 250 to 405 degree days after flowering	-	experimental results
58		cap.growth.3	0.35	percentage equivalent of photosynthetic carbon receptacles use from 405 to 415 degree days after flowering	-	experimental results
59		Kernel.oil.percent	0.63	percentage equivalent of oil in kernel	-	experimental results
60		t0.proteins	7.29	thermal time start of proteins accumulation in grains	°C days	Andrianasolo et al. (2014)
61		t1.proteins	509	thermal time end of proteins accumulation in grains	°C days	Andrianasolo et al. (2014)
62		hull.growth.1	0.33	percentage equivalent of photosynthetic carbon hulls use from flowering to t0.oil_0	-	experimental results
63		hullKernel.growth.2	0.165	ratio of photosynthetic carbon hulls and kernels use from t0.oil_0 and t0_oil.1	-	experimental results
64		hull.growth.3	0.15	ratio of photosynthetic carbon hulls use from t0.oil_1 and t1.hull	-	experimental results
65		Kernel.growth.1	0.55	ratio of photosynthetic carbon kernels use from t0.oil_1 to t1.hull	-	experimental results
66		Kernel.growth.3	0.65	ratio of photosynthetic carbon kernels use from t1.hull to t1.cap	-	experimental results

3. Results

Parameterization results will be presented first, followed by evaluation and sensitivity analysis.

3.1. Parameterization results

We plotted dynamic simulated and observed data from parameterization, for receptacles, stems, oil, proteins, hulls weights per m², and oil, hulls and proteins percent (Fig.3; Table.3). Receptacles patterns were well reproduced, though values were over-estimated (from 34 to 87 g m^{-2} error). LAI patterns and range were also well simulated except in N+ D2 (both genotypes) where error was higher than in other conditions (0.90 vs 0.45). Simulated oil and hulls weights were very close to observed ones; only oil in N+ D1 was highly overestimated (20 oil points higher than other situations). Proteins weights were globally over-estimated (from 11 to 30 g m^{-2}). Hulls percent fitted very well to observed data. Simulated stems dynamics did not correspond to observed stems patterns and were only little depleted compared to real data. Model efficiency was very good (>0.90) for all outputs except for LAI where it was lower (EF= 0.28 to 0.63) and intermediate for proteins percent (from 0.62 to 0.91). Simulated oil percent fitted quite well to observed data (RMSE = 6.06 for whole dynamics and 4.4 for harvest date). Proteins percent fitted well the dynamics except in the plateau phase in N- (RMSE= 5.71 vs 5.09 for harvest date).

We performed an ANOVA on RMSE values of those outputs for detecting genotype, nitrogen, density, or any interaction effects explaining the variability of RMSE values (Table.4). Significant nitrogen (**), plant density (*) and nitrogen by density (*) effects were found on RMSE of LAI: error was lower in D1 (0.35) compared to D2 (0.60), and higher in N+ (0.63) compared to N- (0.33). There was little plant density effect (.) on stems and receptacles RMSEs, with error tending to be lower in D1 compared to D2 (41.6 vs 66.2 and 69.9 vs 73.2 for stems and receptacles respectively). Several interactions affected oil weight

RMSE; when computing mean oil RMSE per genotype, we found that *cv*.Olledy displayed higher error (48.7 vs 36.4 for *cv*.Kerbel). Both nitrogen (***) and genotype (**) factors affected proteins weight; *cv*.Olledy had lower proteins error (14.4 vs 21.6 g m^{-2} for *cv*.Kerbel) and higher error on proteins was made in N- compared to N+ (22.6 vs 13.40).

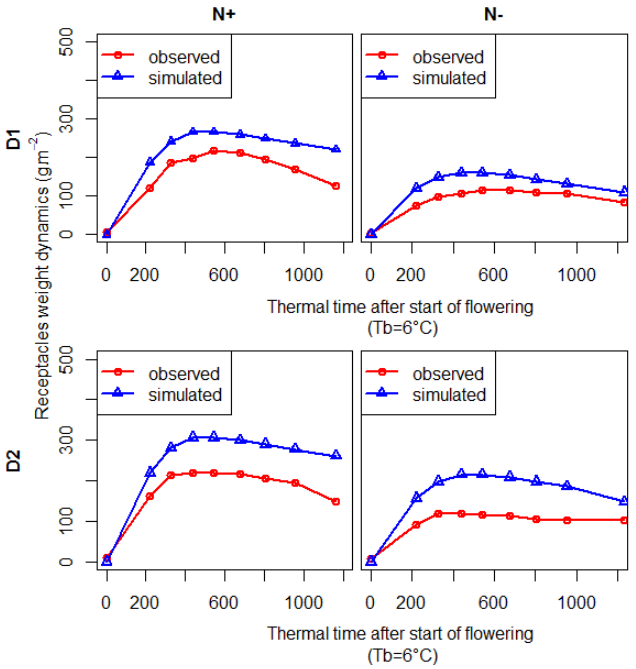
Nitrogen effect clearly influenced oil percent (RMSE: 5.2 in N+, 6.3 in N-). Proteins percent was strongly affected by both genotype (***), nitrogen (***) and genotype by nitrogen interaction (**). Error was lower in *cv*.Olledy (4.7 vs 6.6 proteins points for *cv*.Kerbel) and lower in N+.

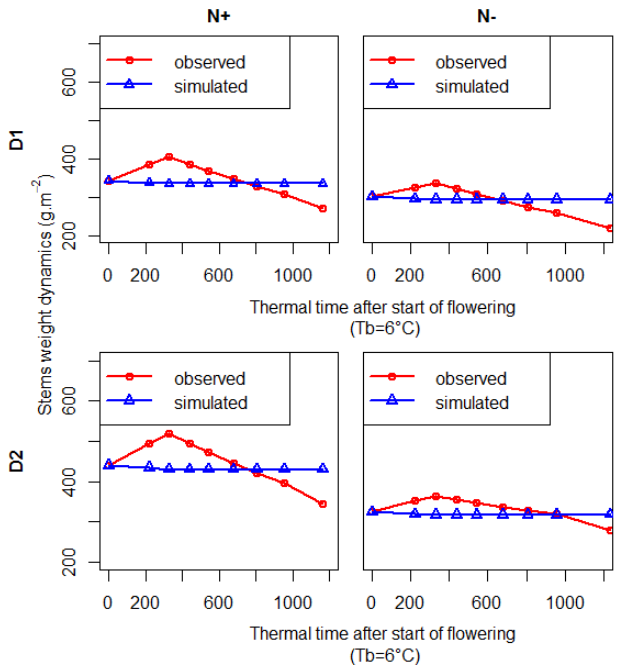
3.2. Evaluation results

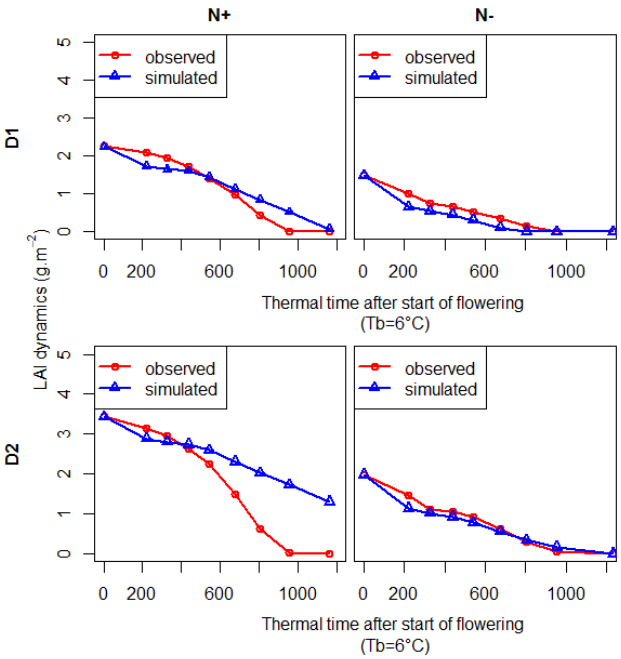
Less data was available for dynamic description in evaluation dataset, for which we compared simulated and observed LAI, oil and proteins percent (%) (Fig.4). The model predicted satisfactorily oil percent and LAI in their dynamic trend. Highest errors were observed in Agen 2013 experiment for the 3 variables except proteins percent (RMSEP = 2.62) and lowest in Le Magneraud. Quality of prediction in Auzeville 2011 and 2012 were similar. Efficiencies were very poor for LAI prediction in all sites (from -2.75 to 0.42), and low in Agen and Le Magneraud for proteins percent (-0.21 and 0.05 respectively).When computed over all sites and years (Table.5), we observed that all variables tended to be overestimated (negative bias).

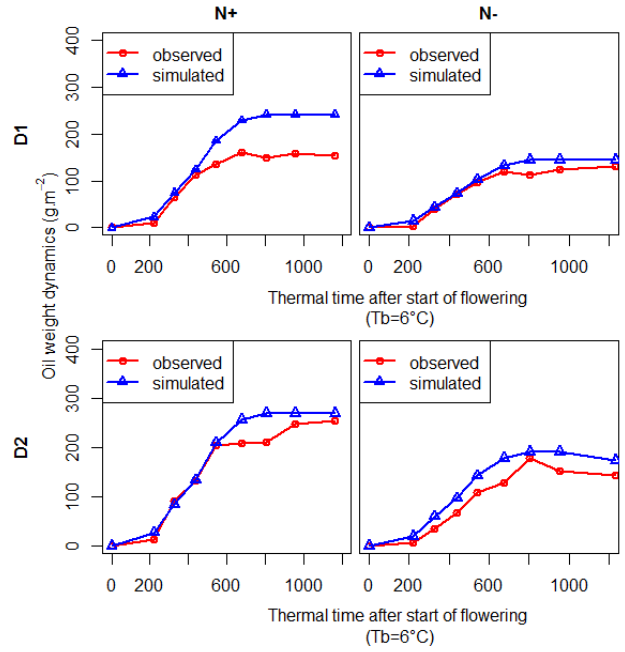
Table.3. Goodness of fit of source-sink model per genotype, water and nitrogen conditions and plant density in 2012 Auzeville parameterization dataset. Root mean squared error (RMSE), and model efficiency (EF) are provided for LAI, stems, receptacles, oil, proteins weights and oil and proteins percent.

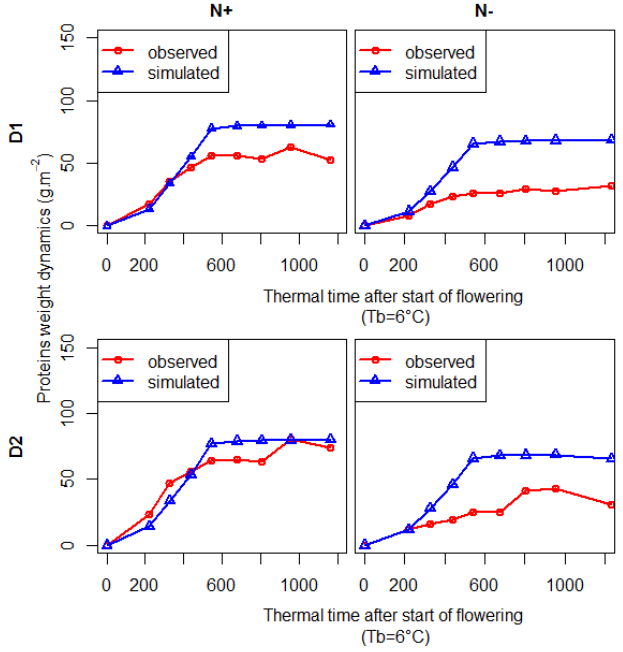
genotype	water.cond	N.cond	plant.density	RMSE						
				LAI	stems (g m^{-2})	receptacles (g m^{-2})	oil weight (g m^{-2})	protein weight (g m^{-2})	oil percent (%)	protein percent (%)
Kerbel	IRR	N+	D1	0.28	45.17	61.50	58.83	18.37	5.61	4.62
Kerbel	IRR	N+	D2	0.92	56.61	81.79	30.51	12.15	5.17	4.15
Kerbel	IRR	N-	D1	0.21	41.12	40.97	24.95	30.12	7.82	10.25
Kerbel	IRR	N-	D2	0.30	86.39	80.56	35.95	27.40	7.01	7.70
Olledy	IRR	N+	D1	0.58	71.60	84.77	50.15	11.58	5.10	4.26
Olledy	IRR	N+	D2	0.91	62.89	72.44	44.27	12.69	5.14	4.18
Olledy	IRR	N-	D1	0.42	33.78	103.28	54.57	20.04	4.94	4.87
Olledy	IRR	N-	D2	0.40	84.79	80.93	51.86	15.27	7.02	5.72
genotype	water.cond	N.cond	plant.density	EF						
				LAI	stems	receptacles	oil weight	protein weight	oil percent	protein percent
Kerbel	IRR	N+	D1	0.63	0.99	0.99	0.99	0.97	0.99	0.91
Kerbel	IRR	N+	D2	0.49	0.99	0.99	1.00	0.99	0.99	0.90
Kerbel	IRR	N-	D1	0.28	0.98	0.98	0.99	0.80	0.99	0.62
Kerbel	IRR	N-	D2	0.40	0.96	0.96	0.99	0.90	0.99	0.71
Olledy	IRR	N+	D1	0.51	0.99	0.99	0.99	0.98	0.99	0.90
Olledy	IRR	N+	D2	0.50	0.96	0.99	1.00	0.98	0.99	0.89
Olledy	IRR	N-	D1	0.49	0.95	0.97	0.99	0.89	0.99	0.80
Olledy	IRR	N-	D2	0.46	0.98	0.98	0.99	0.92	0.99	0.76

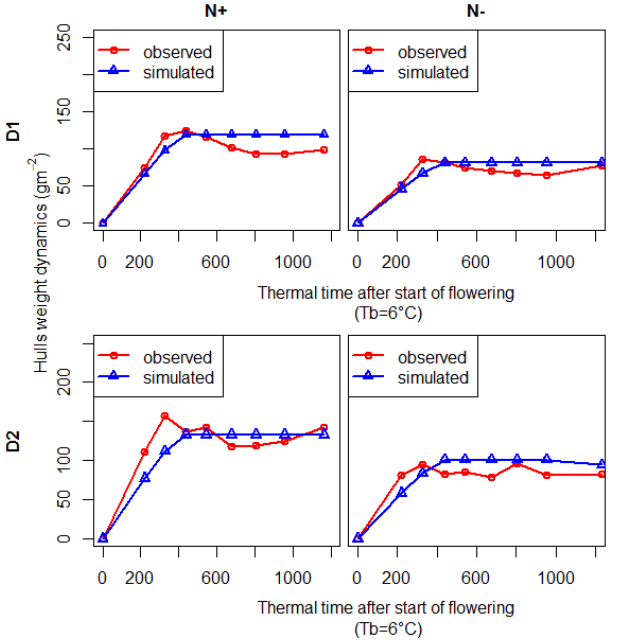

N+
N-
D1
D2
observed
simulated
Receptacles weight dynamics (g.m-2)
Thermal time after start of flowering
(Tb=6°C)


N+
N-
D1
D2
observed
simulated
Stems weight dynamics (g.m-2)
Thermal time after start of flowering
(Tb=6°C)


N+
N-
D1
D2
observed
simulated
LAI dynamics (g.m-2)
Thermal time after start of flowering
(Tb=6°C)


N+
N-
D1
D2
observed
simulated
Oil weight dynamics (g.m-2)
Thermal time after start of flowering
(Tb=6°C)


N+
N-
D1
D2
observed
simulated
Proteins weight dynamics (g.m-2)
Thermal time after start of flowering
(Tb=6°C)


N+
N-
D1
D2
observed
simulated
Hulls weight dynamics (gm-2)
Thermal time after start of flowering
(Tb=6°C)

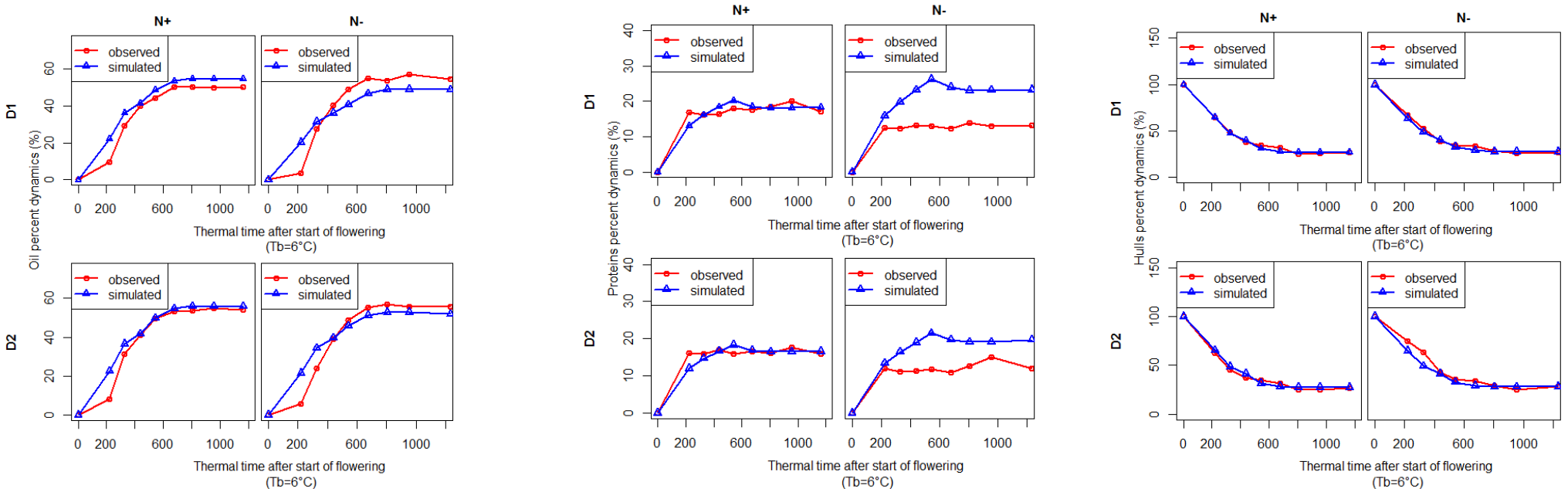


Fig.3. Results of simulations from parameterization dataset as illustrated for *cv*.Kerbel genotype. Observed and simulated data are compared: receptacles, stems, oil, proteins, hulls weights (g m^{-2}) and LAI and oil, proteins and hulls percent.

Table.4. Summary table of ANOVA performed on RMSE of LAI, stems, receptacles, oil, proteins weights per m², oil and proteins percent. Level of significance is indicated by «***» (p-value < 0.1%); «**» (p-value < 1%); «*» (p-value < 5%) and «.» (p-value < 10%). P-values higher than 10% were considered non-significant («-»symbol).

RMSE	LAI	stems weight ($g\ m^{-2}$)	receptacles weight ($g\ m^{-2}$)	oil weight ($g\ m^{-2}$)	protein weight ($g\ m^{-2}$)	oil percent (%)	protein percent (%)
genotype	-	-	-	*	**	.	***
N.cond	**	-	-	-	***	*	***
plant.density	*	-	-	-	-	-	-
genotype * N.cond	-	.	.	*	.	-	-
genotype * plant.density	-	-	-	-	-	-	**
N.cond*plant.density	*	-	-	*	-	-	.
genotype * N.cond*plant.de	-	-	-	.	-	-	-

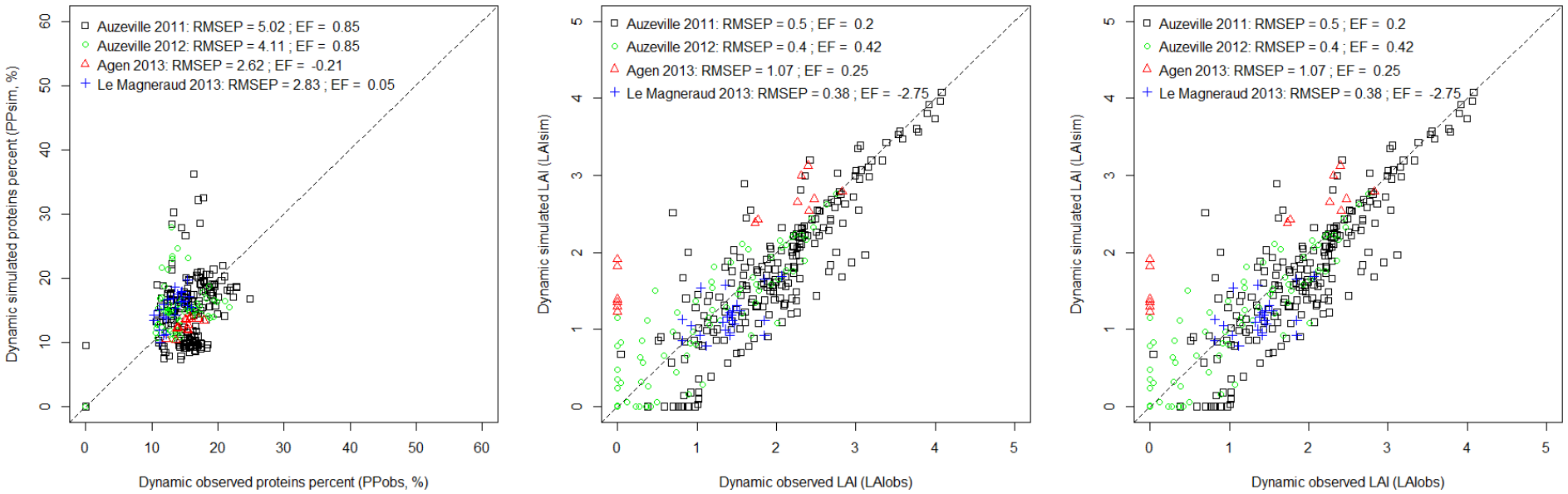


Fig.4. Comparison of simulated and observed values of oil and proteins percent and LAI in evaluation dataset (dynamic). Dots are distinguished by site-year. RMSEP, EF and R^2 are provided.

Table.5. Summary table of quality of prediction of the source-sink model regarding LAI, oil and proteins weights, oil and proteins percent. RMSEP, EF, and bias are provided. Values of indicators were averaged across all situations for all sites and years of the evaluation dataset.

	RMSEP	EF	R^2	bias
LAI	0.48	-0.71	0.73	0.08
oil weight (g m^{-2})	39.78	0.98	0.87	-3.91
protein weight (g m^{-2})	14.31	0.58	0.78	-0.64
oil percent (%)	6.10	0.97	0.94	-0.06
protein percent (%)	3.90	0.61	0.72	0.02

3.3. Sensitivity analysis results

We computed mean and standard deviation of μ* for 6 contrasted situations we selected (2 situations in 2011 (*cv*.Kerbel N+ D1, rainfed and irrigated) and 4 situations in 2013 (Le Magneraud: *cv*.Kerbel and *cv*.LG5451_HO, rainfed and irrigated). Parameters highly influencing either oil, proteins percent or oil and proteins weights per m² are represented on dynamic of accumulation of oil and proteins (Fig.6).

Most sensitive parameters could be categorized into 4 groups: N.uptake, photosynthesis, C.stems remobilization and oil. Some commonly shared parameters for oil weight and percent are those concerning photosynthesis: *b.RUEpot*, *a.RUEpot*, *t1.RUEpot*, *photosynthesis.TT.threshold*, as well as parameters concerning C.stems remobilization (*TT.C.stems.half*, *C.stems.rate*). Some parameters related to oil output were specific to oil percent such as *t0.oil_1*, *Kernel.oil.percent*, *hull.growth.1*, *Kernel.growth.1*, *t1.hull*.

For proteins, similar sensitive parameters could be found between weights and percents: *pool.N.threshold*, *N.uptake.TT*, *N.abs2.pot*. Though, it appears that proteins percent was influenced by changes in photosynthesis parameters (*b.RUEpot*, *a.RUEpot*, *t1.RUEpot*) and C.stems remobilization (*C.stems.rate*, *TT.C.stems.half*) and *t1.hull*, while proteins weight is

not. Some original sensitive parameters of proteins weight were *cap.growth.2*, *Kernel.growth.1*, *t0.oil_1*.

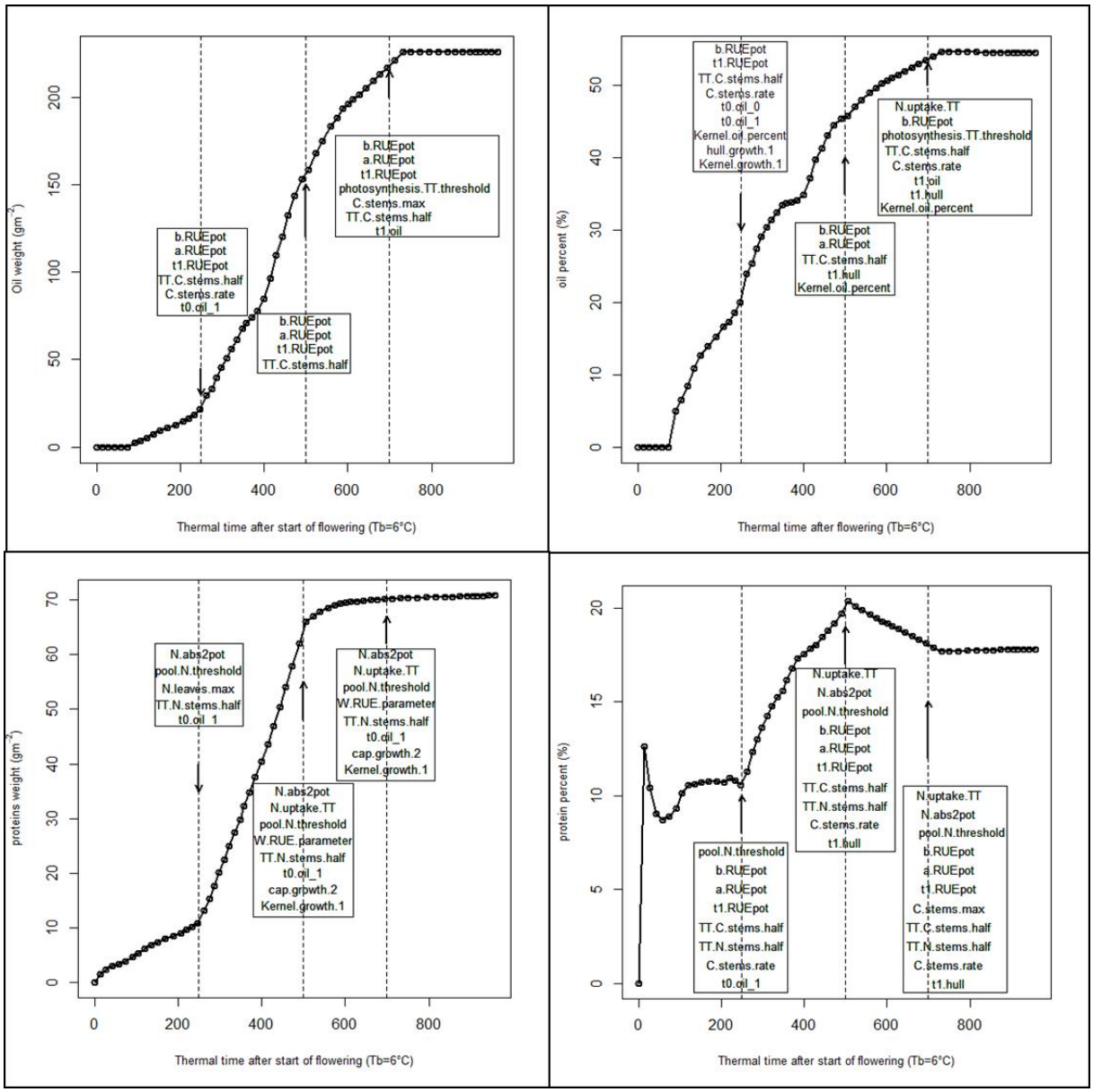


Fig.6. Summary graphics of sensitivity analysis performed on oil, proteins weights and percent. Boxes indicate parameters that were categorized "high" on output sensitivity. Arrows indicate the 3 thermal times we selected for running Morris screening method: 250, 500 and 700 °C days after flowering. Curves were obtained from mean simulated values in Auzeville 2011 for illustration.

4. Discussion

This work aimed at constructing a dynamic source-sink model for simulating oil and proteins accumulation in sunflower grains.

We opted for a simplified approach for describing acquisition of daily photosynthetic carbon after flowering (RUE). This is commonly used in other crop models such as CERES, EPIC or STICS (Palosuo et al., 2011). Though, this approach has already been criticized by Loomis and Amthor (1999) who argued that RUE models should be replaced by proper photosynthesis and respiration models. Indeed, respiration costs were not taken into account in this modelling approach. High respiratory load from receptacles was observed during grain filling (Hall et al., 1995; Ploschuk and Hall, 1997) and was considered to be the cause of the break of the relationship between RUE and SLN in sunflower.

Though, replacing the RUE model by photosynthesis and respiration model could not be adapted to our modeling objectives, *i.e.* simulating oil and proteins accumulation at square meter and daily scales. Photosynthesis models are usually described at hourly step and leaf level (Farquhar et al., 1980; Baldocchi and Amthor, 2001). Scaling up to canopy level was already attempted in other species (big-leaf models) but led to over-estimations of photosynthesis; de Pury and Farquhar (1997) and Chen et al. (2005) suggested to correct canopy photosynthesis by considering shaded and sunlit leaves proportions in canopy and compute consequent photosynthesis. This perspective appears to be time-consuming: it should be possible to mix RUE and biochemical or biophysical approaches, as attempted by Choudhury (2001) who distinguished between RUE and CUE (carbon use efficiency) such

that CUE = (RUE/gross photosynthesis)*IPAR/100 (in maize, wheat and rice). We could expect lower over-estimations of photosynthetic carbon from this approach.

Indeed, our results suggested that the main source of over-estimation of hulls, receptacles and kernels was the C.photo. In this study, we proposed unique RUE with thermal time relationship independently of nitrogen or genotype conditions. Apart from the fact that our RUE values appear to be overestimated (3.54 g/MJ IPAR) compared to those in literature (2.24 to 2.62 in Hall et al., 1995), RUE variability with nitrogen factor needs to be taken into account in the first part of grain filling. We could also suggest that since plants in non-limiting nitrogen conditions display larger demand, they could have higher respiratory load, which should be leading to diminished RUE values. Over-estimation of proteins is likely linked to the uncertainty of the parameters for which they are sensitive (N uptake rate, pool of usable nitrogen). Some growth chambers experiments could be needed to properly evaluate nitrogen dynamics in sunflower. At least, N uptake after flowering behaved as expected: higher in non-limiting nitrogen conditions. It could not be excluded that limiting nitrogen conditions lead to higher nitrogen remobilization while it is the reverse in N+. There could be a differential potential absorption at flowering that depends on the nitrogen status.

Carbon and nitrogen labelling are usually used for monitoring components fluxes in the plant. Though we simply used total carbon and nitrogen percent, our results appear to be close to reality. Since total carbon and nitrogen are variables that are lighter to use and collect, we suggest that total carbon and nitrogen could be a new way for characterizing nitrogen and carbon flux in other species for agronomic modeling purposes.An original result is the place stems carbon remobilization holds for determining oil weight and percent. Hall et al. (1989) demonstrated that assimilates stored before flowering contribution increased in water-stressed conditions. López Pereira et al. (2008) found by contrast that stems and receptacles contribution to grain weight was negligible. Our data suggest that differences in carbon

remobilization, probably related to genotypic (timings) or water stress (remobilization rate) differential responses, make the difference on oil quantity and percentage. López Pereira et al. (2008) also demonstrated that receptacles and stems weights could increase from flowering to physiological maturity; this concerns storage capacity of these organs resulting from remobilizations between reserve compartments or back from grains to vegetative organs (assimilates excess). Such behavior is not considered in our model, and could be later integrated as in Villalobos et al. (1996) or Lescourret et al. (1998). There could be likely a negative feedback on photosynthesis (Iglesias, 2012). Priority rules based on thermal times was relevant because it permitted to reproduce all components dynamics quite well, both at source and sink level. Combined with the minimum law (Tabourel-Tayot and Gastal, 1998a, b), it appeared to us to be really respecting source-sink principle. Another interesting result is the fact that parameters influencing proteins quantity are not the same as for proteins percent. The latter is more influenced by factors "belonging" to oil percent (*b.RUEpot*, *a.RUEpot*, *t1.RUEpot*). This proves that oil and proteins accumulate independently and are determined by different factors; since oil and proteins percent rely on similar parameters, it is not surprising to find final negative correlation between oil and proteins percentages (Connor and Hall, 1997; Aguirrezábal et al., 2009). We previously demonstrated that genotypic effects played through timings and durations of mainly sink processes (Andrianasolo et al., 2014; chapter II). That is, the appearance of timing parameters at sink percentage level (*t1.hull*, *t0_oil.1* …) confirms that genotypic differences are marked at the sink level, particularly the ratio between a sink component and the seed dry weight.

Hulls percent evolution was well simulated in all situations. It shows that despite over-estimations of oil and proteins quantities and percent, we were able to reproduce the ratio between those components. We believe we are well on track to achieve an extrapolated model.

The assumptions we made on LAI dynamics led to good predictions of LAI. It is consistent with the 2 phases (low and then high senescence rate) described by de la Vega et al. (2011). We consider that after flowering, nitrogen content of leaves inevitably diminishes till reaching an N content at which leaves begin their second phase of rapid senescence. Differential behavior could exist since senescence rate is very low in limited-nitrogen plants, which makes difficult to achieve 2% of initial N content. This justifies the use of a physiological thermal time threshold at which leaves die independently of nitrogen conditions.

LAI was overestimated particularly in N+ x D2 in the second part of senescence; we might further consider senescence that is caused by excess of assimilates (Thomas, 2003; 2013), or senescence caused by shading in high density canopy (Gabrielle et al., 1998).

Although the prediction quality is not comparable to that of the statistical model (RMSEP = 6.06 vs 3.8 for the model of Casadebaig, 2008), we can consider that this model is satisfying for what it was first aimed for: describing daily oil and proteins accumulation in sunflower grains. The source-sink model required more parameters (66 vs 8 for the statistical model). Sensitivity analysis on several contrasted situation indicated that part of them were little sensitive and we may be able to simplify the formalism of the model to make it more operational and integrate it into SUNFLO model.

Eventually it should be used to predict, for any date of interest to the farmer, the oil content and proteins that could be achieved with an acceptable margin of error, knowing the genotypes and the environment in which the crop was grown.

5. Conclusions

*The aim of this work was to build a dynamic source-sink model for simulating oil and proteins accumulation in sunflower grains. It relies on the description of carbon and nitrogen dynamics among sources (stems, leaves, receptacles) and sinks (receptacles, hulls and kernels).

*A parameterization dataset consisting of 24 situations and an evaluation dataset (50 situations) were used for building the model.

*Quality of prediction was 6 and 4 points for oil and proteins percent respectively.

* From results of sensitivity analysis, further investigation should be focused on N uptake, photosynthesis, stems carbon remobilization and genotypic variability at sink level.

Bilan du chapitre IV

- Le chapitre IV a été dédié à la construction, le paramétrage et l'évaluation du modèle dynamique. Ce dernier est basé sur la description des relations source-puits régies par des règles de priorité (ordre chronologique et/ou en fonction de l'état d'épuisement des ressources).
- Les informations acquises dans les chapitres II (timings, durées, vitesses, effets des facteurs) et III (valeur du paramètre de régulation photosynthétique en réponse au stress hydrique) sont pris en compte dans le modèle. Il n'y a pour le moment pas de variabilité génotypique explicitée dans le modèle.
- Le modèle source-puits a été couplé au modèle SUNFLO afin de simuler les disponibilités hydrique et azotée journalières.
- Le modèle source-puits a été paramétré sur une base de 24 situations et évalué sur un jeu indépendant de 50 autres situations, avec une diversité au niveau des années, des génotypes, des conditions azotées et hydriques et des densités de peuplement.
- Les résultats montrent que le modèle reproduit fidèlement les dynamiques des sources et des puits mais les valeurs sont sur-estimées, notamment pour la teneur en huile et en protéines (RMSEP= 6.1 et 3.9 respectivement). Un travail d'optimisation des paramètres devrait permettre de réduire l'erreur de prédiction.
- L'analyse de sensibilité (méthode de Morris) a permis d'identifier les paramètres les plus sensibles sur les quantités et teneurs en huile et protéines ; ces paramètres sont liés aux processus d'absorption d'azote, de photosynthèse, de remobilisation carbonée depuis les tiges et de timings au niveau des puits carbonés (coque, huile).

Dans la discussion générale, des éléments complémentaires seront apportés concernant la capacité du modèle à reproduire des effets agronomiques (notamment azote) tels qu'ils ont été observés au champ (Chapitre II) ainsi que les relations entre assimilations/remobilisations carbonées et azotées.

Discussion générale, conclusions et perspectives

Ce travail de thèse a porté sur l'analyse et la modélisation de l'élaboration de la qualité des graines (teneurs en huile et en protéines) chez le tournesol. En effet, l'huile (alimentaire et celle incorporée dans le biodiesel) et les protéines (via les tourteaux) sont les principaux débouchés de l'utilisation de ses graines, et font l'objet de recherches agronomiques et industrielles ayant pour but d'optimiser leurs teneurs, ce qui permettra au tournesol de rester compétitif sur le marché des oléagineux.

Pourtant, la variabilité des rendements et teneurs en huile et protéines des graines en fonction des années, conditions pédoclimatiques et interactions génotype et environnement crée une instabilité économique pour le producteur. Les teneurs en huile et protéines sont sous contrôle génétique et environnemental, mais les parts de chacun de ces facteurs semblent variables en fonction des conditions de cultures. Afin d'optimiser ces teneurs, il nous a paru essentiel de disséquer les effets de chacun de ces facteurs sur les teneurs en huile et en protéines. Ceci nous a permis de proposer un modèle fonctionnel et dynamique d'élaboration de l'huile et des protéines dans des conditions de disponibilités hydrique et azotée et chez des génotypes contrastés.

Le chapitre I (modélisation statistique) nous a permis d'identifier et hiérarchiser les facteurs déterminants de la teneur en huile, tandis que les chapitres II et III étaient consacrés à la compréhension du mode d'action de ces facteurs (azote, densité et génotype pour le chapitre II ; eau et génotype pour le chapitre III) dont le déterminisme et/ou les interactions ne sont que peu documentés dans la littérature. Ces nouveaux éléments nous ont permis d'asseoir les hypothèses fortes de notre modèle dynamique (chapitre IV).

Nous allons passer en revue les principales avancées dans chacun de ces chapitres.

1. La modélisation statistique pour hiérarchiser et trier les facteurs importants de la teneur en huile

Nous avons proposé de construire 3 types de modèles statistiques pour la teneur en huile (régression linéaire multiple, modèle additif généralisé (GAM) et arbre hiérarchique (Regression tree - RT) et de comparer leurs performances avec le modèle de Pereyra-Irujo et Aguirrezábal (2007) ; ce dernier simule la teneur en huile à partir de trois co-variables (teneur en huile potentielle, rayonnement intercepté entre 250 et 450°C jours après la floraison et densité de peuplement) et a été évalué chez un seul génotype dans des milieux non-limitants (azote et eau). Les qualités de prédiction (RMSEP) sont de 3.3, 2.4, 1.9 et 2.5 points d'huile pour les modèles de Pereyra-Irujo et Aguirrezábal (2007), linéaire multiple, GAM et arbre, respectivement. De plus, nous avons procédé à une analyse de la contribution relative des différentes co-variables retenues dans chaque modèle. Les résultats marquants sont les suivants :

- Les trois nouveaux modèles construits ont en commun la hiérarchie des facteurs les plus importants pour la teneur en huile : il s'agit en premier lieu du génotype (avec une contribution de 25 à 56% pour les 3 nouveaux modèles construits), suivi du fonctionnement en post-floraison (LAD2, MRUE2 –contribution de l'ordre de 20%), des facteurs azote, eau, température et peuplement.
- Les modèles statistiques prédisent bien les effets moyens (tendance générale) et sont conformes à la littérature: plus la teneur en huile potentielle (génotypique) est élevée, plus la teneur en huile observée l'est aussi (Izquierdo et al., 2008) ; de même les teneurs élevées correspondent à des durées de fonctionnement de la source longues (Merrien, 1992). A de fortes doses d'azote, les teneurs en huile diminuent (Zheljazkov et al., 2009). Le stress hydrique et les fortes températures (supérieures à 34°C) pénalisent la teneur en huile (Chimenti et al., 2001 ; Rondanini et al., 2003).
- Les modèles montrent une diversité de réponses vis-à-vis de l'effet densité (Fig.1); en général plus la densité est élevée, plus la teneur en huile l'est aussi (sauf aux très fortes densités de l'ordre de 8 plantes par m²). Cette diversité de réponses serait davantage liée à la manière dont les modèles sont construits plutôt qu'une capacité à

reproduire les variations des effets de la densité, qui sont généralement positifs mais peuvent être «effacés » ou « pénalisés » par d'autres facteurs de stress.

Néanmoins, cet effet densité qui diffère selon les modèles montre déjà une première limitation de la modélisation statistique : tous les modèles proposés sont en partie faux, dans le sens où ils ne reproduisent pas de façon correcte toute la gamme de réponses de la teneur en huile à une gamme de doses d'azote, de densités. Notre explication est la suivante : plus la gamme de stress (hydrique, azoté) est fort, plus la gamme de réponses génotype x eau (G x W), génotype x densité (G x D), génotype x azote (G x N) voire génotype x azote x eau (G x N x W)… est élevée (Gallais, 1992), issue de l'adaptation propre des génotypes à ces conditions de cultures.. Différentes interactions entre les facteurs déterminant la teneur en huile ont été mises en évidence (Fig 2). Pourtant, même avec une bonne caractérisation de chacune de ces situations de stress, il serait impossible de trouver un modèle statistique qui satisfasse à toutes les situations, la contribution relative des facteurs déterminants étant aussi variable que la diversité des réponses décrites. Le modèle statistique ne permet que peu ou pas de reproduire les IGEC (il y a potentiellement autant d'IGEC que de niveaux de stress pour un génotype dans un milieu donné, Fig.2). Lorsqu'il le fait, le modèle ne pourrait être extrapolé et ne serait valide que dans les mêmes sites ou sites similaires à ceux utilisés pour la construction du modèle.

La modélisation statistique s'est avérée essentielle dans notre démarche de modélisation dynamique car elle a permis l'identification et la hiérarchisation des facteurs déterminant la qualité (teneur en huile). Non seulement elle permet de proposer une version améliorée du modèle SUNFLO existant (Casadebaig et al., 2011), mais elle a été nécessaire dans la démarche de conceptualisation d'un modèle plus fonctionnel. Elle constitue de fait la première étape de toute modélisation en aidant au tri de facteurs importants pour la qualité face à une bibliographie parfois contradictoire. Cette partie répond donc à la question : « Quels sont les variables/facteurs importants et comment sont-ils hiérarchisés ? »

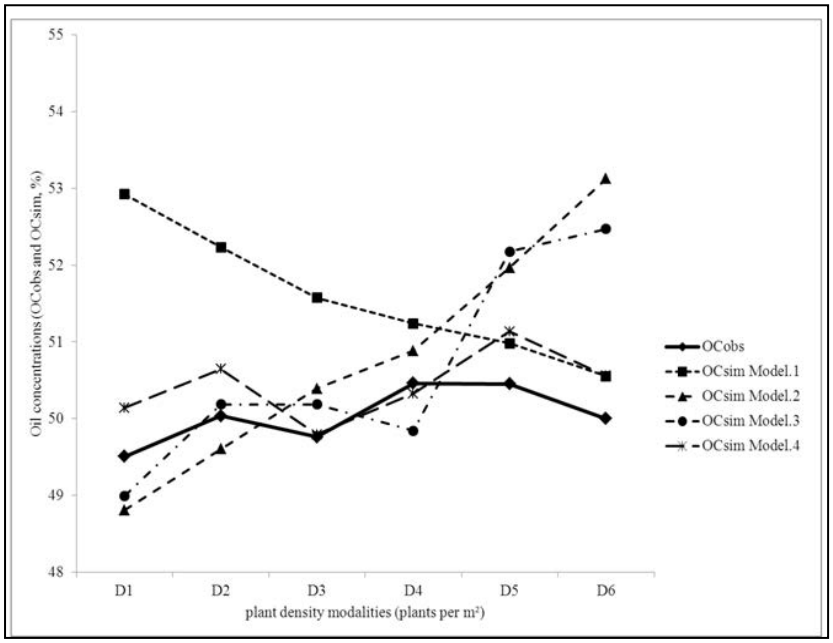


Fig.1. Observations et simulations des teneurs en huile (%) en fonction de la densité de peuplement (D1 à D6, correspondant à 3 à 8 plantes par m²). Model.1, 2, 3 et 4 correspondent au modèle de Pereyra-Irujo et Aguirrezábal (2007), modèle linéaire multiple, GAM et arbre de régression hiérarchique, respectivement. La dynamique de la teneur en huile moyenne observée est indiquée en gras. Schéma repris d'Andrianasolo et al. (2014)

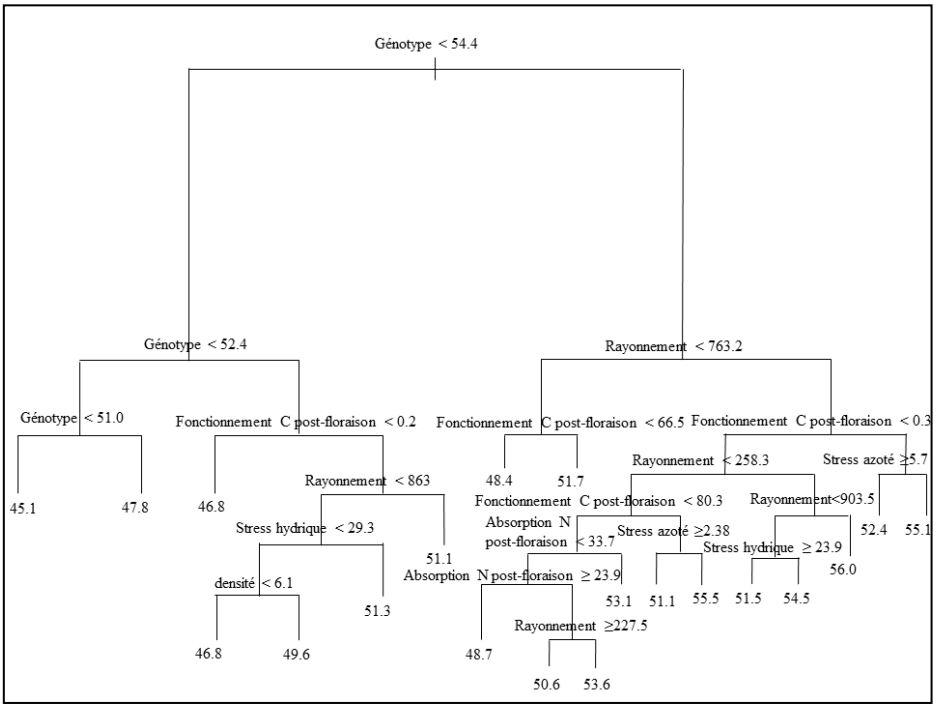


Fig.2. Illustration des résultats de la régression hiérarchique (regression tree) obtenus dans le cadre de la modélisation statistique. Schéma issu d'Andrianasolo et al. (2014), les indicateurs (prédicteurs) ont été remplacés par la catégorie à laquelle ils appartiennent pour plus de lisibilité.

2. Des expérimentations au champ et en serre pour décortiquer les effets des facteurs sur les déterminants des teneurs en huile et protéines

Une fois les facteurs les plus importants sélectionnés, notre deuxième étape vers la modélisation fonctionnelle a été de mettre en place des expérimentations afin de mieux décrire et comprendre le mode d'action des différents facteurs. Cette deuxième partie répond à la question : « Comment les facteurs agissent-ils sur les variables déterminant l'huile et les protéines » ?

2.1. L'analyse des relations source-puits et des effets génotype, azote et densité

Le chapitre II a été consacré à l'analyse des dynamiques « source » (feuilles, tiges, réceptacles) et « puits » (huile et protéines) chez 2 génotypes soumis à des conditions d'azote (N- : non fertilisé ; N+ : 150 kg par ha), et de densités de peuplement (D1 : 3 et D2 : 4.5 plantes par m²) contrastées. La disponibilité en eau n'était pas limitante (culture bien irriguée). L'hypothèse forte de cette expérimentation a été la suivante : l'huile et les protéines exprimées en teneurs/concentrations ne reflètent pas systématiquement les effets des différents facteurs sur l'accumulation de l'huile et des protéines en quantités, car ces facteurs affectent les sources et puits à des périodes (timings) et des intensités différentes. Cette hypothèse a été vérifiée car nos résultats montrent que:

- L'apport d'azote a un effet positif sur les quantités d'huile, de péricarpe et de protéines, mais son effet est plus fort sur les protéines, puis les coques, puis l'huile. Ainsi, la teneur en huile apparaît plus faible dans des conditions de fortes disponibilités en azote.
- Les situations en condition azotée non-limitante montrent des quantités d'azote absorbées en post-floraison, des durées de surface foliaire et des quantités d'azote remobilisées plus élevées, comparées à celles en condition non-fertilisée.
- L'apport d'azote change les proportions des sources d'allocation de carbone vers les graines (davantage de remobilisation carbonée en N- et moins de carbone issu du fonctionnement photosynthétique en post floraison). De plus, l'apport d'azote entraîne une augmentation des vitesses de remobilisation carbonée et azotée dans les feuilles,

tiges et réceptacles, et de vitesse d'accumulation des quantités d'huile, de coques et de protéines.

- L'effet densité se traduit par une plus forte vitesse d'accumulation des protéines en D1 et une plus forte vitesse de sénescence des feuilles en D2.
- Les génotypes diffèrent en termes de démarrage de la période « source » (timing) et de durée de remobilisation pour les feuilles, tiges et réceptacles. Seules les teneurs finales (à la récolte) des deux génotypes sont significativement différentes à l'échelle de la source.

Cette partie apporte beaucoup d'éléments nouveaux, aussi bien en termes de méthodes que de résultats concernant la physiologie et l'écophysiologie de la qualité. A notre connaissance, les contributions des remobilisations vs assimilations azotées et carbonées n'ont pas été quantifiées chez le tournesol. L'originalité de notre démarche réside dans le fait d'analyser les effets des facteurs sur des dynamiques en comparant les paramètres (au nombre de 4) d'un modèle bi-linéaire. Une approche similaire a déjà été proposée par Yin et al. (2009) concernant un modèle sigmoïde amélioré afin de comparer des génotypes de blé ; le modèle bi-linéaire s'est avéré adapté à toutes les dynamiques de source et puits que nous voulions comparer et est à nos yeux moins lourd, tout en possédant des paramètres interprétables biologiquement. Il reste à vérifier que le modèle plus fonctionnel est apte à reproduire ces mêmes effets.

2.2. L'analyse de la réponse à la contrainte hydrique et les effets stade et génotype

Deux expérimentations en serre (2009 et 2012) ont permis d'analyser la réponse de génotypes de tournesol (2 en 2009 et 3 en 2012, dont 1 en commun pour les deux années) à la contrainte hydrique (dessèchement progressif). En 2009, 3 étages foliaires sont suivis en phase de pré-floraison tandis qu'en 2012, un seul étage foliaire est utilisé pour comparer la régulation des flux transpiratoires en pré- et post-floraison. L'argument fort de cette partie est le suivant : les différences de régulation transpiratoire avant et après la floraison peuvent être expliquées par des différences dans les proportions de feuilles qui coexistent sur une même plante et dont les âges, donc le fonctionnement, diffèrent. De nouveaux éléments sont également apportés concernant la réponse de la transpiration et de la photosynthèse à un dessèchement progressif du sol :

- Avant la floraison, un dessèchement progressif du sol réduit tout d'abord la transpiration puis la photosynthèse, confirmant une plus forte sensibilité de la transpiration à la contrainte hydrique comparativement à la photosynthèse. En revanche, cette différence

de sensibilité des processus à la contrainte hydrique apparaît moins marquée en phase post-floraison ; la photosynthèse s'avérant aussi sensible que la transpiration lorsque la contrainte hydrique s'applique après floraison. Ce résultat suggère une limitation stomatique et/ou non stomatique (altérations des photosystèmes/ limitation biochimique à la fixation du CO2) plus marquée de la photosynthèse en réponse à une contrainte hydrique en phase post-floraison.

- Des différences génotypiques ont été observées concernant la sensibilité de la transpiration et de la photosynthèse à la contrainte hydrique. En post-floraison, le cultivar Melody semble maintenir une certaine activité photosynthétique après le début de fermeture progressive des stomates. Il existe donc des interactions (Génotype x eau, G x W) au niveau de la source concernant la régulation des flux transpiratoires et photosynthétiques.
- Des différences de sensibilité au stress hydrique ont également été mises en évidence selon l'âge des feuilles : les feuilles jeunes complètement développées régulent de façon plus tardive leur flux transpiratoire, par rapport à des feuilles plus âgées (perte de contrôle stomatique) ou en expansion (stomates incomplètement développés).

Nous avons démontré que les dynamiques de réponse de la transpiration et la photosynthèse à la contrainte hydrique sont différentes avant et après floraison. Ces différences ont été déjà évoquées par Connor et Hall (1997). Dans le cadre de la représentation de la variabilité génotypique de la réponse au stress hydrique dans le modèle SUNFLO (Casadebaig et al., 2011), il serait ainsi plus pertinent de tenir compte des valeurs de paramètres en post-floraison pour représenter des effets de stress hydrique sur des variables s'élaborant en post-floraison (huile, protéines). La mise en relation de ces paramètres de régulation avec les variables finales que sont la teneur en huile et en protéines devrait être réalisée afin d'évaluer l'impact du stress hydrique foliaire sur la qualité. Nous notons néanmoins que le phénotypage est beaucoup plus difficile à réaliser en conditions de post-floraison et qu'un suivi jusqu'à la récolte paraît peu envisageable. Une piste proposée serait de faire un prélèvement de graines pour une mesure de la teneur en huile et en protéines à une date où l'on sait que ces teneurs sont corrélées à celles obtenues à la récolte. La modélisation dynamique pourrait permettre d'identifier la fenêtre thermique/date adéquate pour cette mesure.

L'ensemble des résultats de ces expérimentations nous permet de proposer le schéma relationnel suivant :

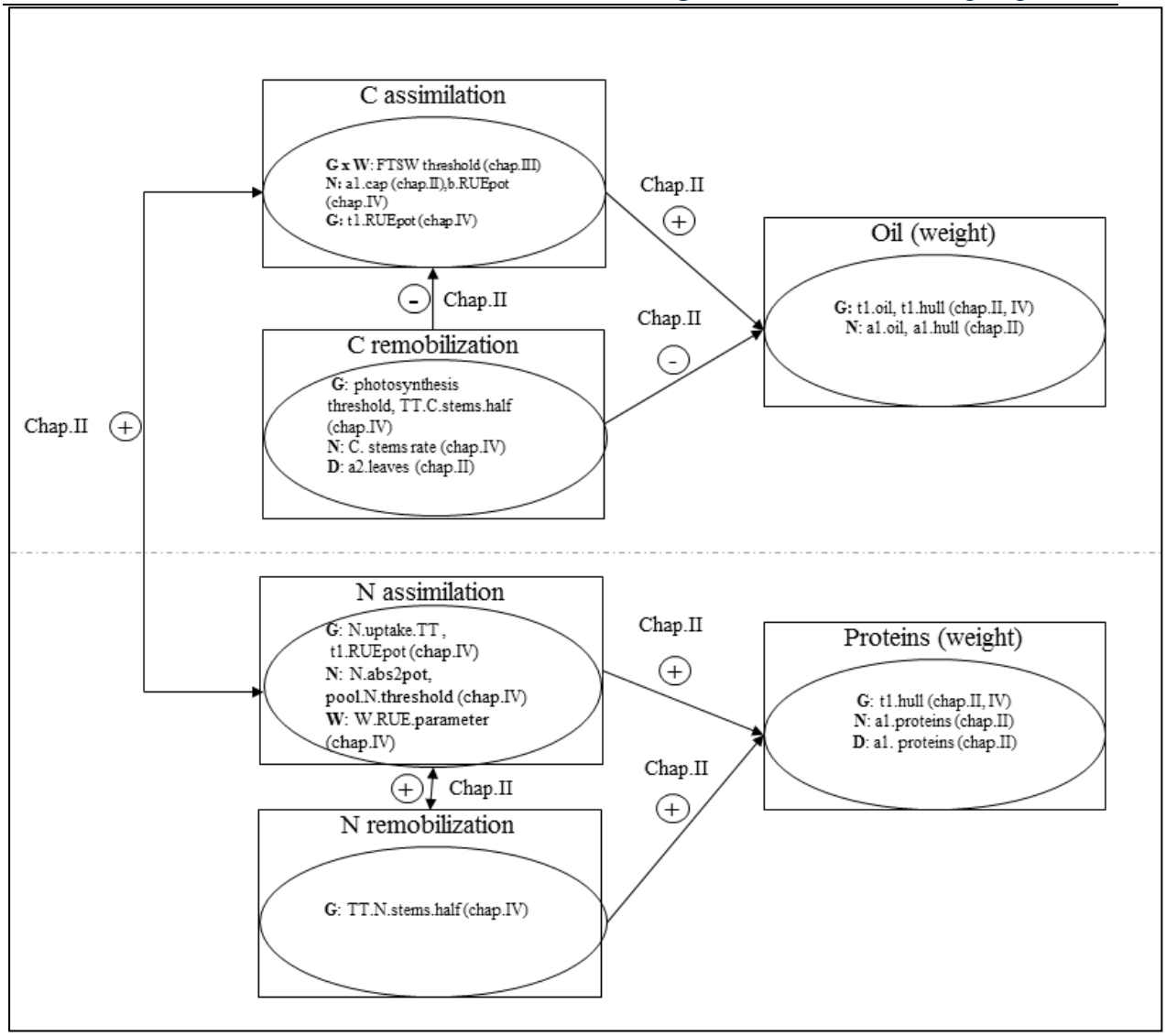


Fig.3. Schéma récapitulatif des principales relations et paramètres établis entre les différentes composantes « source » et « puits » pour la détermination des teneurs en huile et protéines. G, N, W et D correspondent au facteur génotype, azote, eau (stress hydrique) et densité de peuplement, respectivement. Les chapitres qui ont permis d'établir les relations entre composantes et/ou d'identifier les facteurs et paramètres importants sont indiqués.

Les effets « densité » et « température » ont été retenus dans la modélisation statistique (chapitre I) et devraient être analysés dans le cadre d'autres expérimentations sur la « source » et sur le « puits ».

3. La modélisation dynamique basée sur les relations source-puits

Un modèle fonctionnel basé sur les relations source-puits (flux de carbone et d'azote) est proposé pour simuler de façon journalière (à partir de la floraison) l'accumulation de l'huile et des protéines dans les graines de tournesol. Les principaux processus représentés sont les phénomènes de remobilisation carbonée et azotée depuis les tiges, feuilles et réceptacles, ainsi que les processus d'assimilation (photosynthèse et absorption d'azote). Le modèle tient compte de conditions de stress hydrique (simulées par SUNFLO) et azotée (INN mesuré à la floraison, la dynamique de disponibilité en azote du sol étant simulée par SUNFLO en post-floraison). Les principales variables de sortie (variables élaborées) sont les quantités (en g m^{-2}) et les teneurs en huile et protéines. Le modèle a été paramétré à partir des données de l'expérimentation au champ 2012 (24 parcelles) puis évalué sur 50 parcelles indépendantes (3 années, 3 génotypes, 5 niveaux de densités de peuplement, différents niveaux de disponibilité hydrique et azotée). Enfin, une analyse de sensibilité (méthode de Morris) a été effectuée sur quelques situations afin d'identifier les paramètres les plus influents sur les teneurs en huile et protéines et de proposer de futurs paramètres génotypiques, non inclus dans le modèle actuel.

Ce dernier a permis les avancées suivantes :

- Proposition d'une trame conceptuelle à la fois nouvelle et cohérente avec les éléments de la littérature (modèle d'intégration) ;
- Réflexion sur la nécessité d'un meilleur paramétrage pour les processus d'acquisition de l'azote, de remobilisation du carbone depuis les tiges, qui ont fait peu ou pas l'objet d'études dans le passé.
- Suite à l'analyse de sensibilité, des pistes sur les paramètres qui méritent une « explicitation » génotypique (*t0.oil_0*, *t0.oil_1*, *t1.hull*,…)

Nous avons également pu vérifier les performances du modèle :

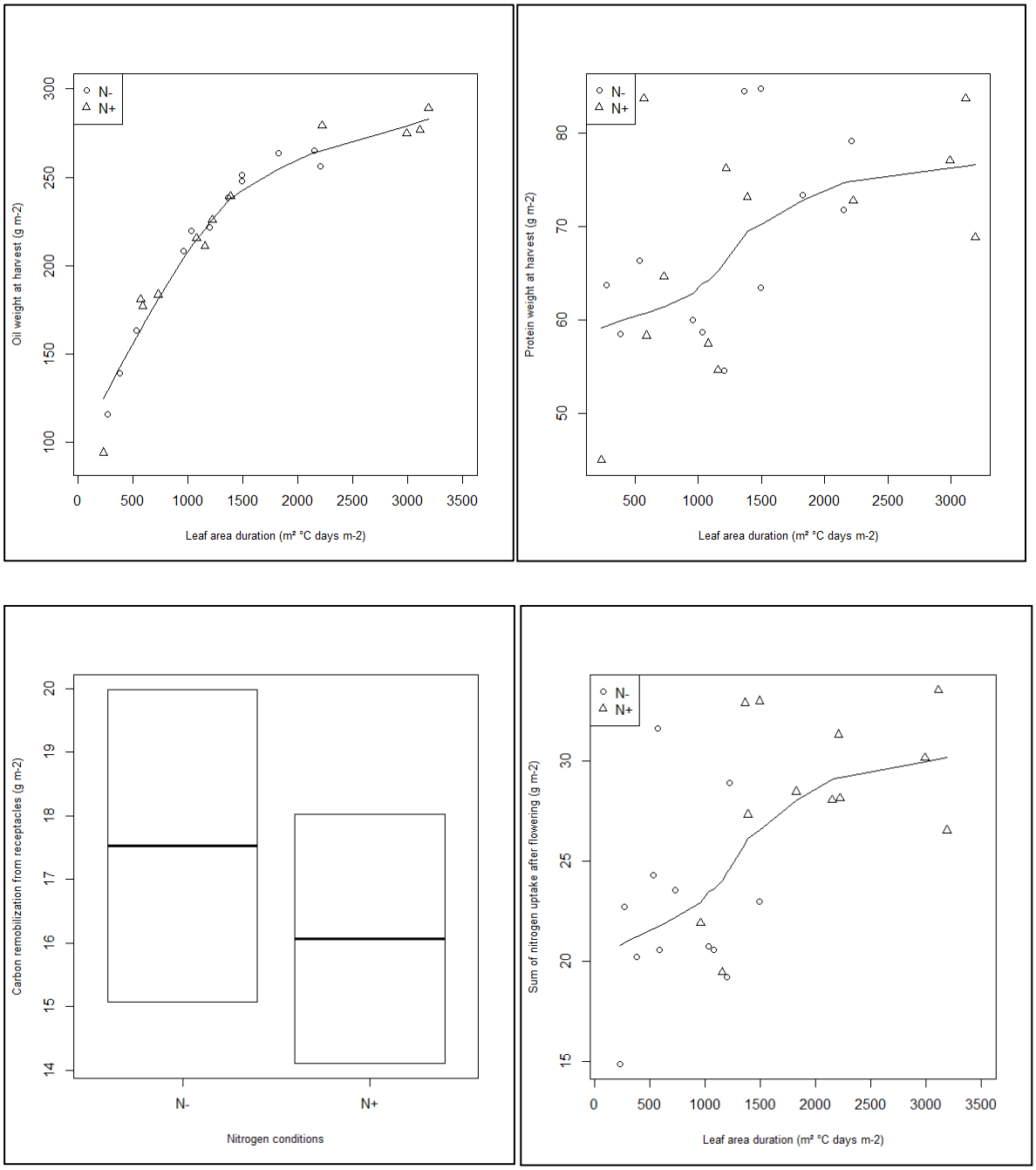


Fig.4. Relation entre durée de surface foliaire (leaf area duration, LAD) et poids d'huile à la récolte (en haut à gauche) et entre LAD et poids de protéines à la récolte (en haut à droite). Représentation de la remobilisation carbonée depuis les réceptacles en fonction de la condition azotée (N+, N-) en bas à gauche, et relation entre LAD et somme d'azote absorbée après floraison (en bas à droite). Toutes les données représentées sont issues de la simulation par le modèle source-puits dynamique. Les courbes de tendance ont été effectuées avec la fonction « *lowess* » sous R.

Les simulations du modèle sont fidèles aux tendances générales observées (chapitre II) ; plus le LAD est élevé, plus les quantités d'huile et de protéines sont élevées. Les situations où les LAD sont plus élevés correspondent à celles où l'absorption d'azote en post-floraison est également plus grande (mais les quantités d'azote absorbées sont fortement surestimées).

Enfin, nous vérifions bien l'observation selon laquelle il y a davantage de remobilisation de carbone en condition azotée limitante (Fig.4).

Néanmoins, les paramètres de ce modèle doivent encore être optimisés d'autant qu'il reste beaucoup d'incertitudes liées au manque de connaissances sur plus de la moitié des paramètres. Dans l'état, le modèle reproduit bien l'évolution au cours du temps des sources et des puits mais surestime les variables de sortie qui nous intéressent, notamment les teneurs en huile (RMSEP= 6.1) et les protéines (RMSEP=3.9).

L'effet des températures n'est pas inclus ; sachant que les températures jouent sur les durées de remplissage (Chimenti et al., 2001), il doit exister une interaction génotype x température (une sensibilité différentielle aux fortes températures qui dépend du génotype) ; preuve en est les différents seuils de sensibilité décrits dans la littérature (Chimenti et al., 2001 (25°C); Rondanini et al., 2003 (34°C) ; Angeloni et al., 2012 (17°C)).

L'existence de paramètres non sensibles questionne sur la pertinence de leur exclusion du modèle ; les exclure devrait permettre d'alléger le modèle en espérant augmenter sa qualité de prédiction (Thornley et Johnson, 1990). Les garder serait aussi justifiable dans la mesure où la non-sensibilité ne signifie pas pour autant «l'inutilité » : l'exclusion de certains paramètres à sens biologique nous semble délicat, d'autant qu'un paramètre peut être non sensible parce que le formalisme choisi n'est pas le bon. L'analyse de sensibilité nous a permis dans un premier temps de simuler des variations que l'on suppose liées au génotype (timings, durées) et aux conditions de l'environnement (vitesses). Il est donc rassurant que les teneurs en huile et protéines soient sensibles aux paramètres pour lesquels nous avions une forte hypothèse de sensibilité génotypique et environnementale.

De plus, certains paramètres n'ont pas de signification biologique propre et ont été obtenus par modélisation statistique/empirique, tels que ceux impliqués dans la relation INN et ratio de tiges ou ceux utilisés pour calculer les ratios d'azote dans chaque compartiment source en début de simulation. Le paramétrage a été effectué sur 24 parcelles en 2012 ; il serait opportun de valider ces relations pour d'autres années et génotypes.

Un des points délicats de cette modélisation fonctionnelle a été la prise en compte de l'échelle à laquelle a lieu chaque processus. Par exemple, Aguirrezábal et al. (2009) argumentent que la teneur en huile et protéines sont déterminées à l'échelle du couvert et qu'elles ne pourraient pas être bien prédites en faisant une simple extrapolation de l'échelle individuelle (plante) à l'échelle du couvert.

Nous avons comparé nos données (exemple de l'année 2012 : génotype Kerbel ; deux conditions azotées, deux densités de peuplement, eau non-limitante) issues des teneurs en huile et protéines obtenues à l'échelle capitule et celles à l'échelle couvert (récolte). En prenant comme référence les teneurs à l'échelle du couvert, celles individuelles plus faibles ont tendance à être surestimées tandis que les teneurs en protéines plus fortes sont sous-estimées (Fig.5). Les corrélations entre les teneurs au niveau des deux échelles sont néanmoins très significatives, à la fois pour l'huile (R^2 = 0.87) et pour les protéines (R^2 = 0.86). Compte-tenu des pertes entraînées par les dégâts d'oiseaux et la verse en cours de remplissage, nous considérons nos teneurs à l'échelle du capitule comme étant plus fiables. De plus, nous considérons qu'il n'y a pas de biais lié à l'échantillonnage des capitules (3 plantes prélevées de façon aléatoire par parcelle et par date de prélèvement). Une source de surestimation de nos valeurs pourrait provenir de la non-prise en compte de la variabilité intra-capitulaire : un gradient de poids de grains et de teneurs en huile existe (Merrien, 1992 ; Alkio et al., 2002). Nous proposerions d'évaluer la contribution des différentes cohortes de graines dans le poids et les teneurs totaux à l'échelle de la plante. Les pourcentages de contribution pourraient ensuite être intégrés dans le modèle fonctionnel.

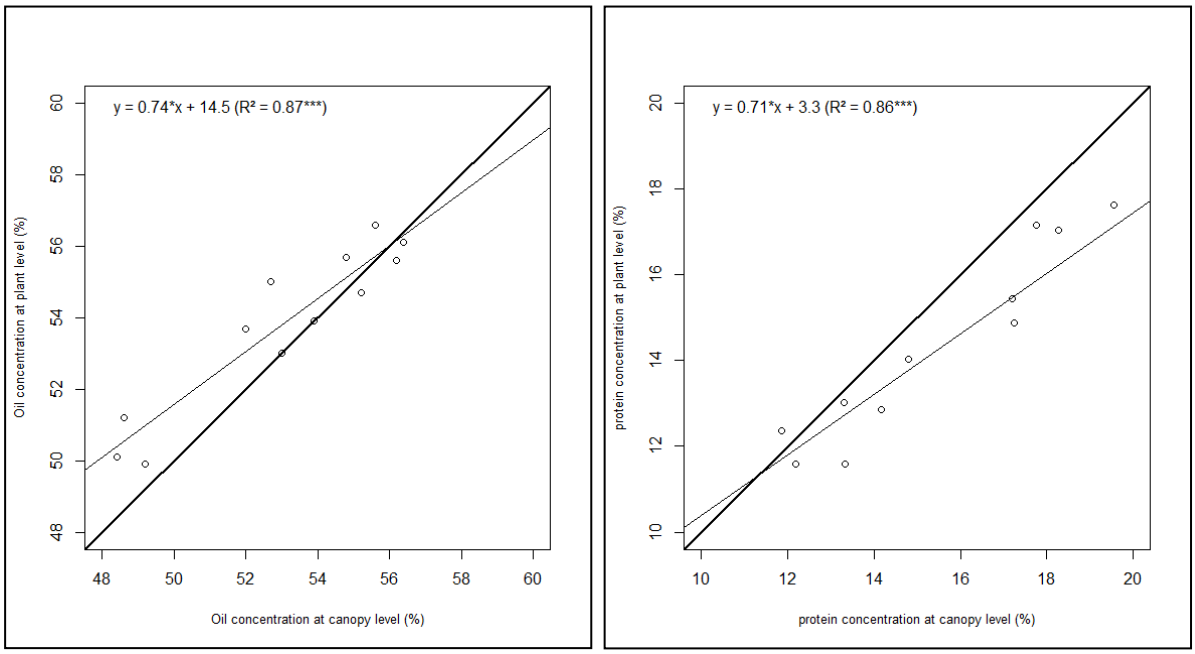


Fig. 5. Comparaison des teneurs en huile et protéines obtenues à l'échelle de la plante entière et du couvert, à la récolte (données Auzeville, 2012, génotype : Kerbel). La première bissectrice et la droite de régression sont représentées. L'équation de la droite de régression est fournie pour chaque relation, ainsi que le R^2.

Des questionnements similaires de changement d'échelle se sont posés concernant la représentation des processus photosynthétiques, d'où le choix *a priori* d'une représentation simplifiée (approche RUE) valable à l'échelle du couvert.

Enfin, la comparaison des deux types de modélisation abordée dans cette thèse permet d'aboutir aux conclusions suivantes :

*La modélisation statistique offre une très bonne qualité de prédiction (2 points d'huile) mais son champ d'application est limité aux sites dans lesquels elle a été paramétrée, ou à des sites avec des conditions pédoclimatiques similaires. Nous avons été un peu plus loin qu'une simple modélisation statistique (empirique), dans le sens où un effort a été mis dans la sélection de prédicteurs ayant un sens biologique. Néanmoins, nous ne serions pas capables d'expliquer toutes les relations « prédicteur-variable expliquée » engendrées par chacun de nos modèles statistiques. Le modèle linéaire multiple pourrait être utilisé par l'agronome souhaitant seulement identifier les variables importantes, tandis que les modèles GAM et arbre hiérarchique (regression tree RT) à partir de relations plus complexes (GAM) ou d'explicitation des lieux d'interaction (RT) pourraient être utilisés par le physiologiste et le modélisateur pour prédire les teneurs en huile et constituer une aide à la décision. Moins lourde à paramétrer, la modélisation statistique prédit bien la tendance générale et quelques IGEC, mais n'a qu'une portée « locale » et ne pourrait être extrapolée hors de son domaine de validité. Une manière d'optimiser la prédiction statistique serait d'identifier et de caractériser de grandes catégories de situations pédoclimatiques et conduites culturales d'intérêt et de paramétrer le modèle statistique pour chaque catégorie.

*La modélisation dynamique (et fonctionnelle), bien que méritant encore quelques ajustements concernant ses paramètres, permet de décortiquer les processus d'élaboration des teneurs en huile et protéines via la description des flux de carbone et d'azote depuis les sources vers les puits. Elle inclut les principaux lieux potentiels d'action (aussi bien spatial que temporel) des facteurs génotypique et environnementaux et leur interaction. Elle serait ainsi plus à même de « capturer » les effets d'interaction des facteurs. De plus, elle s'avère plus souple dans son utilisation comparée à la modélisation statistique ; le diagnostic de l'erreur peut s'effectuer processus par processus, il est ainsi plus facile de formuler de nouvelles hypothèses et de nouvelles pistes d'amélioration. Outre son intérêt heuristique (avancées des connaissances pour le chercheur, l'agronome), le modèle dynamique actuel peut évoluer vers des intérêts pratiques (modélisateur, expert technique, agriculteur) dans une démarche de conseil agronomique (quelle meilleure association génotype-conduite pour optimiser les teneurs en protéines et en huile selon l'objectif de l'agriculteur mais aussi de la coopérative ou de la filière), voire d'accompagnement à l'évaluation variétale (expérimentations virtuelles) dans les réseaux d'inscription et de post-inscription variétale.

Références bibliographiques

Adiredjo, A.L., Navaud, O., Lamaze, T., Grieu, P., 2014. Leaf Carbon Isotope Discrimination as an Accurate Indicator of Water-Use Efficiency in Sunflower Genotypes Subjected to Five Stable Soil Water Contents. Journal of Agronomy and Crop Science 200, 416–424.

Agreste. La statistique, l'évaluation et la prospective agricole [WWW Document], URL. *agreste.agriculture.gouv.fr/ (accessed 08.14)*

Agüera, E., Cabello, P., De La Haba, P., 2010. Induction of leaf senescence by low nitrogen nutrition in sunflower (Helianthus annuus) plants. Physiologia plantarum 138, 256–267.

Agüera, E., Cabello, P., de la Mata, L., Molina, E., de la Haba, P., n.d. Metabolic Regulation of Leaf Senescence in Sunflower (Helianthus annuus L.) Plants. ISBN: 978-953-51-0144-4, InTech, DOI: 10.5772/33671.

Aguirrezábal, L.A., Lavaud, Y., Dosio, G.A., Izquierdo, N.G., Andrade, F.H., González, L.M., 2003. Intercepted solar radiation during seed filling determines sunflower weight per seed and oil concentration. Crop Science 43, 152–161.

Aguirrezábal, L., Martre, P., Pereyra Irujo, G., Izquierdo, N., Allard, V., Press, A., 2009. Management and breeding strategies for the improvement of grain and oil quality. Crop Physiology: Applications for Genetic Improvement and Agronomy.

Alkio, M., Diepenbrock, W., Grimm, E., 2002. Evidence for sectorial photoassimilate supply in the capitulum of sunflower (Helianthus annuus). New phytologist 156, 445–456.

Alkio, M., Grimm, E., 2003. Vascular connections between the receptacle and empty achenes in sunflower (Helianthus annuus L.). Journal of experimental botany 54, 345–348.

Alkio, M., Schubert, A., Diepenbrock, W., Grimm, E., 2003. Effect of source–sink ratio on seed set and filling in sunflower (Helianthus annuus L.). Plant, Cell & Environment 26, 1609–1619.

Anastasi, U., Santonoceto, C., Giuffrè, A.M., Sortino, O., Gresta, F., Abbate, V., 2010. Yield performance and grain lipid composition of standard and oleic sunflower as affected by water supply. Field crops research 119, 145–153.

Andrade, F.H., Ferreiro, M.A., 1996. Reproductive growth of maize, sunflower and soybean at different source levels during grain filling. Field Crops Research 48, 155–165.

Andrianasolo, F.N., Champolivier, L., Maury, P., Debaeke, P., 2012. Plant density contribution to seed oil content the responses of contrasting sunflower genotypes grown in multi-environmental network, in: Proceedings of the 18th International Sunflower Conference. Mar del Plata and Balcarce (Argentina), pp. 724–729.

Andrianasolo, F.N., Casadebaig, P., Maza, E., Champolivier, L., Maury, P., Debaeke, P., 2014. Prediction of sunflower grain oil concentration as a function of variety, crop management and environment using statistical models. European Journal of Agronomy 54, 84–96.

Angeloni, P., Echarte, M.M., Aguirrezábal, L.A.N., 2012. Temperature during grain filling affects grain weight and oil concentration in sunflower hybrid both directly and through the reduction of radiation interception, in: Proceedings of the 18th International Sunflower Conference. Mar del Plata and Balcarce (Argentina), pp. 354–359.

Archontoulis, S.V., Vos, J., Yin, X., Bastiaans, L., Danalatos, N.G., Struik, P.C., 2011. Temporal dynamics of light and nitrogen vertical distributions in canopies of sunflower, kenaf and cynara. Field Crops Research 122, 186–198.

Ayerdi-Gotor, A., Berger, M., Labalette, F., Centis, S., Eychenne, V., Daydé, J., Calmon, A., 2008. Variabilité des teneurs et compositions des composés mineurs dans l'huile de tournesol au cours du développement du capitule. Oléagineux, Corps Gras, Lipides 15, 400–406.

Baenziger, P.S., Clements, R.L., McIntosh, M.S., Yamazaki, W.T., Starling, T.M., Sammons, D.J., Johnson, J.W., 1985. Effect of Cultivar, Environment, and Their Interaction and Stability Analyses on Milling and Baking Quality of Soft Red Winter Wheat1. Crop Sci. 25, 5–8.

Baldocchi, D.D., Amthor, J.S., 2001. Canopy Photosynthesis: History. Terrestrial global productivity 9.

Bannayan, M., Kobayashi, K., Marashi, H., Hoogenboom, G., 2007. Gene-based modelling for rice: An opportunity to enhance the simulation of rice growth and development? Journal of theoretical biology 249, 593–605.

Bauchot, A., Merrien, A., 1988. Teneur en proteines des graines de tournesol et etat proteique foliaire: revue bibliographique. Informations Techniques CETIOM.

Bellocchi, G., Rivington, M., Donatelli, M., Matthews, K., 2010. Validation of biophysical models: issues and methodologies. A review. Agronomy for Sustainable Development 30, 109–130.

Berger, M., Ayerdi-Gotor, A., Sarrafi, A., Maury, P., Daydé, J., Calmon, A., 2010. Compréhension du determinisme de la qualité des huiles du tournesol face aux nouvelles attentes. Oléagineux, Corps Gras, Lipides 17, 171–184.

Bergez, J.-E., Debaeke, P., Deumier, J.-M., Lacroix, B., Leenhardt, D., Leroy, P., Wallach, D., 2001. MODERATO: an object-oriented decision tool for designing maize irrigation schedules. Ecological Modelling 137, 43–60.

Bertheloot, J., Andrieu, B., Fournier, C., Martre, P., 2008. A process-based model to simulate nitrogen distribution in wheat (Triticum aestivum) during grain-filling. Functional Plant Biology 35, 781–796.

Blanchet, R., Merrien, A., Gelfi, N., Rollier, M., 1982. Evolution des biosyntheses au cours du cycle de developpement de tournesol, et repartition des assimilats selon les organes et leurs constituants., in: Proceedings of the International Sunflower Conference.

Blanchet, R., Piquemal, M., Cavalié, G., Hernandez, M., Quinones, H., 1988. Influence de contraintes hydriques sur la répartition des assimilats entre les organes du tournesol, in: Proc 12th Int Sunf Conf Novi-Sad (Yugoslavia). pp. 124–9.

Blum, A., 1998. Improving wheat grain filling under stress by stem reserve mobilisation. Euphytica 100, 77–83.

Bock, B.R., Sikora, F.J., 1990. Modified-quadratic/plateau model for describing plant responses to fertilizer. Soil Science Society of America Journal 54, 1784–1789.

Boote, K.J., Jones, J.W., Pickering, N.B., 1996. Potential Uses and Limitations of Crop Models. Agronomy Journal 88, 704–716.

Boote, K.J., Jones, J.W., Batchelor, W.D., Nafziger, E.D., Myers, O., 2003. Genetic Coefficients in the CROPGRO–Soybean Model. Agronomy Journal 95, 32–51.

Borra, S., Di Ciaccio, A., 2010. Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods. Computational statistics & data analysis 54, 2976–2989.

Borrás, L., Curá, J.A., Otegui, M.E., 2002. Maize kernel composition and post-flowering source-sink ratio. Crop Science 42, 781–790.

Borrás, L., Slafer, G.A., Otegui, M.E., 2004. Seed dry weight response to source–sink manipulations in wheat, maize and soybean: a quantitative reappraisal. Field Crops Research 86, 131–146.

Borredon, M.E., Berger, M., Dauguet, S., Labalette, F., Merrien, A., Mouloungui, Z., Raoul, Y., 2011. Débouchés actuels et futures du tournesol produit en France– Critères de qualité. Innovations Agronomiques 14, 19–38.

Borrell, A., Hammer, G., Oosterom, E., 2001. Stay-green: A consequence of the balance between supply and demand for nitrogen during grain filling? Annals of Applied Biology 138, 91–95.

Borum, J., Murray, L., Michael Kemp, W., 1989. Aspects of nitrogen acquisition and conservation in eelgrass plants. Aquatic Botany 35, 289–300.

Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and regression trees.Belmont, Chapman and Hall (Wadsworth, Inc.), New York, USA.

Brisson, N., Gary, C., Justes, E., Roche, R., Mary, B., Ripoche, D., Zimmer, D., Sierra, J., Bertuzzi, P., Burger, P., 2003. An overview of the crop model STICS. European Journal of agronomy 18, 309–332.

Brun, F., Makowski, D., Wallach, D., Jones, JW.,2013. R package ZeBook. Working with dynamic models for agriculture and environment (v.0.5). http://cran.r-project.org/web/packages/ZeBook/index.html.

Burnham, K.P., Anderson, D.R., 2002. Model selection and multi-model inference: a practical information-theoretic approach. Springer Verlag, New York.

Cabelguenne, M., Debaeke, P., Bouniols, A., 1999. EPICphase, a version of the EPIC model simulating the effects of water and nitrogen stress on biomass and yield, taking account of developmental stages: validation on maize, sunflower, sorghum, soybean and winter wheat. Agricultural Systems 60, 175–196.

Campolongo, F., Cariboni, J., Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models. Environmental modelling & software 22, 1509–1518.

Cantagallo, J.E., Chimenti, C.A., Hall, A.J., 1997. Number of seeds per unit area in sunflower correlates well with a photothermal quotient. Crop Science 37, 1780–1786.

Casadebaig, P., 2008. Analyse et modélisation de l'interaction génotype-environnement-conduite de culture: application au tournesol (Helianthus annuus L.). PhD Thesis, Institut National Polytechnique, Toulouse.

Casadebaig, P., Debaeke, P., Lecoeur, J., 2008. Thresholds for leaf expansion and transpiration response to soil water deficit in a range of sunflower genotypes. European Journal of Agronomy 28, 646–654.

Casadebaig, P., Guilioni, L., Lecoeur, J., Christophe, A., Champolivier, L., Debaeke, P., 2011. SUNFLO, a model to simulate genotype-specific performance of the sunflower crop in contrasting environments. Agricultural and Forest Meteorology 151, 163–178.

Casey, R., De Jong, H., Gouzé, J.-L., 2006. Piecewise-linear models of genetic regulatory networks: Equilibria and their stability. Journal of mathematical biology 52, 27–56.

Cechin, I., de Fátima Fumis, T., 2004. Effect of nitrogen supply on growth and photosynthesis of sunflower plants grown in the greenhouse. Plant Science 166, 1379–1385.

Cechin, I., Rossi, S.C., Oliveira, V.C., Fumis, T.F., 2006. Photosynthetic responses and proline content of mature and young leaves of sunflower plants under water deficit. Photosynthetica 44, 143–146.

Cechin, I., Corniani, N., Fumis, T. de F., Cataneo, A.C., 2010. Differential responses between mature and young leaves of sunflower plants to oxidative stress caused by water deficit. Ciência Rural 40, 1290–1294.

CETIOM - Centre technique des oléagineux [WWW Document], URL *http://www.cetiom.fr/* (accessed 7.12.13).

Champolivier, L., Merrien, A., 1996. Évolution de la teneur en huile et de sa composition en acides gras chez deux variétés de tournesol (oléique ou non) sous l'effet de températures différentes pendant la maturation des graines. OCL. Oléagineux, corps gras, lipides 3, 140–144.

Champolivier, L., Debaeke, P., Thibierge, J., Dejoux, J.F., Ledoux, S., Ludot, M., Berger, F., Casadebaig, P., Jouffret, P., Vogrincic, C., 2011. Construire des stratégies de production adaptées aux débouchés à l'échelle du bassin de collecte. Innovations Agronomiques 14, 39–57.

Chapman, S.C., Hammer, G.L., Meinke, H., 1993. A Sunflower Simulation Model: I. Model Development. Agron. J. 85, 725–735.

Chapman, S.C., 2008. Use of crop models to understand genotype by environment interactions for drought in real-world and simulated plant breeding trials. Euphytica 161, 195–208.

Chaves, M.M., Pereira, J.S., Maroco, J., Rodrigues, M.L., Ricardo, C.P.P., Osório, M.L., Carvalho, I., Faria, T., Pinheiro, C., 2002. How plants cope with water stress in the field? Photosynthesis and growth. Annals of Botany 89, 907–916.

Chen, J.M., Liu, J., Cihlar, J., Goulden, M.L., 1999. Daily canopy photosynthesis model through temporal and spatial scaling for remote sensing applications. Ecological modelling 124, 99–119.

Chervet, B., Vear, F., 1989. Evolution des caractéristiques de la graine et du capitule chez le tournesol au cours de la maturation. Agronomie 9, 305–313.

Chimenti, C.A., Hall, A.J., Sol López, M., 2001. Embryo-growth rate and duration in sunflower as affected by temperature. Field Crops Research 69, 81–88.

Choudhury, B.J., 2001. Modeling radiation-and carbon-use efficiencies of maize, sorghum, and rice. Agricultural and Forest Meteorology 106, 317–330.

Ciampitti, I.A., Vyn, T.J., 2011. A comprehensive study of plant density consequences on nitrogen uptake dynamics of maize plants from vegetative to reproductive stages. Field crops research 121, 2–18.

Connor, D.J., Fereres, E., 1999. A dynamic model of crop growth and partitioning of biomass. Field crops research 63, 139–157.

Connor, D.J., Sadras, V.O., 1992. Physiology of yield expression in sunflower. Field Crops Research 30, 333–389.

Connor, D.J., Hall, A.J., 1997. Sunflower Physiology. in: Sunflower Technology and Production Schneiter A.A (ed), ASA, Madison, Wisconsin, USA,113–182.

Crawley, M.J., 2012. The R Book, 2nd ed. John Wiley & Sons.

Ćupina et al (1992) Influence of N P K nutrition on sunflower plants 3 accumulation and transformation of sugars, proteins and oil in developing seeds.

Danuso, F., Vedove, G., Peressotii, A., 1988. Photosynthetic response of sunflower leaves to PPFD under field conditions, with relation to their age and position, in: International Sunflower Conference. pp. 95–I02.

de la Vega, A.J., Cantore, M.A., Sposaro, M.M., Trápani, N., López Pereira, M., Hall, A.J., 2011. Canopy stay-green and yield in non-stressed sunflower. Field Crops Research 121, 175–185.

de Carvalho Lopes, D., Steidle Neto, A.J., 2011. Simulation models applied to crops with potential for biodiesel production. Computers and electronics in agriculture 75, 1–9.

de Pury, D. de, Farquhar, G.D., 1997. Simple scaling of photosynthesis from leaves to canopies without the errors of big-leaf models. Plant, Cell & Environment 20, 537–557.

de Wit, C. de, 1959. Potential photosynthesis of crop surfaces. Neth. J. Agric. Sci 7, 141–9.

de Wit, C.T., 1966. Photosynthesis of leaf canopies. Centre for Agricultural Publications and Documentation.

de Wit, C. de, Setlík, I., 1970. Dynamic concepts in biology., in: Prediction and Measurement of Photosynthetic Productivity. Proceedings of the IBP/PP Technical Meeting, Tfebon,[Czechoslovakia], 14-21 September, 1969. Wageningen, Netherlands, PUDO, pp. 12–23.

Debaeke, P., Mailhol, J.-C., Bergez, J.-E., 2006. Adaptations agronomiques à la sécheresse. Systèmes de grande culture, in: Sécheresse et agriculture. Réduire la vulnérabilité de l'agriculture à un risque accru de manque d'eau. INRA, France, pp. 258–360.

Debaeke, P., Casadebaig, P., Haquin, B., Mestries, E., Palleau, J.-P., Salvi, F., 2010. Simulation de la réponse variétale du tournesol à l'environnement à l'aide du modèle SUNFLO. Oléagineux, Corps Gras, Lipides 17, 143–151.

Debaeke, P., van Oosterom, E.J., Justes, E., Champolivier, L., Merrien, A., Aguirrezabal, L.A.N., González-Dugo, V., Massignam, A.M., Montemurro, F., 2012. A species-specific critical nitrogen dilution curve for sunflower (Helianthus annuus L.). Field Crops Research 136, 76–84.

Debaeke P., Mestries E., Desanlis M., Seassau C., 2014. Effects of crop management on the incidence and severity of fungal diseases in sunflower. In: Arribas JE (ed) Sunflowers: Growth and Development, Environmental Influences and Pests/Diseases, Nova Science Pubs., New York, USA, pp 201-226.

Denis, L., Vear, F., 1994. Environmental effects on hullability of sunflower hybrids. Agronomie, 589–597.

Denis, L., Vear, F., 1996. Variation of hullability and other seed characteristics among sunflower lines and hybrids. Euphytica 87, 177–187.

Diepenbrock, W., Long, M., Feil, B., 2001. Yield and quality of sunflower as affected by row orientation, row spacing and plant density. Bodenkultur-Wien and Munchen 52, 29–36.

Dordas, C., 2012. Variation in dry matter and nitrogen accumulation and remobilization in barley as affected by fertilization, cultivar, and source–sink relations. European Journal of Agronomy 37, 31–42. doi:10.1016/j.eja.2011.10.002

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36, 027–046.

Dosio, G.A., Andrade, F.H., Pereyra, V., 2000. Solar radiation intercepted during seed filling and oil production in two sunflower hybrids.

Dosio, G.A., Rey, H., Lecoeur, J., Izquierdo, N.G., Aguirrezábal, L.A., Tardieu, F., Turc, O., 2003. A whole-plant analysis of the dynamics of expansion of individual leaves of two sunflower hybrids. Journal of Experimental Botany 54, 2541–2552.

Dreccer, M.F., Grashoff, C., Rabbinge, R., 1997. Source-sink ratio in barley (*Hordeum vulgare* L.) during grain filling: effects on senescence and grain protein concentration. Field Crops Research 49, 269–277.

Drouet, J.-L., Pagès, L., 2007. GRAAL-CN: A model of GRowth, Architecture and ALlocation for Carbon and Nitrogen dynamics within whole plants formalised at the organ level. Ecological Modelling 206, 231–249.

Dumas, J.B.A., 1831. Procedes de I'analyse organique. Ann. Chim. Phys 47, 198–205.

Ebrahimi, A., Maury, P., Berger, M., Poormohammad Kiani, S., Nabipour, A., Shariati, F., Grieu, P., Sarrafi, A., 2008. QTL mapping of seed-quality traits in sunflower recombinant inbred lines under different water regimes. Genome 51, 599–615.

Echarte, M.M., Pereyra-Irujo, P.-I., Covi, M., Izquierdo, N.G., Aguirrezábal, L.A.N., 2010. Producing better sunflower oils in a changing environment, in: Advances in Fats and Oil Research, Transworld Research Network. Mabel Cristina Tomás, Argentina, pp. 1–23.

Efron, B., Tibshirani, R., 1997. Improvements on cross-validation: the 632+ bootstrap method. Journal of the American Statistical Association 92, 548–560.

Ercoli, L., Lulli, L., Mariotti, M., Masoni, A., Arduini, I., 2008. Post-anthesis dry matter and nitrogen dynamics in durum wheat as affected by nitrogen supply and soil water availability. European Journal of Agronomy 28, 138–147.

FAO. Food and Agriculture Organization of the United Nations [WWW Document], URL http://www.fao.org/home/en/ (accessed 7.12.13).

Farquhar, G.D., von Caemmerer, S. von, Berry, J.A., 1980. A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species. Planta 149, 78–90.

Ferraro, D.O., Rivero, D.E., Ghersa, C.M., 2009. An analysis of the factors that influence sugarcane yield in Northern Argentina using classification and regression trees. Field Crops Research 112, 149–157.

Ferreira, A.M., Abreu, F.G., 2001. Description of development, light interception and growth of sunflower at two sowing dates and two densities. Mathematics and Computers in simulation 56, 369–384.

Fick, G.N., 1975. Heritability of oil content in sunflowers. Crop Science 15, 77–78.

Fick, G.N., 1978. Sunflower breeding and genetics. In J.F. Carter (ed.) Sunflower science and technology. Agron. Monogr. 19. ASA, CSSA, and SSSA, Madison, WI, 279-337

Fick, G.N., Miller, J.F., 1997. Sunflower Breeding. in:Sunflower Technology and Production. Schneiter A.A (ed), ASA, Madison, Wisconsin, USA, 395–439.

Flexas, J., Medrano, H., 2002. Drought-inhibition of photosynthesis in C3 plants: stomatal and non-stomatal limitations revisited. Annals of Botany 89, 183–189.

Fushiki, T., 2011. Estimation of prediction error by using K-fold cross-validation. Statistics and Computing 21, 137–146.

Gallais, A., 1992. Bases génétiques et stratégie de sélection de l'adaptation générale. Le Sélectionneur Français 42, 59-78.

Gabrielle, B., Denoroy, P., Gosse, G., Justes, E., Andersen, M.N., 1998. Development and evaluation of a CERES-type model for winter oilseed rape. Field Crops Research 57, 95–111. doi:10.1016/S0378-4290(97)00120-2

Goffner, D., Cazalis, R., du Sert, C.P., Calmes, J., Cavalie, G., 1988. 14C photoassimilate partitioning in developing sunflower seeds. Journal of experimental botany 39, 1411–1420.

Goldberg, R.A., 1968. Agribusiness Coordination: a systems approach to the wheat, soybean, and Florida orange economies. Agribusiness Coordination: a systems approach to the wheat, soybean, and Florida orange economies.

Grieu, P., Maury, P., Debaeke, P., Sarrafi, A., 2008. Améliorer la tolérance à la sécheresse du tournesol: apports de l'écophysiologie et de la génétique. Revue Innovations Agronomiques 2, 37–51.

Grimm, S.S., Jones, J.W., Boote, K.J., Herzog, D.C., 1994. Modeling the occurrence of reproductive stages after flowering for four soybean cultivars. Agronomy Journal 86, 31–38.

Grömping, U., 2006. Relative importance for linear regression in R: the package relaimpo. Journal of Statistical Software 17, 1–27.

Grossman, Y.L., DeJong, T.M., 1994. PEACH: a simulation model of reproductive and vegetative growth in peach trees. Tree Physiology 14, 329–345.

Gubbels, G.H., Dedio, W., 1986. Effect of plant density and soil fertility on oilseed sunflower genotypes. Canadian journal of plant science 66, 521–527.

Hall, A.J., Connor, D.J., Whitfield, D.M., 1989. Contribution of pre-anthesis assimilates to grain-filling in irrigated and water-stressed sunflower crops I. Estimates using labelled carbon. Field Crops Research 20, 95–112.

Hall, A.J., Whitfield, D.M., Connor, D.J., 1990. Contribution of pre-anthesis assimilates to grain-filling in irrigated and water-stressed sunflower crops II. Estimates from a carbon budget. Field Crops Research 24, 273–294.

Hall, A.J., Connor, D.J., Sadras, V.O., 1995. Radiation-use efficiency of sunflower crops: effects of specific leaf nitrogen and ontogeny. Field Crops Research 41, 65–77.

Hammadeh, I., Maury, P., Debaeke, P., Lecoeur, J., Nouri, L., Kiani, S.P., Grieu, P., 2005. Canopy nitrogen distribution and photosynthesis during grain filling in irrigated and water stressed sunflower genotypes. Interdrought-II, in: The Second International Conference on Integrated Approaches to Sustain and Improve Plant Production under Drought Stress. pp. 24–28.

Hammer, G., Cooper, M., Tardieu, F., Welch, S., Walsh, B., van Eeuwijk, F., Chapman, S., Podlich, D., 2006. Models for navigating biological complexity in breeding improved crop plants. Trends in plant science 11, 587–593.

Hawkins, D.M., Basak, S.C., Mills, D., 2003. Assessing model fit by cross-validation. Journal of chemical information and computer sciences 43, 579–586.

Hocking, P.J., Steer, B.T., 1983. Distribution of nitrogen during growth of sunflower (Helianthus annuus L.). Annals of Botany 51, 787–799.

Horie, T., 1977. Simulation of sunflower growth. I. Formulation and parametrization of dry matter production, leaf photosynthesis, respiration and partitioning of photosynthates. Bulletin. Series A (Physics and statistics).

Hsiao, T.C., 1973. Plant responses to water stress. Annual review of plant physiology 24, 519–570.

Iglesias, D.J., Lliso, I., Tadeo, F.R., Talon, M., 2002. Regulation of photosynthesis through source: sink imbalance in citrus is mediated by carbohydrate content in leaves. Physiologia Plantarum 116, 563–572.

Izaurralde, R.C., Williams, J.R., McGill, W.B., Rosenberg, N.J., Jakas, M.C.Q., 2006. Simulating soil C dynamics with EPIC: Model description and testing against long-term data. Ecological Modelling 192, 362–384. doi:10.1016/j.ecolmodel.2005.07.010

Izquierdo, N.G., Dosio, G.A.A., Cantarero, M., Luján, J., Aguirrezábal, L.A.N., 2008. Weight per Grain, Oil Concentration, and Solar Radiation Intercepted during Grain Filling in Black Hull and Striped Hull Sunflower Hybrids. Crop Science 48, 688-699.

Jamieson, P.D., Semenov, M.A., 2000. Modelling nitrogen uptake and redistribution in wheat. Field Crops Research 68, 21–29.

Jeuffroy, M.H., Ney, B., Ourry, A., 2002. Integrated physiological and agronomic modelling of N capture and use within the plant. Journal of Experimental Botany 53, 809–823.

Jeuffroy, M.-H., Valantin-Morison, M., Champolivier, L., Reau, R., 2006. Azote, rendement et qualité des graines: mise au point et utilisation du modèle Azodyn-colza pour améliorer les performances du colza vis-à-vis de l'azote. Oléagineux, Corps Gras, Lipides 13, 388–392.

Jiang, W., Simon, R., 2007. A comparison of bootstrap methods and an adjusted bootstrap

approach for estimating the prediction error in microarray classification. Statistics in medicine 26, 5320–5334.

Jones, C.A., Kiniry, J.R., Dyke, P.T., 1986. CERES-Maize: A simulation model of maize growth and development.

Jouffret, P., Labalette, F., Thibierge, J., 2011. Atouts et besoins en innovations du tournesol pour une agriculture durable. Innovations agronomiques 14, 1–17.

Kang, M., Yang, L., Zhang, B., De Reffye, P., 2011. Correlation between dynamic tomato fruit-set and source–sink ratio: a common relationship for different plant densities and seasons? Annals of botany 107, 805–815.

Keong, Y.K., Keng, W.M., 2012. Statistical Modeling of Weather-based Yield Forecasting for Young Mature Oil Palm. APCBEE Procedia 4, 58–65.

Khamis, A., Ismail, Z., Haron, K., Mohammed, A.T., 2006. Modeling oil palm yield using multiple linear regresion and robust M- regression. Journal of Agronomy 5, 32–36.

Kiani, S.P., Grieu, P., Maury, P., Hewezi, T., Gentzbittel, L., Sarrafi, A., 2007a. Genetic variability for physiological traits under drought conditions and differential expression of water stress-associated genes in sunflower (Helianthus annuus L.). Theoretical and applied genetics 114, 193–207.

Kiani, S.P., Talia, P., Maury, P., Grieu, P., Heinz, R., Perrault, A., Nishinakamasu, V., Hopp, E., Gentzbittel, L., Paniego, N., 2007b. Genetic analysis of plant water status and osmotic adjustment in recombinant inbred lines of sunflower under two water treatments. Plant Science 172, 773–787.

Kiniry, J.R., Blanchet, R., Williams, J.R., Texier, V., Jones, C.A., Cabelguenne, M., 1992. Sunflower simulation using the EPIC and ALMANAC models. Field Crops Research 30, 403–423.

Knowles, P.F., 1978. Morphology and anatomy. Sunflower science and technology 55–87.

Kobayashi, K., Salam, M.U., 2000. Comparing simulated and measured values using mean squared deviation and its components. Agronomy Journal 92, 345–352.

Kutner, M.H., Nachtsheim, C., Neter, J., 2004. Applied Linear Regression Models. McGraw-Hill/Irwin 5th Ed.

Landau, S., Mitchell, R.A.C., Barnett, V., Colls, J.J., Craigon, J., Payne, R.W., 2000. A parsimonious, multiple-regression model of wheat yield response to environment. Agricultural and Forest Meteorology 101, 151–166.

Lawlor, D.W., 2002. Limitation to Photosynthesis in Water-stressed Leaves: Stomata vs. Metabolism and the Role of ATP. Annals of botany 89, 871–885.

Léchaudel, M., Génard, M., Lescourret, F., Urban, L., Jannoyer, M., 2005. Modeling effects of weather and source–sink relationships on mango fruit growth. Tree physiology 25, 583–597.

Lee, A., Robertson, B., 2012. R330 package. URL http://cran.r-project.org/web/packages/R330/R330.pdf

Lescourret, F., Ben Mimoun, M., Génard, M., 1998. A simulation model of growth at the shoot-bearing fruit level: I. Description and parameterization for peach. European Journal of Agronomy 9, 173–188.

Li, W., Zhou, Z., Meng, Y., Xu, N., Fok, M., 2009. Modeling boll maturation period, seed growth, protein, and oil content of cotton (*Gossypium hirsutum* L.) in China. Field crops research 112, 131–140.

Lindeman, R.H., Merenda, P.F., Gold, R.Z., 1980. Introduction to bivariate and multivariate analysis. Scott, Foresman Glenview, IL.

Lindström, L.I., Pellegrini, C.N., Aguirrezábal, L.A.N., Hernández, L.F., 2006. Growth and development of sunflower fruits under shade during pre and early post-anthesis period. Field Crops Research 96, 151–159.

Lindström, L.I., Pellegrini, C.N., Hernández, L.F., 2007. Histological development of the sunflower fruit pericarp as affected by pre-and early post-anthesis canopy shading. Field crops research 103, 229–238.

Lobell, D.B., Ortiz-Monasterio, J.I., Asner, G.P., Naylor, R.L., Falcon, W.P., 2005. Combining field surveys, remote sensing, and regression trees to understand yield variations in an irrigated wheat landscape. Agronomy Journal 97, 241–249.

Loomis, R.S., Amthor, J.S., 1999. Yield potential, plant assimilatory capacity, and metabolic efficiencies.

López Pereira, M., Trapani, N., Sadras, V.O., 2000. Genetic improvement of sunflower in Argentina between 1930 and 1995: Part III. Dry matter partitioning and grain composition. Field Crops Research 67, 215–221.

López Pereira, M., Berney, A., Hall, A.J., Trápani, N., 2008. Contribution of pre-anthesis photoassimilates to grain yield: Its relationship with yield in Argentine sunflower cultivars released between 1930 and 1995. Field Crops Research 105, 88–96.

Maindonald, J., Braun, W.J., 2010. Data analysis and graphics using R: an example-based approach. Cambridge University Press.

Makowski, D., Hillier, J., Wallach, D., Andrieu, B., Jeuffroy, M.H., 2006. Parameter estimation for crop models. Working with dynamic crop models: Evaluation, analysis, parameterization, and application. Elsevier, Amsterdam 101–103.

Mantese, A.I., Medan, D., Hall, A.J., 2006. Achene structure, development and lipid accumulation in sunflower cultivars differing in oil content at maturity. Annals of botany 97, 999–1010.

Marchand, G., Mayjonade, B., Varès, D., Blanchet, N., Boniface, M.-C., Maury, P., Andrianasolo, F.N., Burger, P., Debaeke, P., Casadebaig, P., 2013. A biomarker based on gene expression indicates plant water status in controlled and natural environments. Plant, cell & environment 36, 2175–2189.

Marra, G., Wood, S.N., 2011. Practical variable selection for generalized additive models. Computational Statistics & Data Analysis 55, 2372–2387.

Martre, P., Porter, J.R., Jamieson, P.D., Triboï, E., 2003. Modeling grain nitrogen accumulation and protein composition to understand the sink/source regulations of nitrogen remobilization for wheat. Plant Physiology 133, 1959–1967.

Massignam, A.M., Chapman, S.C., Hammer, G.L., Fukai, S., 2009. Physiological determinants of maize and sunflower grain yield as affected by nitrogen supply. Field Crops Research 113, 256–267.

Maury, P., Mojayad, F., Berger, M., Planchon, C., 1996. Photochemical response to drought acclimation in two sunflower genotypes. Physiologia Plantarum 98, 57–66.

Maury P. 1997. Adaptation a la sécheresse et photosynthèse chez le tournesol (*Helianthus annuus* L.). PhD thesis.

Maury, P., Berger, M., Mojayad, F., Planchon, C., 2000. Leaf water characteristics and drought acclimation in sunflower genotypes. Plant and Soil 223, 155–162.

Meinke, H., Hammer, G.L., Chapman, S.C., 1993. A sunflower simulation model: II. Simulating production risks in a variable sub-tropical environment. Agronomy Journal 85, 735–742.

Merrien, A., Quinsac, A., Maisonneuve, C., 1988. Variabilité de la teneur en protéines des graines de tournesol en relation avec l'état protéinique foliaire.

Merrien, A., 1992. Physiologie du tournesol. 65 p. Centre Technique Interprofessionnel des Oléagineux Métropolitain (CETIOM), Paris, France.

Monteith, J.L., Moss, C.J., 1977. Climate and the efficiency of crop production in Britain [and discussion]. Philosophical Transactions of the Royal Society of London. B, Biological Sciences 281, 277–294.

Morot-Gaudry, J.-F., 2001. Nitrogen assimilation by plants: physiological, biochemical and molecular aspects. Science Publishers, Inc.

Morot-Gaudry, J.-F., Moreau, F., Prat, R., Maurel, C., Sentenac, H., 2012. Biologie végétale: Nutrition et métabolisme-2e édition. Dunod.

Morris, M.D., 1991. Factorial sampling plans for preliminary computational experiments. Technometrics 33, 161–174.

Moschen, S., Bengoa Luoni, S., Paniego, N.B., Hopp, H.E., Dosio, G.A.A., Fernandez, P., Heinz, R.A., 2014. Identification of Candidate Genes Associated with Leaf Senescence in Cultivated Sunflower (Helianthus annuus L.). PLoS ONE 9, e104379. doi:10.1371/journal.pone.0104379

Nickerson, D.M., Facey, D.E., Grossman, G.D., 1989. Estimating physiological thresholds with continuous two-phase regression. Physiol. Zool 62, 866–887.

Noodén, L.D., 2012. Senescence and aging in plants. Elsevier.

OIL WORLD. The Independent Forecasting Service for Oilseeds, Oils and Meals Providing Primary Information- Professional Analysis- Unbiased Opinion.[WWW Document], URL *http://www.oilworld.de/* (accessed 08/14).

Palosuo, T., Kersebaum, K.C., Angulo, C., Hlavinka, P., Moriondo, M., Olesen, J.E., Patil, R.H., Ruget, F., Rumbaur, C., Takáč, J., 2011. Simulation of winter wheat yield and its variability in different climates of Europe: A comparison of eight crop growth models. European Journal of Agronomy 35, 103–114.

Pan, J., Zhu, Y., Jiang, D., Dai, T., Li, Y., Cao, W., 2006. Modeling plant nitrogen uptake and grain nitrogen accumulation in wheat. Field crops research 97, 322–336.

Pan, J., Zhu, Y., Cao, W., 2007. Modeling plant carbon flow and grain starch accumulation in wheat. Field crops research 101, 276–284.

Pantin, F., Simonneau, T., Muller, B., 2012. Coming of leaf age: control of growth by hydraulics and metabolics during leaf ontogeny. New Phytologist 196, 349–366.

Paul, M.J., Driscoll, S.P., 1997. Sugar repression of photosynthesis: the role of carbohydrates in signalling nitrogen deficiency through source: sink imbalance. Plant, Cell & Environment 20, 110–116.

Penhale, P.A., Thayer, G.W., 1980. Uptake and transfer of carbon and phosphorus by eelgrass (*Zostera marina* L.) and its epiphytes. Journal of Experimental Marine Biology and Ecology 42, 113–123.

Pereyra-Irujo, G.A., Aguirrezábal, L.A.N., 2007. Sunflower yield and oil quality interactions and variability: Analysis through a simple simulation model. Agricultural and Forest Meteorology 143, 252–265.

Peyronnet, C., Lacampagne, J.-P., Le Cadre, P., Pressenda, F., 2014. Les sources de protéines dans l'alimentation du bétail en France: la place des oléoprotéagineux. OCL 21, D402.

Picq, G., Abramovsky, P., 1989. Indicateurs et conditions de croissance associes a la teneur et au rendement en huile et en proteines des akenes de tournesol (Helianthus annuus). Inf. Tech. CETIOM 108, 18–29.

Pilorgé, É., 2010. Nouveau contexte environnemental et réglementaire: quel impact pour la culture du tournesol? Oléagineux, Corps Gras, Lipides 17, 136–138.

Ploschuk, E.L., Hall, A.J., 1997. Maintenance respiration coefficient for sunflower grains is less than that for the entire capitulum. Field crops research 49, 147–157.

Pommel, B., Gallais, A., Coque, M., Quillere, I., Hirel, B., Prioul, J.L., Andrieu, B., Floriot, M., 2006. Carbon and nitrogen allocation and grain filling in three maize hybrids differing in leaf senescence. European Journal of Agronomy 24, 203–211.

Prado, P., Collier, C.J., Lavery, P.S., 2008. C-13 and N-15 translocation within and among shoots in two Posidonia species from Western Australia. Marine Ecology Progress Series 361, 69–82.

PROLEA. La filière française des huiles et protéines végétales[WWW Document], URL http://www.prolea.com/ (accessed 08/14).

Prost, L., Makowski, D., Jeuffroy, M.-H., 2008. Comparison of stepwise selection and Bayesian model averaging for yield gap analysis. Ecological Modelling 219, 66–76.

Pujol, G., Iooss, B., Janon, A., 2014 R package Sensitivity. Sensitivity analysis (v.1.9). http://cran.r-project.org/web/packages/sensitivity/index.html.Pury, D. de, Farquhar, G.D., 1997. Simple scaling of photosynthesis from leaves to canopies without the errors of big-leaf models. Plant, Cell & Environment 20, 537–557.

Quere, L., 2004. Des facteurs clés limitants pour le tournesol identifiés en 2003. Oléoscope 31–32.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rao, R.B., Fung, G., Rosales, R., 2008. On the Dangers of Cross-Validation. An Experimental Evaluation, in: SDM. pp. 588–596.

Razi, M., Athappilly, K., 2005. A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models. Expert Systems with Applications 29, 65–74.

Rizzardi, M.A., da Silva, P.R.F., da Rocha, A.B., 1992. Dry matter and oil partitioning in sunflower achenes as a function of cultivar and plant density, in: Proceedings of the 13th International Sunflower Conference, Pisa (Italy), pp. 7–11.

Rengel, D., Arribat, S., Maury, P., Martin-Magniette, M.-L., Hourlier, T., Laporte, M., Varès, D., Carrère, S., Grieu, P., Balzergue, S., 2012. A gene-phenotype network based on genetic variability for drought responses reveals key physiological processes in controlled and natural environments. PloS one 7, e45249.

Rinaldi, M., Losavio, N., Flagella, Z., 2003. Evaluation and application of the OILCROP–SUN model for sunflower in southern Italy. Agricultural Systems 78, 17–30.

Rizzardi, M.A., da Silva, P.R.F., da Rocha, A.B., 1992. Dry matter and oil partitioning in sunflower achenes as a function of cultivar and plant density, in: Proceedings of the 13th International Sunflower Conference, Pisa, Italy. pp. 7–11.

Roche, J., 2005. Composition de la graine de tournesol (Helianthus annuus L.) sous l'effet conjugué des contraintes agri-environnementales et des potentiels varietaux. PhD Thesis, Institut National Polytechnique, Toulouse.

Rondanini, D., Savin, R., Hall, A.J., 2003. Dynamics of fruit growth and oil quality of sunflower (Helianthus annuus L.) exposed to brief intervals of high temperature during grain filling. Field Crops Research 83, 79–90.

Ruiz, R.A., Maddonni, G.A., 2006. Sunflower Seed Weight and Oil Concentration under Different Post-Flowering Source-Sink Ratios. Crop Science 46, 671-680.

Sadras, V.O., Connor, D.J., Whitfield, D.M., 1993. Yield, yield components and source-sink relationships in water-stressed sunflower. Field Crops Research 31, 27–39.

Saltelli, A., Chan, K., Scott, E.M., 2000. Sensitivity analysis. Wiley New York.

Santonoceto, C., Anastasi,U., Riggi, E., Abbate, V.,2003. Accumulation dynamics of dry matter, oil and major fatty acids in sunflower seeds in relation to genotype and water regime. Ital. J. Agron. 7, 3–14.

Schmidt, M., Lipson, H., 2009. Distilling Free-Form Natural Laws from Experimental Data. Science 324, 81–85.

Schmidt, M., Lipson, H., 2013. Eureqa (Version 0.98 beta) [Software]. Available from http://www.eureqa/com/ / (accessed 7.12.13).

Schneiter, A.A., Miller, J.F., 1981. Description of sunflower growth stages. Crop Science 21, 901–903.

Schnyder, H., 1993. The role of carbohydrate storage and redistribution in the source-sink relations of wheat and barley during grain filling—a review. New Phytologist 123, 233–245.

Seassau, C., 2010. Etiologie du syndrome de dessèchement précoce du tournesol: implication de Phoma macdonaldii et interaction avec la conduite de culture. INPT.

Shabana, R., 1974. Genetic variability of sunflower varieties and inbred lines, in: Proc. of the 6th Int. Sunflower Conf., Bucharest, Romania. pp. 263–269.

Shatar, T.M., McBratney, A.B., 1999. Empirical modeling of relationships between sorghum yield and soil properties. Precision Agriculture 1, 249–276.

Sinclair, T.R., Goudriaan, J., De Wit, C.T., 1977. Mesophyll resistance and CO2 compensation concentration in leaf photosynthesis models. Photosynthetica.

Sinclair, T.R., Horie, T., 1989. Leaf nitrogen, photosynthesis, and crop radiation use efficiency: a review. Crop science 29, 90–98.

Sinclair, T.R., 2005. Theoretical analysis of soil and plant traits influencing daily plant water flux on drying soils. Agronomy journal 97, 1148–1152.

Sobol′, I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and computers in simulation 55, 271–280.

Steer, B.T., Hocking, P.J., 1984. Nitrogen nutrition of sunflower (*Helianthus annuus* L.): Acquisition and partitioning of dry matter and nitrogen by vegetative organs and their relationship to seed yield. Field Crops Research 9, 237–251.

Steer, B.T., Hocking, P.J., Kortt, A.A., Roxburgh, C.M., 1984. Nitrogen nutrition of sunflower (Helianthus annuus L.): Yield components, the timing of their establishment and seed characteristics in response to nitrogen supply. Field Crops Research 9, 219–236.

Steer, B.T., Milroy, S.P., Kamona, R.M., 1993. A model to simulate the development, growth and yield of irrigated sunflower. Field Crops Research 32, 83–99.

Stöckle, C.O., Kjelgaard, J., Bellocchi, G., 2004. Evaluation of estimated weather data for calculating Penman-Monteith reference crop evapotranspiration. Irrigation science 23, 39–46.

Stoyanova, J., Ivanov, P., 1975. Inheritance of oil and protein content in first hybrid progeny of sunflower. Rasteniev" dni Nauki (Bulgaria).

Tabourel-Tayot, F., Gastal, F., 1998a. MecaNiCAL, a supply–demand model of carbon and nitrogen partitioning applied to defoliated grass: 1. Model description and analysis. European Journal of Agronomy 9, 223–241.

Tabourel-Tayot, F., Gastal, F., 1998b. MecaNiCAL, a supply–demand model of carbon and nitrogen partitioning applied to defoliated grass: 2. Parameter estimation and model evaluation. European journal of agronomy 9, 243–258.

Tanaka, W., Maddonni, G.Á., 2009. Maize kernel oil and episodes of shading during the grain-filling period. Crop Science 49, 2187–2197.

Tardieu, F.C., Durant, P., Triboï, J., Zivy, E., n.d. M. 2006. Perception de la sécheresse par la plante. Conséquences sur la productivité et sur la qualité des produits récoltés. J, Amigues. P, Debaeke. G, Lemaire. B, Seguin. F, Tardieu. A, Thomas (eds). Sécheresse et agriculture. Réduire la vulnérabilité de l'agriculture à un risque accru de manque d'eau. Expertise scientifique collective, Rapport, INRA, France 49–67.

Thomas, H., 2013. Senescence, ageing and death of the whole plant. New Phytologist 197, 696–711.

Thomas, H., Ougham, H.J., Wagstaff, C., Stead, A.D., 2003. Defining senescence and death. Journal of Experimental Botany 54, 1127–1132. doi:10.1093/jxb/erg133

Thornley, J.H.M., 1972. A balanced quantitative model for root: shoot ratios in vegetative plants. Annals of Botany 36, 431–441.

Thornley, J.H., Johnson, I.R., 1990. Plant and crop modelling. A mathematical approach to plant and crop physiology. Clarendon Press.

Tittonell, P., Shepherd, K., Vanlauwe, B., Giller, K., 2008. Unravelling the effects of soil and crop management on maize productivity in smallholder agricultural systems of western Kenya—An application of classification and regression tree analysis. Agriculture, Ecosystems & Environment 123, 137–150.

Toms, J.D., Lesperance, M.L., 2003. Piecewise regression: a tool for identifying ecological thresholds. Ecology 84, 2034–2041.

Triki, S., Ben Hamida, J., Mazliak, P., 1997. Study of the metabolism of lipid reserves in ripening sunflower seeds: a tracing experiment with (1-14C) acetate. Oleagineux Corps Gras Lipides (France).

Tulbure, M.G., Wimberly, M.C., Boe, A., Owens, V.N., 2012. Climatic and genetic controls of yields of switchgrass, a model bioenergy species. Agriculture, Ecosystems & Environment 146, 121–129.

Uhart, S.A., Andrade, F.H., 1995. Nitrogen and carbon accumulation and remobilization during grain filling in maize under different source/sink ratios. Crop Science 35, 183–190.

USDA – US Department of Agriculture [WWW Document], URL http://www.usda.gov/ (accessed 08.14).

Utz, H.F., Melchinger, A.E., Schön, C.C., 2000. Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. Genetics 154, 1839–1849.

van Oosterom, E.J., Chapman, S.C., Borrell, A.K., Broad, I.J., Hammer, G.L., 2010. Functional dynamics of the nitrogen balance of sorghum. II. Grain filling period. Field Crops Research 115, 29–38.

Vear, F., Bony, H., Joubert, G., Tourvieille de Labrouhe, D.T., Pauchet, I., Pinochet, X., 2003. 30 years of sunflower breeding in France. Oléagineux, Corps Gras, Lipides 10, 66–73.

Vega, C.R., Andrade, F.H., Sadras, V.O., 2001. Reproductive partitioning and seed set efficiency in soybean, sunflower and maize. Field Crops Research 72, 163–175.

Valentinuz, O.R., Tollenaar, M., 2004. Vertical profile of leaf senescence during the grain-filling period in older and newer maize hybrids. Crop Science 44, 827–834.

Villalobos, F.J., Sadras, V.O., Soriano, A., Fereres, E., 1994. Planting density effects on dry matter partitioning and productivity of sunflower hybrids. Field crops research 36, 1–11.

Villalobos, F.J., Hall, A.J., Ritchie, J.T., Orgaz, F., 1996. OILCROP-SUN: A Development, Growth, and Yield Model of the Sunflower Crop. Agron. J. 88, 403–415.

Wallach, D., Makowski, D., Jones, J., 2006. Working with Dynamic Crop Models: Evaluation. Analysis, Parameterization and Applications. Elsevier BV.

Wallach, D., Makowski, D., Jones, J. W., Brun, F., 2013. Working with Dynamic Crop Models, 2nd Edition Methods, Tools and Examples for Agriculture and Environment. Academic Press.

Whittingham, M.J., Stephens, P.A., Bradbury, R.B., Freckleton, R.P., 2006. Why do we still use stepwise modelling in ecology and behavior? Journal of Animal Ecology 75, 1182–1189.

Williams, J.R., 1990. The Erosion-Productivity Impact Calculator (EPIC) Model: A Case History. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 329, 421–428. doi:10.1098/rstb.1990.0184

Wood, S.N., 2003. Thin plate regression splines. Journal of the Royal Statistical Society. Series B, statistical methodology 65, 95–114.

Wood, S.N., 2004. Stable and efficient multiple smoothing parameter estimation for generalized additive models. Journal of the American Statistical Association 99.

Wullschleger, S.D., Davis, E.B., Borsuk, M.E., Gunderson, C.A., Lynd, L.R., 2010. Biomass Production in Switchgrass across the United States: Database Description and Determinants of Yield. Agronomy Journal 102, 1158-1168.

Yegappan, T.M., Paton, D.M., Gates, C.T., Müller, W.J., 1982. Water Stress in Sunflower (Helianthus annuus L.) 2. Effects on Leaf Cells and Leaf Area. Annals of Botany 49, 63–68.

Yin, X., Guo, W., Spiertz, J.H., 2009. A quantitative approach to characterize sink–source relationships during grain filling in contrasting wheat genotypes. Field Crops Research 114, 119–126.

Zheljazkov, V.D., Vick, B.A., Baldwin, B.S., Buehring, N., Astatkie, T., Johnson, B., 2009. Oil content and saturated fatty acids in sunflower as a function of planting date, nitrogen rate, and hybrid. Agronomy journal 101, 1003–1011.

Zheng, H., Chen, L., Han, X., Zhao, X., Ma, Y., 2009. Classification and regression tree (CART) for analysis of soybean yield variability among fields in Northeast China: The importance of phosphorus application rates under drought conditions. Agriculture, Ecosystems & Environment 132, 98–105.

Zuur, A.F., Ieno, E.N., Elphick, C.S., 2010. A protocol for data exploration to avoid common statistical problems. Methods in Ecology and Evolution 1, 3–14.

Annexe

Table.1. Results of sensitivity analysis following Morris screening method (chapter IV) for parameters of the dynamic source-sink model. 4 outputs (oil and protein weights per m² and oil and protein percent) at 3 selected thermal times (250, 500 and 700°C days after flowering) were tested in 6 field situations. Means and standard deviations of µ* for those situations were computed and categorized into 3 groups: "0", meaning parameters having no influence on outputs; "low" corresponding to parameters representing up to 25% of maximum µ*, and "high" for those higher than 25% of maximum µ*.

factors	oil weight			protein weight			oil percent			protein percent		
	250°C days after flowering	500 °C days after flowering	700°C days after flowering	250°C days after flowering	500 °C days after flowering	700°C days after flowering	250°C days after flowering	500 °C days after flowering	700°C days after flowering	250°C days after flowering	500 °C days after flowering	700°C days after flowering
N.hull.percent	0	0	0	0	0	0	0	0	0	0	0	0
N.Kernel.percent	0	0	0	0	0	0	0	0	0	0	0	0
N.percent.INN_0	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
N.percent.INN_stems	0	0	0	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
N.percent.INN_cap	0	0	0	0	0	0	0	0	0	0	0	0
N.percent.INN_leaves	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
stems_INN_1	0	0	0	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
stems_INN_2	0	0	0	LOW	0	0	0	0	0	0	0	0
N.uptake.TT	0	0	0	0	HIGH	HIGH	0	LOW	HIGH	0	HIGH	HIGH
N.abs2pot	0	0	0	HIGH	HIGH	HIGH	LOW	LOW	LOW	LOW	HIGH	HIGH
nk	0	0	0	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
pool.N.threshold	0	0	0	HIGH	HIGH	HIGH	LOW	LOW	LOW	HIGH	HIGH	HIGH
N.uptake.sink.growth	0	0	0	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
k	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
b.RUEpot	HIGH	HIGH	HIGH	LOW	LOW	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH
a.RUEpot	HIGH	HIGH	HIGH	LOW	LOW	LOW	LOW	HIGH	LOW	HIGH	HIGH	HIGH
t1.RUEpot	HIGH	HIGH	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	HIGH	HIGH	HIGH
photosynthesis.TT.threshold	0	LOW	HIGH	0	LOW	LOW	0	LOW	HIGH	0	LOW	LOW
W.RUE.parameter	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW
N.leaves.limit2.percent	0	0	0	0	0	0	0	0	0	0	0	0
LAI.TT.threshold	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
a2.LAI	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
SLAfv	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
C.stems.max	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH
C.cap.max	0	LOW	LOW	0	LOW	LOW	0	LOW	LOW	0	LOW	LOW
C.leaves.max	0	LOW	LOW	0	LOW	LOW	0	LOW	LOW	0	LOW	LOW
N.stems.max	0	0	0	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
N.cap.max	0	0	0	0	0	0	0	0	0	0	0	0
N.leaves.max	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
TT.C.stems.half	HIGH	HIGH	HIGH	LOW	LOW	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH
TT.C.cap.half	0	LOW	LOW	0	LOW	LOW	0	LOW	LOW	0	LOW	LOW
TT.C.leaves.half	0	LOW	LOW	0	LOW	LOW	0	LOW	LOW	0	LOW	LOW
TT.N.stems.half	0	0	0	HIGH	HIGH	HIGH	LOW	LOW	LOW	HIGH	HIGH	HIGH
TT.N.cap.half	0	0	0	0	0	0	0	0	0	0	0	0
TT.N.leaves.half	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
C.stems.rate	HIGH	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	HIGH	HIGH	HIGH	HIGH
C.leaves.rate	0	LOW	LOW	0	LOW	LOW	0	LOW	LOW	0	LOW	LOW
C.cap.rate	0	LOW	LOW	0	LOW	LOW	0	LOW	LOW	0	LOW	LOW
N.stems.rate	0	0	0	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
N.leaves.rate	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
N.cap.rate	0	0	0	0	0	0	0	0	0	0	0	0
C.leaves.limit.percent	0	0	0	0	0	0	0	0	0	0	0	0
C.cap.limit.percent	0	0	0	0	0	0	0	0	0	0	0	0
N.leaves.limit.percent	0	0	0	0	0	0	0	0	0	0	0	0
N.stems.limit.percent	0	0	0	LOW	0	0	0	0	0	0	0	0
N.cap.limit.percent	0	0	0	0	0	0	0	0	0	0	0	0
coef.proteins	0	0	0	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	HIGH
t0.oil_0	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW
t0.oil_1	HIGH	LOW	LOW	HIGH	HIGH	HIGH	HIGH	LOW	LOW	HIGH	LOW	LOW
t1.oil	0	0	HIGH	0	0	LOW	0	0	HIGH	0	0	LOW
t1.hull	0	LOW	LOW	0	LOW	LOW	0	HIGH	HIGH	0	HIGH	HIGH
cap.growth.1	0	0	0	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
cap.growth.2	0	0	0	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW
cap.growth.3	0	0	0	0	LOW	LOW	0	LOW	LOW	0	LOW	LOW
Kernel.oil.percent	LOW	LOW	LOW	0	0	0	HIGH	HIGH	HIGH	LOW	LOW	LOW
t0.proteins	0	0	0	0	0	0	0	0	0	0	0	0
t1.proteins	0	0	0	0	0	0	0	0	0	0	0	0
hull.growth.1	0	0	0	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW
hullKernel.growth.2	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
hull.growth.3	0	0	0	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
Kernel.growth.1	LOW	LOW	LOW	LOW	HIGH	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW
Kernel.growth.3	0	LOW	LOW	0	LOW	LOW	0	LOW	LOW	0	LOW	LOW

AUTEUR: Fety Nambinina ANDRIANASOLO

TITRE: Modélisation statistique et dynamique de la composition de la graine de tournesol (*Helianthus annuus* L.) sous l'influence de facteurs agronomiques et environnementaux
DIRECTEURS DE THESE: Philippe DEBAEKE, Pierre MAURY
LIEU ET DATE DE SOUTENANCE: Auzeville, 14 Novembre 2014

RESUME

Pour répondre à la demande mondiale croissante en huile et en protéines, le tournesol apparaît comme une culture très compétitive grâce à la diversification de ses débouchés et son attractivité environnementale et nutritionnelle. Pourtant, les teneurs en huile et protéines sont soumises à des effets génotypiques et environnementaux qui les rendent fluctuantes et difficilement prédictibles. Nous argumentons qu'une meilleure connaissance des effets les plus importants et leurs interactions devrait permettre de mieux prédire ces teneurs. Deux approches de modélisation ont été développées. Dans la première, trois modèles statistiques ont été construits puis comparés à un modèle simple existant. L'approche dynamique est basée sur l'analyse des relations source-puits au champ et en serre (2011 et 2012) pendant le remplissage. Les performances et domaines de validité des deux types de modélisation sont comparés.

TITLE: Statistical and dynamic modeling of sunflower (*Helianthus annuus* L.) grain composition under agronomic and environmental factors effects

SUMMARY

Considering the growing global demand for oil and protein, sunflower appears as a highly competitive crop, thanks to the diversification of its markets and environmental attractiveness and health. Yet the protein and oil contents are submitted to genotypic and environmental effects that make them fluctuating and hardly predictable. We argue that a better knowledge of most important effects and their interactions should permit to improve prediction.
Two modeling approaches are proposed: statistical one, where we compared three types of statistical models with a simple existing one. The dynamic approach is based on source-sink relationships analysis (field and greenhouse experiments in 2011 and 2012) during grain filling. Performances of both modeling types and their validity domain are compared.

Mots-clés : activité photosynthétique, huile, modélisation dynamique, modélisation statistique, protéines, régulation des flux transpiratoires, source-puits, stress hydrique

DISCIPLINE ADMINISTRATIVE : ED SEVAB : Agrosystèmes, écosystèmes et environnement

INTITULE ET ADRESSE DE L'U.F.R. OU DU LABORATOIRE : UMR 1248 AGIR