



HAL
open science

SUNFLO, a model to simulate genotype-specific performance of the sunflower crop in contrasting environments

Pierre Casadebaig, Lydie Guilioni, Jeremie Lecoer, Angélique Christophe, Luc Champolivier, Philippe Debaeke

► To cite this version:

Pierre Casadebaig, Lydie Guilioni, Jeremie Lecoer, Angélique Christophe, Luc Champolivier, et al. SUNFLO, a model to simulate genotype-specific performance of the sunflower crop in contrasting environments. *Agricultural and Forest Meteorology*, 2011, 151 (2), pp.163 - 178. 10.1016/j.agrformet.2010.09.012 . hal-01506231

HAL Id: hal-01506231

<https://hal.science/hal-01506231>

Submitted on 3 Jul 2023

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

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



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

1 **SUNFLO, a model to simulate genotype-specific performance of the**
2 **sunflower crop in contrasting environments**

3
4 Pierre Casadebaig¹, Lydie Guilioni², Jérémie Lecoœur²,
5 Angélique Christophe³, Luc Champolivier⁴, Philippe Debaeke¹

6
7 ¹ INRA, UMR 1248 AGIR, BP 52627, F-31326 Castanet-Tolosan, France

8 ² Montpellier SupAgro, UMR 759 LEPSE, 2 place Viala, F-34060 Montpellier, France

9 ³ INRA, UMR 759 LEPSE, 2 place Viala, F-34060 Montpellier, France

10 ⁴CETIOM, Centre INRA de Toulouse, BP 52627, F-31326 Castanet-Tolosan, France

11
12
13 **ABSTRACT**

14 Yield improvement certainly depends on breeding new genotypes, but also on identifying the
15 best genotype for a given location and crop management system. Hence we need to quickly
16 evaluate the performance of each new variety in different cropping systems and
17 environmental conditions.

18 Our objective was to develop a model which can help to improve genotypic assessment in the
19 sunflower crop (SUNFLO). The present work aimed at identifying, quantifying and
20 modelling the phenotypic variability of crop performance in response to the main abiotic
21 stresses occurring in the field (light, temperature, water, nitrogen) but also in the expression
22 of genotypic variability.

23 We therefore include just enough genetic information to enable the models to be used with
24 new genotypes. Each genotype was thus defined by chosen phenotypic traits which were
25 transcribed into a set of 12 genotype-specific parameters.

26 The model's performance was evaluated in both specific field experiments and generic multi-
27 environment trials (MET). The first evaluation assessed model robustness: no variables had a
28 large prediction error, indicating that the final output error results more from poor prediction
29 for all variables than from error compensation. An ANOVA on the simulated MET dataset
30 showed that although the model simulates less variability than in reality (60%), there was
31 genotype-environment interaction and the ranking of the ANOVA factors was identical in
32 both observed and simulated networks. The model's accuracy was sufficient to discriminate

33 between genotypes from different breeding periods, but was similar to the difference in
34 performance between actual genotypes ($\sim 0.2 \text{ t}\cdot\text{ha}^{-1}$).

35 To improve the understanding of crop physiology and crop-environment interactions, this
36 kind of model shows weaknesses, especially when dealing with environmental stress
37 integration or biomass allocation. On the other hand, SUNFLO seems sufficiently robust to
38 estimate the influence on yield of breeding traits or to explore new management practices.

39

40 **Key words:** crop model, genotypic variability, water stress, nitrogen stress, yield, oil
41 content, sunflower (*Helianthus annuus* L.)

42 INTRODUCTION

43

44 Sunflower is a major world oilseed crop (ca. 35 Mt, 9% oil production after palm and
45 rapeseed oil) which recently acquired new interest for biodiesel development (Pereyra-Irujo
46 et al., 2009). Considering sunflower seed production (source : FAOSTAT 2009), major
47 production regions are Europe (62 % of world production, mainly by Ukraine, Russia,
48 France), Americas (19 %, mainly Argentina, USA) and Asia (15%, mainly China, India).

49 Yield improvement certainly depends on breeding new genotypes, but also on identifying the
50 best genotype for a given location and crop management. The improvement of genotype
51 assessment is obvious at the three steps of genotype development: (1) breeding: evaluation of
52 the impact of a morpho-physiological trait (or an ideotype) on the field performance, (2)
53 cultivar registration: assessment of the performance of elite plant materials in multi-
54 environment trials and (3) cultivar dissemination: providing advice on stable and high-
55 yielding combinations of genotype, environment and crop management to farmers. Although
56 Genotype x Environment x Management (G x E x M) interactions may sometimes appear to
57 breeders as an hindrance to the stable phenotypic response of a genotype, they could be
58 exploited more fully by advisers to recommend the best cultivar-management combination in
59 a given environment, especially in a context of low-input production.

60 *Modelling can help in genotype assessment.* During breeding, yield improvement does not
61 result directly from the accumulation of sequencing or functional genomic information due to
62 very variable and unpredictable cropping environments (Miflin, 2000; Sinclair *et al.*, 2004).
63 For example, in sorghum, a QTL for the stay-green physiological trait was identified (Tao *et al.*
64 *al.*, 2000) but its effect on yield was shown to vary greatly (Borell and Hammer, 2000).
65 Predicting the effects of genotypic traits on yield for contrasting environments is thus a
66 halfway step for the exploitation of genomic results (Hammer *et al.*, 2005; Hammer *et al.*,
67 2006). During cultivar evaluation for official registration and advice, field experimentation
68 remains an essential tool, but modelling could provide additional information on genotype
69 response to varying soil-weather combinations and could limit the number of trials needed by
70 siting them in the most appropriate environments (Messina *et al.*, 2006).

71 Crop models, by their capacity to simulate phenotypic plasticity in response to environmental
72 variability (weather, soil or practices) can help to solve the inherent difficulty of genotype
73 evaluation. For sunflower, crop physiology has been incorporated in different and
74 complementary ways in a few simulation models. In most of these (Steer *et al.*, 1993 ;
75 Villalobos *et al.*, 1996; Pereyra-Irujo and Aguirrezabal, 2007) biomass accumulation is

76 driven by radiation (RUE); only Q-Sun (Chapman *et al.*, 1993) switches to a water-driven
77 (WUE) submodel if water is the main limiting factor. These models include limiting factors
78 based on their target environments: in those of Steer *et al.* (1993) and Pereyra-Irujo and
79 Aguirrezabal (2007) yield is only limited by temperature and light. To these two factors,
80 Chapman *et al.* (1993) added water stress and Villalobos *et al.* (1996) widened its scope by
81 the inclusion of nitrogen stress.

82 *How do crop models deal with genotypic variability?* Genotypic variability does not appear to
83 be as explicitly included as environmental responses in current sunflower crop models
84 (Connor and Hall, 1997). Two models consider different parameterization sets between
85 genotypes: Villalobos *et al.* (1996) introduced a statistical optimization to represent three
86 genotypes through five parameters (phenological stages, yield potential) and Chapman *et*
87 *al.*'s (1993) model did not isolate the values for genotypic parameters but rather indicated
88 genotypic variability for key parameters (water use efficiency, phyllochrons, phenological
89 stages). Moreover, the values of these parameters are mostly estimated by optimization,
90 which limits the parameterization at a larger scale for numerous genotypes. Several generic
91 crop models have been adapted to the sunflower crop, but they lack genotypic refinement
92 (Kiniry *et al.*, 1992).

93 The value of a crop model for evaluating genotypes results mainly from its predictive quality
94 but also from its ability to be updated for yearly cultivar releases from breeding companies
95 and the official registration process. A model's predictive quality is usually linked to the
96 relevance of simulated stress factors, the description of canopy growth, the robustness of
97 parameterization and the quality of input data. Updating can be made possible by improving
98 the ease of genotypic parameterization, which depends as much on the total number of
99 genotypic parameters as on the methods used to estimate their values. Therefore, an ideal
100 genotypic parameterization would aim to limit the total parameter number while maintaining
101 a useful predictive capacity.

102 Two steps can be distinguished when modelling the link between a phenotype (i.e. leaf area
103 or grain yield) and a genotype. Firstly, a physiological trait (i.e. reduced conductance with
104 water stress) could be transcribed as a model parameter (slope of the response curve)
105 (Casadebaig *et al.*, 2008). Then the robustness of the relation between the trait and the plant
106 genome should be evaluated (Yin *et al.*, 2004). The study of Chapman *et al.* (2002) illustrates
107 this view well, although the sorghum lines evaluated by the model differed only by four
108 alleles.

109 Our approach here concerns only the first step: to link a complex phenotype to a set of

110 accessible genotype traits. Each genotype is thus defined by chosen traits which were
111 transcribed into a set of genotype-specific parameters. These genotypic parameters are,
112 despite their name, under uncertain genetic control (Slafer, 2003).

113 Unlike estimating parameters by optimization, direct measurement allows parameter values
114 to be more representative of crop physiology than the paired data / optimization algorithm
115 (Jeuffroy *et al.*, 1996). In this way, the time and complex procedures needed to introduce new
116 genotypes can be reduced, improving the accessibility of the model among technical services
117 (Mavromatis *et al.*, 2001).

118 *Why develop a new sunflower crop model?* It was the need to differentiate genotypic
119 response with as few parameters as possible, rather than a lack of suitability of existing
120 sunflower crop models, that led us to develop a new crop model for sunflower. An analysis of
121 the comparative performance of genotypes from different breeding generations (Vear *et al.*,
122 2003) through the generic approach of biomass production of Monteith (1977) allowed
123 phenotypic variability to be quantified in this system (Debaeke *et al.*, 2004). That study
124 confirmed the role of (1) plant leaf area and its control (persistence, efficiency) after
125 flowering, (2) vertical distribution of this leaf area and (3) phenology (duration from
126 flowering to maturity) for genotype performance. This knowledge was incorporated into a
127 crop model intended to evaluate the contribution of identified parameters to the variability of
128 yield potential (Lecoeur *et al.*, 2009).

129 Our objective was to identify, quantify and model phenotypic variability of sunflower at the
130 individual plant level in response to the main abiotic stresses occurring at field level (light,
131 temperature, water, nitrogen) but also in the expression of genotypic variability (G x E
132 interactions).

133 This objective underlies a parsimonious addition of genetic information to increase the
134 model's versatility and usefulness. Using a crop model to help to evaluate genotypes thus
135 requires (1) choosing a suitable modelling framework, (2) setting up genotypic parameters,
136 (3) estimating their values with a simple and robust methodology, (4) evaluating the model's
137 performance on a small-scale evaluation trial network.

138

139 **MATERIALS AND METHODS**

140

141 1. Experimental design and crop conditions

142 Crop measurements from 56 experiments, carried out on 27 trials, were used to develop,
143 calibrate and evaluate the model's performance.

144 These experiments, described in **Table 1**, covered a wide range of genotypic variability (2-20
145 genotypes) and cultural conditions: potential growth conditions (exp. #01, 02, 04, 05), water
146 (exp. #03, 23, 24-40) and nitrogen (exp. #09-14) stressed situations, water-nitrogen
147 interaction trials (exp. #15-22). Design was split-plot with 3-4 replicates depending on trials,
148 plot size was ranging from 20 to 30 m², on mainly silty clay to silt soils (see Tabl for
149 available water content estimation).

150 The first dataset (trials #1-40) was used to estimate the model's adjustment capacity to
151 observed datas. It consisted of a detailed phenotyping (7 variables : phenology, absorbed
152 nitrogen, leaf area index, radiation interception efficiency, above-ground biomass, achene
153 yield and oil content) performed on different genotypes in various environmental conditions,
154 together with a precise soil, weather and management description, and was used to reduce
155 uncertainty from model inputs. Evaluating the model on this dataset would provide insight on
156 the lowest prediction error attainable. For each intermediate variable, non-genotypic model
157 parameters were optimized on a subset of experiments chosen to evaluate the variable. In this
158 way, no single experiment was used to estimate all model parameters. Parameters for the two
159 statistical models predicting output variables (yield, oil content) were estimated using all the
160 information available in this dataset, which could be regarded as a "model development" set.

161 A second dataset (trials #41-56) was created to estimate the model's prediction capacity and
162 discriminate genotypic performance. Observations of yield and oil content came from a
163 multi-environment experimental network (16 trials of 20 genotypes from the combination of
164 15 sites x 2 years, 2000-2001) which was used in an earlier study to assess the genetic
165 progress made in France on the sunflower crop (Vear *et al.*, 2003). Environmental description
166 was less accurate than for the first dataset as these were simply cultivar comparison trials.
167 These data were not used to estimate the model parameters, but allowed us to test the
168 predictive capacity of the model for a standard application. This dataset may be viewed as a
169 "model evaluation" set.

170

171 2. Plant measurements

172 The timing of developmental stages was determined on six plants per genotype, twice a
173 week, using the scale proposed by CETIOM for sunflower (CETIOM, 2004). In addition, the
174 number of visible, senescent (yellowish surface > 50% leaf area) and dead leaves was also
175 counted. A particular crop phenological stage was considered as having been reached when
176 more than 50 % of the plants evinced the features of that phenotype.

177 Leaf ranking was counted from the base to the top of the stem. Architectural measurements

178 were made during successive steps from flower bud appearance to the end of flowering,
179 provided that leaves were fully expanded. The rank, length, and width of each leaf blade were
180 measured on 6 plants per genotype with a ruler (± 0.5 mm). Individual leaf area (ILA, mm²)
181 was estimated from leaf largest length (L) and largest width (W) using a break linear model
182 [EQ1] calibrated using a planimeter (LI-3100, Li-cor inc., Lincoln, NE, USA).

183

184 [EQ1] $ILA = a * LW$ if $LW < c/(a-b)$ else $b * LW + c$;

185 where $a=0.684$; $b=0.736$ and $c=-8.860$ ($R^2=0.992$, $RMSE=5.9$, $n=304$)

186

187 Significant genotypic variability was not observed for the parameters of this relationship.

188 Total leaf area (TLA) was calculated as the sum of the individual leaf areas.

189 The radiation interception efficiency (RIE) of the non-senescent fraction of the canopy was
190 measured (400–700 nm wavelength) during crop growth using a hand-held *Picqhelios*
191 apparatus (AERIC, Balma) [Picq1988]. Cumulative intercepted radiation was computed from
192 daily incident PAR and daily RIE, interpolated between the sampling dates. The light
193 extinction coefficient (k) was obtained from non-linear regressions with Beer's law (Monsi
194 and Saeki, 2005).

195 Total shoot dry biomass (TDM) was measured on square meter quadrats (5-7 plants).
196 Radiation use efficiency (RUE) was then calculated for different phenological stages from
197 total dry biomass divided by cumulative intercepted radiation (PAR). Total plant nitrogen
198 content was assessed by the Dumas combustion method, allowing the calculation of the total
199 absorbed nitrogen in aerial parts.

200 The photosynthetic parameter (PHS) was estimated from leaf photosynthetic activity
201 measured with a portable photosynthesis system (CIRAS, PP system, UK) with a control
202 radiation level of $1500 \mu\text{mol m}^{-2} \cdot \text{s}^{-1}$. All the photosynthesis genotypic values were
203 normalized with respect to those obtained for the genotype cv. Melody.

204 Harvest index (HI) was the ratio of dry grain weight to total dry matter on sampled plants.
205 Yield was expressed on a dry basis (0% moisture) cleaned of trash. Oil content was
206 determined by Nuclear Magnetic Resonance (Bruker NMS 110 Minispec NMR Analyzer) on
207 20g dry achene samplings.

208

209 3. Environment and management description.

210 The input variables required for simulation were environmental (weather and soil) and related
211 to crop management. Four weather variables were recorded daily: mean air temperature ($^{\circ}\text{C}$,

212 2m height), global incident radiation ($\text{MJ}\cdot\text{m}^{-2}$), potential evapotranspiration (mm, Penman-
213 Monteith) and precipitation (mm). The distance between the crop and the weather station was
214 greater in the evaluation dataset (10-20 km) than in the development set (0-5 km). The
215 available soil water content over the maximum rooting depth (the difference between
216 volumetric moisture content at field capacity and wilting point) was estimated from soil
217 texture (clay content), apparent soil density and stone content. Crop management (Table 1)
218 was summarized by sowing operations (date, seed depth, plant density ; 6-7 pnt/m²), fertilizer
219 regime (dates and amounts), and irrigation schedule (dates and amounts). Soil water and
220 nitrogen (mean = 40.7 kg/ha) at sowing were initialized from water balance and nitrogen
221 balance sheet models : in >90% of the cases, initial water content was equal to field capacity.

222

223 4. Statistical analysis and naming conventions.

224 Observed and simulated values were compared using a range of statistical criteria using mean
225 squared deviation and its derivative (Kobayashi and Salam, 2000). Relative Root Mean
226 Square of Errors (RRMSE) was calculated as the quotient of RMSE divided by the mean of
227 observations in order to facilitate comparison between variables and other models. Statistical
228 modelling was done using multiple linear regression. All data analysis, statistics and graphics
229 were done using R software (R Development Core Team, 2006).

230 Given a variable named V , the rate of increase of V – or its daily increase in the case of a fixed
231 time-step integration – was named dV . Variables related to the vertical dimension (soil and
232 rooting depth) were prefixed with z . The name of environmental stress related variables
233 combined a letter {T, L, W, N} for the source stress (for temperature, light, water and
234 nitrogen) and the target variable (ex : the effect of water stress on radiation use efficiency is
235 W.RUE,).

236

237 **RESULTS : MODEL DEVELOPMENT**

238

239 1. Modelling approach

240 Starting from a simple modular modelling basis, we used the model SUNFLO_V0 (Lecoeur
241 *et al.*, 2009) as an initial modelling framework. It estimates the above-ground biomass
242 production of a sunflower crop from incident radiation and mean air temperature. It works on
243 a daily time step and describes plant phenology, leaf expansion and biomass production and
244 its allocation to the grain. Rather classically, actual growth depends on potential growth
245 modulated by temperature and radiation. Genotypic variability is taken into account by a

246 small number (12) of parameters (Lecoeur *et al.*, 2009). Moreover, a fine resolution in
247 describing biomass production is not always the best starting point (Hammer *et al.*, 2006).
248 When building a model framework, several groups of equations and parameters can be
249 proposed to describe a given physiological process. If their inputs and outputs are measurable
250 and effectively measured, treating these modules as independent sub-models can have several
251 advantages. Firstly it allows easier knowledge sharing among the scientific community
252 (Jones *et al.*, 2001). Secondly, evaluating the model response module by module gives to the
253 overall performance more credibility throughout the simulated growth period and prevents
254 accidental error compensation (Sinclair and Seligman, 2000). SUNFLO_V1, described
255 below, consists of 7 modules whose structure appears in **Figure 1**.

256

257 *Adding some genotypic variability.* Our objective was to evaluate the possibility of coupling a
258 biophysical model and a phenotyping approach to quantify phenotypic variability of yield
259 and oil content in contrasting cropping environments. Each genotype was defined by a set of
260 parameters whose values are assumed to be constant among environments, thus trying to
261 mimic gene functioning (Colson *et al.*, 1995; Boote *et al.*, 2003).

262 When using experimentation to measure genotypic parameters, there are two kinds of
263 constraint. The first is the amount of work involved, where the number of new genotypes and
264 time available for phenotyping limit the number of genotypic parameters which can be
265 investigated in the future. The second is due to the environmental effects on the values of
266 measured parameters: the use of phenotypic information, as parameter values can only be
267 taken to be genotypic if they prove to be stable in different environments. We finally
268 proposed four criteria before choosing a parameter as genotype-specific; it should (1) present
269 a significant amount of genotypic variability, (2) be stable over environments, (3) markedly
270 affect the model outputs and (4) be easily measurable.

271

272 2. Model structure

273

274 2.1 Phenology

275 Plant phenology is driven by thermal time (Aiken, 2005). Cumulative thermal time since
276 emergence (TTE, °C.d) was calculated [EQ2] as the sum of the daily mean air temperature
277 from emergence using a base temperature (T_b) of 4.8°C (Granier and Tardieu, 1998)
278 common to all genotypes. Four key stages, expressed as genotype-dependent thermal dates,
279 delimited periods of plant growth with changes in plant physiology: floral initiation (FI),

280 beginning of flowering (F), beginning of grain filling (early maturation, EM) and
281 physiological maturity (PM) (CETIOM, 2004; Lecoecur *et al.*, 2009).

282 Photoperiodic effects on phenology were not included in the model: (Leon *et al.*, 2001)
283 showed that flowering date was unaffected by photoperiods between 14.5 and 16h at
284 emergence. Extreme locations and practices for French sunflower crops could lead to
285 photoperiods at emergence from 13.8 h (southern early sowings) to 16.5 h (northern late
286 sowings), thus showing a very slight effect on flowering date in a few cases.

287 Through the reduction of transpiration, water stress causes overheating of the plant. This
288 heating can accelerate crop development and was modelled using a multiplicative effect with
289 thermal time accumulation [EQ2].

290

291 [EQ2] $TTE = \sum((T_m - T_b) * (1 + a(1 - W.TR)))$; where $T_b = 4.8$ °C and $a = 0.1$

292

293 The duration of the sowing-emergence period (E, °C.d) [EQ3], provided soil water content at
294 seed depth is adequate, is assumed to depend only on temperature (Angus *et al.*, 1981).
295 Adequacy of water content for emergence was assumed and the sowing - emergence phase
296 was modelled as in (Villalobos *et al.*, 1996) as a function of air temperature and sowing
297 depth. Hypocotyl elongation (dHE) and germination time (G) were computed with a common
298 base temperature.

299

300 [EQ3] $E = G + dHE * zSOW$; where $G = 86$ °C.d and $dHE = 1.19$ °C.d/mm

301

302 *2.2 Environmental factors limiting crop production*

303

304 *2.2.1 Temperature*

305 Temperature effects on processes are mostly accounted for in the model by using thermal
306 time based relationships. However, we needed to include a direct effect of temperature on
307 radiation use efficiency (RUE, g.MJ⁻¹) and on nitrogen net mineralization rate (dMin,
308 Kg.j.ha⁻¹).

309 A depressive function of non-optimal temperature (T.RUE) was applied to RUE, calculated
310 [EQ4] from daily mean air temperature. This relationship was adapted from Horie (1977) to a
311 bilinear function as in (Villalobos *et al.*, 1996) using upper and lower optimal (T_{ou}, T_{ol}) and
312 critical (T_c) temperature ranges.

313

314 [EQ4] $T.RUE = T_m * (1 / (Tol - T_b)) - (T_b / (Tol - T_b))$ if $T_m < Tol$;

315 1 if $Tol < T_m < Tou$

316 $T_m * (1 / (Tou - T_c)) - (T_c / (Tou - T_c))$ if $T_m > Tou$

317 else 0

318 where $Tol = 20$ °C, $Tou = 28$ °C, $T_c = 37$ °C.

319

320 Nitrogen net mineralization rate (dMin, EQ25) was affected by a logistic function [EQ5] of
321 daily mean air temperature (Valé *et al.*, 2007).

322

323 [EQ5] $T.NM = 36 / (1 + (36 - 1) * \exp(-0.119 * (T_m - 15)))$

324

325 2.2.2 Light

326 Competition for light interception, modified by planting density, affects individual leaf
327 expansion rate [EQ] in sunflower (Rawson and Hindmarsh, 1983; Rey *et al.*, 2008). In an
328 experimental study (Rey, 2003), leaf expansion rate was related to absorbed PAR (APAR)
329 per unit leaf area. By assuming (1) that absorbed PAR is equal to intercepted PAR and (2)
330 that all leaves are affected the same way by light restriction, this previously developed model
331 can be used to drive canopy response to light in a simpler crop model. This relation [EQ6]
332 was initially developed using potential leaf areas (at 1 plant.m^{-2}) as a reference. In the model a
333 scaling parameter ($s = 2.5$) allows it to be applied to potential leaf areas in cropping
334 conditions (at 6 plant.m^{-2}). This parameter is measured as the ratio of plant leaf area at 1
335 plant.m^{-2} to the leaf area at 9 plant.m^{-2} ; Sadras and Hall (1988) have indicated a larger
336 response to density ($s = 3.1$).

337

338 [EQ6] $L.LE = s * (a + (b / (1 + \exp(-(IPAR - c) / d))))$;

339 where $s = 2.5$, $a = -0.14$, $b = 1.13$, $c = 4.13$ and $d = 2.09$

340

341 2.2.3 Water

342 a. Soil and water budget

343 The soil water budget model is based on a previously developed model (Lecoeur and Sinclair,
344 1996) that was used for irrigation scheduling (Sarr *et al.*, 2004). Water movement in the soil
345 is assumed to be only vertical, with runoff and lateral flow being ignored. It is possible to
346 simulate the fraction of transpirable soil water (FTSW), which accounts for the amount of
347 soil water available to the plant within the root zone (Sinclair, 2005). The model treats the

348 soil as a reservoir with three layers (C_i), whose thicknesses (z_{C_i}) change as the roots grow.
349 The required soil physical characteristics are gravimetric moisture content at field capacity
350 (H_{fc} , % of dry soil) and at permanent wilting point (H_{wp} , % of dry soil) and the maximum
351 rooting depth (z_{PR} , mm).

352 Evaporation and N mineralization only take place in the uppermost layer (0-300 mm), in
353 which water is supplied by precipitation (P_r) and irrigation (I_{rr}) and lost by evaporation (EV),
354 transpiration (TR) and drainage (D). Drainage (DC_i , mm) occurs when the water content of a
355 layer exceeds its water retention capacity (H_{fc}).

356 The depth to the bottom of the second layer (z_{C2} , mm) is equal to the rooting depth (z_R ,
357 mm). z_R is linearly related to thermal time [EQ7] until the potential rooting depth (z_{PR} , mm)
358 is reached. In the model, neither soil moisture nor plant biomass may limit root growth, so the
359 advance of the rooting front (dR) was estimated at 0.7 mm/C.d. The thickness of the bottom
360 layer (z_{C3} , mm) is defined as the difference between potential and actual rooting depth.

361

362 [EQ7] $z_R = \sum(dR * T_m)$ if $z_R < z_{PR}$;

363 else $z_R = z_{PR}$;

364 where $dR = 0.7$ mm/C.d

365

366 The water content of each soil layer results from the water balance as described in [EQ8-10]

367

368 [EQ8] $WC_1 = P_r + I_{rr} - dTRC_1 - dEV - DC_1$

369 [EQ9] $WC_2 = +DC_1 - DC_2 - dTRC_2 + (dR * T_m) * (WC_3/z_{C3})$

370 [EQ10] $WC_3 = +DC_2 - DC_3 - (dR * T_m) * (WC_3/z_{C3})$

371

372 Soil evaporation is separate from plant transpiration and is calculated [EQ12] as the product
373 of the fraction of radiation reaching the soil, reference evapotranspiration and daily relative
374 soil evaporation (dRE). dRE [EQ11] accounts for a reduction of soil hydraulic conductivity
375 with time from the last effective rain (DWW , Days Without Water), DWW being
376 incremented if ($P_r + I_{rr} < 3$ mm).

377

378 [EQ11] $dRE = (DWW + 1)^{0.5} - (DWW)^{0.5}$

379 [EQ12] $dEV = (1 - RIE) * ETP * dRE$

380

381 Water loss due to plant transpiration [EQ15-16] is a function of potential transpiration rate

382 [EQ13], water effect on transpiration [EQ19] and root distribution over the two first layers
383 (fR). This distribution is made proportional to the thickness of each soil layer [EQ14].

384

385 [EQ13] $dPTR = Kc * ETP * RIE$; where $Kc = 1.2$

386 [EQ14] $fR = zC1/(zC1+zC2)$

387 [EQ15] $dTRC1 = fR * dPTR * W.TR$ if ($zR > zC1$) else $dPTR * W.TR$

388 [EQ16] $dTRC2 = (1-fR)* dPTR * W.TR$ if ($zR > zC1$) else 0

389

390 The fraction of transpirable soil water (FTSW), the main output from the water budget
391 module [EQ17], is used as a water stress index. FTSW is the ratio of actual to total water
392 content [EQ18] of both root-explored layers.

393

394 [EQ17] $FTSW = (WC1 + WC2) / (PWC1 + PWC2)$

395 [EQ18] $PWCi = zCi * (Hfc.Ci - Hwp.Ci) * sd$

396 where sd is soil apparent density

397

398 b. Modelling the effects of water stress on plant growth

399 Crop growth is explicitly affected by water stress considering three processes: leaf expansion
400 (LE), plant transpiration (TR) and biomass production (RUE). Three water stress factors [0-
401 1] are thus based on FTSW and a genotypic parameter which differs according to the process
402 [EQ19] ; the effect of water on RUE is assumed to be the same as on transpiration. A
403 significant genotypic variability was found in sunflower water stress response when the
404 parameters were estimated in three glasshouse experiments (Casadebaig *et al.*, 2008).

405

406 [EQ19] $W.LE ; W.TR ; W.RUE = -1 + 2/(1+\exp(a*FTSW))$

407 where a is a measured genotypic parameter different for expansion (W.LE) and gas exchange
408 (W.TR and W.RUE) .

409

410 Net nitrogen mineralization was affected by soil moisture in the surface layer by using a
411 linear relationship previously parameterized by Mary *et al.* (1999) and used in a dynamic N
412 leaching model.

413

414 [EQ20] $W.NM = 1 - (1-Fpf) * (1 - RWCC1)$

415 where $Fpf = 0.2$ is the value of the function at Hpf and $RWCC1$ is the relative water content

416 in the surface layer (actual by maximal water content).

417

418 2.2.4 Nitrogen

419

420 a. Nitrogen budget

421 The mineral nitrogen content of the soil layers (kg/ha) depends on mineralization (dMin),
422 fertilization (Fer), leaching (LCi), denitrification (-dDenit) and plant uptake [EQ21-23].

423

$$424 \text{ [EQ21] } NC1 = +dMin + Fer - dMFC1 - dAAC1 - LC1 - dDenit$$

$$425 \text{ [EQ22] } NC2 = +LC1 - LC2 - dMFC2 - dAAC2 + dR * (NC3 / zC3)$$

$$426 \text{ [EQ23] } NC3 = +LC2 - LC3 - dR * (NC3 / zC3)$$

427

428 The time between fertilization and its availability for plant uptake was simply modelled using
429 a threshold of precipitation or irrigation (5 mm) required for N solubilization and
430 incorporation into the surface layer. Moreover, the processes of N immobilization and
431 volatilization were avoided in the model [EQ24] by linking fertilization efficiency (%) and
432 crop growth rate ($\text{g} \cdot \text{m}^{-2} \cdot ^\circ\text{C} \cdot \text{d}^{-1}$). Limaux *et al.* (1999) showed that nitrogen use efficiency is
433 positively correlated to crop growth rate.

434

$$435 \text{ [EQ24] } NUE = 30 + 0.34 * CGR * 100$$

436

437 Nitrogen mineralization [EQ25] was modelled as a potential mineralization rate parameter
438 affected by temperature [EQ5] and soil moisture content [EQ20] as described by Mary *et al.*
439 (1999) and Valé *et al.* (2007). When considering nitrogen loss, leaching (LCi) was the
440 product of drained water and its nitrogen concentration from the soil layer concerned. As the
441 model is to be applied mostly in warm cropping conditions, nitrogen denitrification was
442 added, following Sinclair and Muchow (1995), as an exponential function of air temperature
443 [EQ26].

444

$$445 \text{ [EQ25] } dMin = MINP * T.NM * W.NM$$

$$446 \text{ [EQ26] } dDenit = 6 * \exp(0.07738 * Tm - 6.593)$$

447

448 Generally, plant nutrient uptake in dynamic crop models results from comparing soil supply
449 (determined by the nitrogen budget) and plant demand (based on the crop nitrogen nutrition

450 index). Nitrogen uptake has two components as in the STICS model (Brisson *et al.*, 2003):
451 one part is absorbed in the transpirational stream (mass flow, [EQ27]) while the other is
452 directly absorbed by the roots, simulating active nitrogen influx. This active influx was
453 modelled [EQ28] by using a Michaelis-Menten function of soil solution nitrogen
454 concentration, rooting density and layer thickness, using the shape and affinity constants
455 defined previously by (Brisson *et al.*, 2008). Daily nitrogen uptake for the crop is the sum of
456 mass flow and active nitrogen influx.

457

$$458 \text{ [EQ27] } dMF = dTRC1 * NC1 + dTRC2 * NC2$$

$$459 \text{ [EQ28] } dAA = Vm1 * NCi / (Km1 + NCi) + Vm2 * NCi / (Km2 + NCi) * zCi * fR$$

460

461 The critical crop nitrogen uptake is defined as the minimum nitrogen uptake necessary to
462 achieve maximum biomass accumulation. Across a range of crops, the critical N uptake is
463 related to biomass accumulation by a power function with a coefficient less than unity which
464 suggests that crop N uptake is regulated by both soil N supply and biomass accumulation
465 (Lemaire and Meynard, 1997). Two thresholds (critical and maximal) for plant nitrogen
466 content were thus defined. These thresholds were experimentally determined by monitoring
467 nitrogen accumulation in relation to crop biomass for various fertilization levels (0 - 160 N)
468 in a sunflower crop (Debaeke and Raffaillac, 2006). Critical nitrogen (%) dilution in the plant
469 biomass ($t \cdot ha^{-1}$) was established as EQ29. Maximal nitrogen content of biomass was defined
470 as 1.3 times the critical content (Stockle and Debaeke, 1997).

471

$$472 \text{ [EQ29] } PNCc = \min(a * TDM^{-b})$$

473 $a = 5$ is the plant nitrogen content (%) at $TDM = 1 t \cdot ha^{-1}$ and $b = 0.49$

474

475 b. Crop response to nitrogen stress

476 Two nitrogen stress indexes (NNI, INNI), both based on the ratio of actually absorbed N (N_a ,
477 $kg \cdot ha^{-1}$) to the critical N amount needed to satisfy the demand (N_c), but differing in the way
478 they are calculated, were used : the nitrogen nutrition index (NNI) depends on cumulated
479 absorbed N while the instantaneous nitrogen nutrition index (INNI) is derived from the rate
480 of absorption (dN_a/dN_c).

481 Nitrogen stress factors [0-1] are linearly linked to NNI or INNI (Brisson *et al.*, 2008): the
482 nitrogen stress effect on leaf expansion (N.LE) is governed by NNI while the effect on RUE
483 (N.RUE) is equal to INNI.

484

485 [EQ30] $N.LE = 1.75 * NNI - 0.75$ if $NNI > 0.6$; else $N.LE = 0.3$

486 [EQ31] $N.RUE = INNI$

487

488 2.3 Leaf area dynamics

489 Considering a crop model based on Monteith's notions (Monteith, 1977), where crop growth
490 is driven mainly by light interception, an accurate prediction of leaf area dynamics is
491 essential for a correct estimation of its dependent variables. This is especially true for
492 sunflowers, where light intercepted after flowering is closely related to oil yield (Merrien and
493 Grandin, 1990; Sadras *et al.*, 2000; Aguirrezabal *et al.*, 2003). As leaf area is also an
494 important determinant of plant transpiration, a common adaptive trait for drought tolerance is
495 reduced leaf growth under water stress (Jones, 2004). There are significant differences in leaf
496 area and its vertical distribution in commercial hybrids (Debaeke *et al.*, 2004; Lecoeur *et al.*,
497 2009). When modelling leaf area, it is important to represent the major interactions between
498 genotype, environment and crop management arising at this level.

499 Crop leaf area depends mainly on the processes of leaf appearance, expansion and
500 senescence, which are affected by various limiting factors - light competition (plant density)
501 [EQ6], water stress [EQ19] and nitrogen stress [EQ30]. However, at field level, the particular
502 interplay of the timing of stress (due to environment or crop phenology) and individual organ
503 expansion leads to very different patterns of leaf area development. This being so, describing
504 a process at an organ sub-level should improve crop leaf area prediction.

505 Three main modelling procedures have been used so far to predict crop leaf area : (i) the use
506 of functions of time or thermal time (Gompertz, polynomials, exponentials) (Chapman *et al.*,
507 1993; Brisson *et al.*, 2003; Pereyra-Irujo and Aguirrezabal, 2007), (ii) the concept of specific
508 leaf area, which uses biomass as the driving variable of leaf area (Penning de Vries *et al.*,
509 1989; Villalobos *et al.*, 1996) and (iii) simulating leaf expansion and senescence separately
510 on the scale of the leaf (Stewart and Dwyer 1994; Lizaso *et al.*, 2003).

511 These approaches differ mainly in their concepts, but also in their mathematical complexity
512 and hence in their numbers of parameters. Errors in estimating their values (statistically or
513 experimentally) have differing impacts, depending on the approach concerned.

514 (Lizaso *et al.*, 2003) investigated a cut-down version of the CERES-Maize model which
515 combines accuracy with ease of parameterization. It simulates individual leaf expansion as
516 the difference between growth and senescence rates. Both expansion and senescence follow a
517 logistic function of thermal time with parameters for a final state (potential area for

518 expansion, actual for senescence), a thermal time at 50% of the final state, and a slope. The
519 mathematics are described by Lizaso *et al.* (2003) but the model equations are presented
520 again in this paper.

521 We intend to adapt this framework to the sunflower crop by further simplifying it to facilitate
522 the parameterization of new genotypes as soon as they become available on the market.

523

524 2.3.1 Timing of leaf appearance and senescence

525 The rate of leaf appearance depends on air temperature and two phyllochrons (°Cd) as
526 performed lower leaves appear at a lower rate (Rey, 2003). Thermal time to 50% of final leaf
527 area was defined by [EQ33] where T_{ii} is the thermal time required for leaf appearance.

528

529 [EQ32] $T_{ii} = i * PHY1$ if $i < 7$;

530 else $T_{ii} = (i - 5) * PHY2 + 400$

531 where $PHY1 = 71.4$ and $PHY2 = 16.3$ °C.d [Rey2003].

532

533 [EQ33] $Te_i = T_{ii} + 1 / Ke_i$

534 where $Ke_i = 0.013$

535

536 Leaf longevity [EQ34] was modelled with a bell-shape curve depending on 2 genotypic
537 parameters and 3 estimated (using plant leaf area as variable) parameters: LL_0 (°C.d) was the
538 asymptote of the curve; a (°C.d), the amplitude and b controlled the width of the curve while
539 LLH controlled the rank of the most persistent leaf (identical to the largest leaf) and TLN was
540 the total leaf number of the genotype. Leaf senescence date [EQ35] was calculated by adding
541 leaf longevity to the 50% expansion date (Te_i).

542

543 [EQ34] $LL_i = LL_0 + a * \exp(-((i - LLH)^2) / ((b * TLN)^2))$

544 where $LL_0 = 153$ °C.d, $a = 850$ °C.d, $b = 0.78$, LLH and TLN values are defined in Lecoeur
545 *et al.* (2009)

546 [EQ35] $Ts_i = Te_i + LL_i$

547

548 2.3.2 Leaf expansion and senescence

549 The vertical distribution of leaf area is variable among sunflower genotypes grown in low-
550 stress conditions (Debaeke *et al.*, 2004). The distribution of individual leaf area (ILA) down
551 the stem (leaf area profile) was modelled with a bell-shape curve (Keating and Wafula, 1992;

552 Lizaso *et al.*, 2003). Three parameters were genotypic and came from direct measurements
 553 (Lecoeur *et al.*, 2009): largest leaf size (LLS, cm²) and height (LLH, rank) and total leaf
 554 number (TLN). The two last parameters control the shape : a affects the width and b affects
 555 the skewness of the curve. As the values for parameters a and b were not stable across
 556 genotypes (CV = 39 %), we derived their value from the 3 other genotypic parameters with a
 557 linear model. This model performed 35% better (n = 754, rmse = 68 cm²) for ILA prediction
 558 when compared with mean values (constant) for the shape parameters.

559

$$560 \text{ [EQ36] } A_{ei} = LLS * \exp(a * ((i-LLH)/(LLH-1))^2 + b * ((i-LLH)/(LLH-1))^3)$$

$$561 \text{ where } b = 1.5 - 0.22 * LLH - 3.53E-4 * LLS + 0.082 * TLN$$

$$562 \text{ and } a = -2.31 + 0.018 * LLH - 1.64E-3 * LLS + 0.0199 * TLN + 0.92 * b$$

563

564 Individual leaf expansion rate (dA_{ei}) combined temperature, final individual leaf area
 565 [EQ36], thermal time to half-expansion [EQ33] and slope of expansion in a logistic relation
 566 [EQ37]. The slope parameter (K_{ei}) was constant for all the leaves and also during
 567 senescence. As previously stated (Villalobos *et al.*, 1996) changes in leaf size along the stem
 568 are largely due to variation in the rate of expansion rather than to its duration. Consequently,
 569 expansion rate was limited by the multiplicative effects of light [EQ6], nitrogen [EQ30] and
 570 water [EQ19] without further prioritization.

571

$$572 \text{ [EQ37] } dA_{ei} = T_{eff} * (A_{ei} * K_{ei}) * \exp(-K_{ei} * (TTE - Tei)) / (1 + \exp(-K_{ei} * (TTE - Tei)))^2$$

$$573 \text{ [EQ38] } S_{Fei} = \int (dA_{ei} * W.LE * N.LE * L.LE)$$

574

575 The progression of leaf senescence (dA_{si}) was modelled on the same basis as expansion,
 576 except that variables A_{ei} and T_{ei} were swapped with actual expanded leaf area (A_{si}) and time
 577 to half-senescence [EQ35]. The senescent area of each leaf was the result of direct integration
 578 of senescence rate, without stress effects.

579

$$580 \text{ [EQ39] } dA_{si} = T_{eff} * (A_{si} * K_{ei}) * \exp(-K_{ei} * (TTE - Tsi)) / (1 + \exp(-K_{ei} * (TTE - Tsi)))^2$$

$$581 \text{ [EQ40] } TLA = \sum (S_{Fei} - S_{fsi})$$

582

583 Plant total green leaf area (TLA) was calculated from the difference of total and senescent
 584 individual leaf areas [EQ40].

585

586 *2.4 Light interception*

587 The radiation interception efficiency [EQ41] was estimated as an exponential function of LAI
588 (Monsi and Saeki, 2005) using an extinction coefficient (k) that varied among genotypes but
589 was constant during crop growth. Genotypic values of k were estimated by non-linear
590 regression between measured LAI and RIE in field conditions under low crop stress (Lecoeur
591 *et al.*, 2009).

592

593 [EQ41] $RIE = 1 - \exp(-k * LAI)$

594

595 *2.5 Biomass accumulation*

596 Daily intercepted radiation was converted into biomass increment using radiation use
597 efficiency [EQ42] ($g.MJ^{-1}$, PAR) which changes with crop phenology (Villalobos *et al.*,
598 1996). The pattern of RUE evolution during crop growth was monitored in a field experiment
599 (Lecoeur *et al.*, 2009). In the absence of environmental stress, the model assumed a constant
600 value for RUE at the very beginning of the cycle, a linear increase to a maximal value during
601 flowering and an exponential decrease from early maturity to physiological maturity.
602 Genotypic parameters were used for 3 phenostages : Thermal date to flowering, early
603 maturity and physiological maturity (respectively TDF, TDEM and TDPM). Similarly to leaf
604 expansion, effective RUE resulted from the multiplicative effects of temperature [EQ4],
605 nitrogen [EQ31], water [EQ19] and a genotype-specific coefficient (PHS). PHS differentiated
606 maximum photosynthetic rates among genotypes in controlled conditions (Lecoeur *et al.*,
607 2009).

608 Total aerial dry biomass was finally calculated as the integration of the product of PAR, RIE
609 and effective RUE.

610

611 [EQ42] $RUE = 1$ if $TTE < 300$ °C.d

612 $= 1 + ((TTE - 300) * 2 / (TDF - 300))$ if $TTE < TDF$

613 $= 3$ if $TTE < (TDEM - 100)$

614 $= a * \exp(b * (1 - ((TDE - TDEM) / (TDPM - TDEM))))$ if $TTE < TDPM$

615 where a = 0.015 is the asymptote and b = 4.5 is the slope of the curve. TDF, TDEM and
616 TDPM value were defined in Lecoeur *et al.* (2009)

617

618 [EQ43] $dTDM = 0.48 * R_g * RIE * RUE * W.RUE * N.RUE * T.RUE * PHS$

619

620 *2.6 Allocation : calculating crop variables as a statistical model input.*

621 Up to the simulation of crop aerial biomass, the model used process-based dynamic
622 relationships. We propose to calculate yield and oil content by using simple statistical models
623 that depend on a set of 11 previously simulated crop variables and or used genotypic
624 parameters. New genotypic variability was introduced at this point by measuring 2 additional
625 parameters : harvest index (glasshouse experiment, Lecoecur *et al.*, 2009) and oil content
626 (trials #1-40) in low stress conditions (potential).

627 This information, mostly reused from the mechanistic part of the model could be classified in
628 different categories: climatic, abiotic stress, canopy indexes and genotypic parameters. **Table**
629 **2** describes more precisely these indexes and their method of calculation.

630 Previously calculated indexes were used as covariables (independent variables) in two
631 multiple linear regression models to explain the value of harvest index and oil content at
632 harvest time. When building these statistical models, we applied a simple selection procedure
633 (stepwise selection on BIC) to prune non-contributing variables, going from 13 to 7 or 8
634 covariables, depending on the model.

635 Parameters for the additive statistical models were estimated on the “development” dataset
636 and their robustness was tested in the “evaluation” set of experiments.

637 We adopted statistical modelling as, for example, Hammer and Broad (2003) stated large
638 variations of HI dynamics with environment and, on our data, the statistical approach proved
639 to be more precise for yield prediction (Casadebaig, 2008) than the parametrisation of a
640 simple mechanistic HI model (Moot et al., 1996). We followed a similar statistical approach
641 for modeling oil content to be more exhaustive in the driving factors, whereas the
642 mechanistic models considered only post-flowering light interception (Aguirrezabal et al.,
643 2003) or grain nitrogen content (Triboi and Triboi-Blondel, 2002). In this case the process-
644 based approach performed better on low stress conditions, but showed inconsistencies on a
645 larger panel of stress conditions (Casadebaig, 2008).

646

647 **RESULTS : MODEL PERFORMANCE**

648

649 1. Model performance in cases where the plant environment is well characterized

650 The model was first evaluated on a dataset where crop environment (soil, weather and
651 management) was accurately characterized, thus minimizing errors in input variables. Seven
652 crop variables were sampled and compared to the simulated ones (**Figure 2**) covering nearly
653 every module presented in **Figure 1**.

654 In general, intermediate variables were predicted with an accuracy (relative RMSE) ranging
655 from 4% to 30%. **Table 3** presents an analysis of the model error and its components
656 (Kobayashi and Salam, 2000).

657 Phenology (sowing – flowering period) is accurately predicted (3.5%); further analysis
658 revealed that most of the error came from the sowing-emergence phase. Trends for
659 dynamically sampled variables (absorbed nitrogen, leaf area, RIE and biomass) over the crop
660 season were predicted with most of the error coming from LCS, indicating that the model
661 had neither a significant bias from measurements (low SB, 1 – 6% of mean square deviation)
662 nor a lack or excess of sensitivity (low SDSD). The outputs, more integrative variables (yield
663 and oil content), showed bias (~ 25%) but relative RMSE was in the standard range of crop
664 model performance (8-15%, Pereyra-Irujo and Aguirrezabal, 2007; Villalobos *et al.*, 1996).
665 Regression lines relating these two variables showed that the simulated range was narrower
666 than the observed one. Two groups of prediction confidence can be drawn showing good
667 prediction accuracy (RRMSE ~ 15%) for all variables except for LAI and absorbed nitrogen
668 which had RRMSE above 20%. No variable had a large prediction error, indicating that the
669 final output error results from poor prediction for all variables rather than error
670 compensation.

671

672 2. Model performance in a small-scale trial network

673 Our objective here was to evaluate the model in its usual application conditions, with regional
674 (instead of local) weather data or more uncertainty about soil analysis. Yield and oil content
675 were collected for 20 genotypes over 16 situations (locations x year). Genotypic variability
676 was wide as the genotypes' release date ranged from 1970 to 2000 in a uniform distribution
677 (Vear *et al.*, 2003). Two important model properties for genotype evaluation were analysed:
678 the prediction capacity (ability to predict quantitative values) and the ranking capacity
679 (ability to rank the genotypes) (**Figures 3-4**).

680 Considering all genotypes and trials, RMSEP was 0.45 t ha⁻¹ for yield and 3.7% for oil
681 content (RRMSEP were 16% and 9%, respectively); here values were slightly higher than for
682 the “development” dataset. Considering each trial individually, RRMSEP ranged from 9% to
683 30% for crop yield and from 3% to 14% for oil content, with a skewed distribution towards
684 well-predicted situations. Ranking capacity (Kendall's correlation coefficient) for each trial
685 was between 0.23 and 0.67 (mean=0.46) for yield, and between -0.1 and 0.75 (mean=0.54)
686 for oil content. Surprisingly, prediction capacity (RMSEP) and ranking capacity were not
687 correlated for yield (r=0.05, p=0.83) or for oil content (r=-0.04, p=0.86).

688 Different factors could explain the observed errors: genotypic factors (linked to genotypic
689 parameterization), environmental factors (linked to environment description or the inclusion
690 of stress in the model) and the interaction between these factors. To further unravel these
691 error factors, the model performance was evaluated on mean effects (**Figures 5-6**).

692 Prediction errors were clearly lower for mean effects when discarding G x E interactions (~
693 40% less error for yield). This analysis also pointed out that environmental effects were
694 stronger in reality than modelled: this was mostly the case with oil content where
695 environmental means RMSE showed a five-fold increase compared with genotypic means.
696 From now, we will focus the model evaluation on grain yield, as this is the most commonly
697 recorded variable. The yield E or G prediction error ($\sim 0.28 \text{ t.ha}^{-1}$) was still below the least
698 significant difference which discriminates between situations ($\text{LSD} = 0.40 \text{ t.ha}^{-1}$) or
699 genotypes ($\text{LSD} = 0.42 \text{ t.ha}^{-1}$). This indicated that the model was able to discriminate
700 between trials and genotypes when dealing with mean effects. There was also a trend
701 towards yield overestimation for the most productive genotypes and underestimation for the
702 least productive.

703 An analysis of variance on yield was done on actual and simulated trial networks (**Table 4**)
704 to quantify the G x E contribution to total yield variability. All tested main effects were
705 highly significant whether dealing with observed or simulated data. Interaction effects were
706 not tested (but were calculated) as there were no replications in the simulated data.

707 Mean deviations (Denis and Vincourt, 1982) of the effects were calculated for observed
708 yields: the environmental effect was the biggest (0.71 t.ha^{-1} between trials) followed by the
709 genotypic effect (0.45 t.ha^{-1} between genotypes) and G x E interactions (0.27 t.ha^{-1}).
710 Although the G x E interaction should not be neglected (10% of the overall mean), this
711 network presented a relatively weak G x E effect (3% of G effect) compared with other
712 results on sunflower network trials (Foucteau *et al.*, 2001; de la Vega and Chapman, 2001),
713 in which the G x E effect was between 11% and 39% of the G effect .

714 In the simulated network, mean sum of squares (MS) was about 60% of the observed one,
715 i.e. the crop model predicted less yield variance than actually occurred. Nevertheless, the
716 ranking of the ANOVA factors was identical in both networks with the same MS ratios
717 (**Table 4**). The G x E effect, which was in fact low in the field, was predicted to be 9-fold
718 smaller by the simulation (0.1%).

719 Finally, genotypic stability was computed to investigate whether the same genotypes were
720 subject to G x E interactions in the real and simulated network. Stability was plotted against
721 performance (**Figure 7**) to identify “high performance, low variability” response in the

722 genotype list. Model simulations strongly linked variability and performance ($r=0.81$,
723 $p<0.001$), but this was not the case in reality ($r=-0.05$, $p>0.8$) where no relation was found.
724 Nevertheless a group of four genotypes (Melody, LG5660, Prodisol, Allstar) could be picked
725 out, based on their performance. In both plots, these genotypes outperformed the others but
726 their variability was among the lowest in the actual network and relatively higher in the
727 simulated network. Thus the model helped more to discriminate genotypes due to their
728 intrinsic performance rather than their sensitivity to environmental effects.

729

730 **DISCUSSION**

731 The “model performance” in itself is not an easy term to define, so improving it can be
732 difficult. For statistical models, the error level calculation is a systematic and important part
733 of the analysis. A confidence interval is often estimated and associated with the model but
734 this approach is not easily applied to crop models. The general method is to compare, in
735 different situations (weather patterns, soils), the model predictions and crop observations to
736 estimate a mean error level. But this method is not ideal. First, the model has often been fitted
737 to the data (through parameter estimation); in this case, the measured error level corresponds
738 to the fitting error and underestimates the prediction error. The representativeness or
739 reliability of observed data is rarely discussed, although the error level can vary widely in
740 situations where the model is used. In this section, we will discuss the model’s use-cases in
741 relation with its performance.

742 The SUNFLO crop model can simulate variations in genotypic performance between
743 different environments (GEM interactions). These interactions played a significant part in
744 yield variability in both the actual and simulated network although they were higher in the
745 actual network. Such results raise several methodological points for discussion concerning
746 the relevance of a dynamic crop model for genotype evaluation, the limitations of the
747 modelling approach and finally the prospects for the use of the model.

748 *1. Does SUNFLO succeeded in evaluating a range of sunflower genotypes?*

749 The ability of the model to rank genotypes (relative performance) was independent from its
750 absolute predictive quality which means that the model can still be used even if simulation
751 biases are detected.

752 *Ranking performance.* The ranking capacity of SUNFLO is mainly due to an appropriate
753 description of the phenotypic variability among genotypes and to a correct parameterization
754 (statistical or measurement). Half of the genotypic parameters were related to resource
755 acquisition (leaf area) or resource management (water response parameters) processes, which

756 play a significant role in the environmental conditions in which sunflower is grown. But this
757 overview of crop behaviour through genotypic parameters is incomplete, as a recent study
758 showed extra phenotypic variability in rooting depth and water extraction among commercial
759 sunflower hybrids (Guilioni *et al.*, 2008). Extending the crop description to the rooting
760 system would probably improve the model's capacity to generate more complex interactions
761 in dry environments.

762 *Absolute prediction performance.* In addition, improving the model's capacity to predict
763 absolute yield is a more demanding objective as it depends on the quality of input data and
764 observed values as well as on the nature of the dataset (for instance, the relative importance
765 of yield-limiting factors not simulated by the model). The model prediction error (15%, 0.45
766 t.ha⁻¹) when estimated on a dataset was however within the range of the other sunflower crop
767 models although the input variables lacked site-specific values. This error remained very
768 close to the least significant difference or between genotypes in microplots but the list of
769 genotypes covered a long period of genotype release (30 years) and potential yield range.
770 This error would probably be too large to discriminate between genotypes from the same
771 release period because of their similar performance level (~ 0.2 t.ha⁻¹).

772 The model performance strongly relies on the quality of the dataset used for evaluation (input
773 data or field assessments). Poor reliability of input data may be attributable to weather being
774 recorded too far from the experimental location, from a rough estimation of available soil
775 water content, or, sometimes, from omissions in the components of crop management (timing
776 and amounts). The prediction error was doubled at the less reliable sites of the network (data
777 not shown). The different sampling methods used in the experimental network and the soil
778 variability within each trial resulted in different yield or precision levels, with direct
779 consequences for the evaluation of model performance.

780 As we evaluated the model in multi-environment trials used for official cultivar registration,
781 we accepted a default environmental characterization. Trials on experimental stations would
782 have facilitated model evaluation by giving a more complete and reliable environment
783 description, but the number of locations would have been limited. In spite of its limitations,
784 the users' network was a good opportunity for testing the model's robustness in real
785 conditions.

786 2. *What limitations result from the modelling options?*

787 Genotypic variability was better simulated than environmental variability (**Figure 2**). This
788 may be due to the model's structure (internal limitations) and from processes not considered

789 by the model (external limitations).

790 *Internal limitations.* Modelling the action of combined environmental stresses (temperature,
791 light, water, nitrogen) on a given biological variable is a conceptual problem to which little
792 attention has been paid. Separating the effect of stresses is an oversimplification of field
793 reality which cannot be observed on a daily basis. Experimenting in controlled
794 environments (greenhouse, CE chambers) helps to separate stresses, but plant growth is
795 suboptimal and population effects are lost. A common solution to this problem is to create
796 stress scalars which are multiplicative or to assume a hierarchy of stresses (the main limiting
797 factor concept). These approaches can lead to an erratic sensitivity of the model:
798 multiplicative solutions tend to overestimate stress effects while ordering constraints using
799 min/max functions introduces threshold effects into the simulation. In our approach, the
800 model slightly overestimated crop performance when using multiplicative stresses, which
801 might lead to bias if new stresses are to be included in the future.

802 Environmental stress modelling can also be viewed as a trade-off between a mechanistic and
803 an empirical integration. In the first case, modelling the mechanisms can be complex: for
804 example, assuming an acceleration of leaf senescence under drought stress may result in a
805 lower light interception and a lower grain yield. On the other hand, the need for simplicity
806 leads one to adopt a more implicit and integrative view whereby drought stress can directly
807 affect the harvest index (instead of active leaf area) and thus yield. It is reasonable to believe
808 that the first solution should generate more realism in plant-environment interactions while
809 demanding more physiological insight.

810 We finally decided to mix mechanistic and statistical approaches to deal with highly
811 integrative variables such as harvest index (HI) or oil content (OC). These variables were
812 modelled at harvest using a simple statistical model dependent on covariables previously
813 simulated by the mechanistic part of the crop model throughout the growing season. But by
814 doing so, feedback effects between different dynamic variables of the model and static
815 output variables cannot be taken into account. However, this statistical solution made it
816 possible to account for environmental variability in the prediction of HI and OC and the
817 large dataset used for parameterization conferred some robustness to the prediction of these
818 final variables.

819 *External limitations.* Modelling the detrimental effects of plant diseases (phoma,
820 phomopsis, sclerotinia, mildew) is a bottleneck to any progress in the prediction of yield and
821 oil content in sunflower fields. These diseases are difficult to control by crop management;

822 plant susceptibility is largely genetic. In 2000 and 2001, in the multi-environmental trials,
823 we can be sure that diseases were responsible for some discrepancies between simulated and
824 observed yields. This problem was indirectly brought to light by analysing the contribution
825 of environments (locations x year) to G by E interactions (environmental ecovalence,
826 Becker and Leon, 1988). In the actual network, two kinds of environment were subject to
827 strong interactions: (1) low-stress environments ; (2) water-stressed environments. In the
828 simulated network, the interactions were just apparent in the low-stress environments,
829 probably because of the lack of information about pathogens, which are often more severe
830 in dense canopies resulting from deep soils or wet conditions. On the other hand, the G x E
831 interaction was well predicted in water-stressed environments, as the main components of
832 the response of sunflower to water availability were included in the model.

833 The validity domain of SUNFLO includes environments where fluctuating water and
834 nitrogen levels occur due to combinations of weather patterns and soils in western Europe.
835 Photoperiodic effects were not modelled, leading to a possible bias in the prediction of
836 anthesis date in other geographical areas.

837

838 *3. SUNFLO, a model for engineering in the domain of varietal evaluation ?*

839 SUNFLO has several characteristics of an engineering model as defined by Passioura (1996):
840 ease of use and extensibility (new genotypes) of the parameterization based on
841 measurements, input data easily available from soil and weather records.

842 The field determination of the genotypic parameters requires potential growth conditions but
843 from different experiments: maximal LAI will be obtained in non-limiting conditions before
844 anthesis, maximal HI requires low vegetative growth before anthesis but sufficient water
845 during grain filling; oil content will be maximized under low nitrogen but well-watered
846 conditions. To cope with the regular appearance of new genotypes, these parameterization
847 experiments together with greenhouse measurements could be optimised to take advantage of
848 official genotype assessment trials.

849 From its capacity to evaluate the consequences on yield of different variety types defined by
850 a limited set of parameters, the model might be used for various applications in the field of
851 varietal evaluation (selection of the best variety x management combinations and evaluation
852 of ideotypes) or environmental diagnosis (detection of G x E interactions, identification of
853 stress patterns). If the model should be used during breeding, for variety registration testing
854 and development, more emphasis should be given on reducing uncertainty on inputs (genetic

855 parameters, soil and climate characterization)

856

857 **CONCLUSION**

858 The inclusion of genotypic traits within a crop model succeeded in differentiating the
859 response of a range of varieties grown in a wide range of environmental conditions (soil,
860 climate). G by E interactions were reproduced by dynamic simulation with the E term
861 resulting from crop management acting at the resource level. To have a model useful for
862 newly registered varieties, all the genotypic parameters were meant to be measurable. The
863 model performance was evaluated by using independent data representative of practical
864 application rather than specific research trials. This evaluation showed that both the model
865 framework and the uncertainty on inputs resulting from practical application limits early
866 discrimination of close-performing cultivars.

867

868 **ACKNOWLEDGEMENTS**

869 The authors are grateful to P. Vincourt and D. Wallach for statistical advisory.

870 I am indebted to A. Scaife for his help with the English in this article.

871

Captions

872

873

874 **Table 1.** Characteristics of the experiments used in the model development. Trial number,
875 site name and geographical location, year, sowing and harvest date, number of genotypes,
876 total N fertilization and irrigation, soil average water content usable by the crop, total
877 precipitation, climatic water deficit (precipitation – PET), mean temperature and incident
878 radiation sum. Integrative variables (sums, means) were computed for the whole growth
879 period. Trials #1-40 were partly used to estimate some model parameters while the remaining
880 data were used to estimate the model prediction capacity on a larger genotypic pool.

881

882 **Table 2.** Variables and parameters used as covariables (upper table) in statistical models of
883 harvest index and oil content (lower).

884

885 **Table 3.** Analysis of model prediction for 7 variables based on mean squared deviation
886 (MSD). Statistical criteria are root mean square of error (RMSE) and RMSE as the fraction
887 of the observed mean (Relative RMSE) ; bias (SB), squared difference between standard
888 deviations (SDSD), lack of correlation weighted by the standard deviations (LCS) for the
889 components of MSD.

890

891 **Table 4.** ANOVA tables of yield for actual and simulated trial networks. The ANOVA model
892 is a two way (Genotype, Environment) with interactions (G x E) model. The last column (%
893 MS) is the relative contribution of the effect to the mean square sum.

894

895 **Figure 1.** Chaining for different modules defining yield and oil production. Crop growth,
896 displayed in the centre part, is viewed as the interaction between environmental and
897 management-related limiting factors (left part) and genotypic information (right).
898 Intermediate variables appear in modules (rectangles) with references to equations in the text.
899 The schematic modules reflect the paragraphs' structure in the “model structure” section of
900 the text. Parallelograms (parameters) and ellipses (variables) represent model inputs.

901

902 **Figure 2.** Evaluation of the fit of the model. Estimated Vs observed data for phenology
903 (anthesis, date of year), nitrogen budget (absorbed N, kg.ha⁻¹), leaf area index, radiation
904 interception efficiency (%), aerial dry matter (t.ha⁻¹), achene yield (t.ha⁻¹) and oil content
905 (%). Absorbed nitrogen, leaf area, RIE (Radiation Interception Efficiency) and biomass were

906 sampled dynamically along the crop cycle. Filled symbols correspond to data used for
907 parameter estimation ; solid line represent the 1:1 line ; dotted line is a regression line on all
908 data ; model efficiency and RMSE (calculated on all data) are displayed in upper-left corner
909 of the plots.

910

911 **Figure 3.** Achene yield prediction for 20 genotypes at 16 sites. Prediction capacity was
912 assessed quantitatively through RMSEP and qualitatively with Kendall's correlation
913 coefficient (ranking). The solid line represent the 1:1 line ; the dotted line is a regression line
914 on all data.

915

916 **Figure 4.** Achene oil concentration prediction for 20 genotypes at 16 sites. Prediction
917 capacity was assessed quantitatively through RMSEP and qualitatively with Kendall's
918 correlation coefficient (ranking). Solid line represent the 1:1 line ; dotted line is a regression
919 line on all data.

920

921 **Figure 5.** Predicting mean environmental effects. Model performance was evaluated by
922 RMSEP and Efficiency for the two output variables.

923

924 **Figure 6.** Predicting mean genotypic effects. Model performance was evaluated by RMSEP
925 and Efficiency for the two output variables.

926

927 **Figure 7.** Variability vs Performance plot for the genotype set in the actual and simulated
928 networks. Variability was calculated as the environmental variance (S^2 , t.ha⁻¹) [Becker1981,
929 Piepho1998] and performance is the mean yield for each genotype.

930

931 **References**

- 932 Aguirrezabal, L.A., Lavaud, Y., Guillermo, A.A., Izquierdo, N.G., Andrade, F.H., Gonzalez, L.M.,
933 2003. Intercepted solar radiation during seed filling determines sunflower weight per seed and oil
934 concentration, *Crop Sci.* 43, 152-161.
- 935 Aiken, R.M., 2005. Applying thermal time scales to sunflower development, *Agron. J.* 97, 746-754.
- 936 Angus, J., Cunningham, R., Moncur, M., Mackenzie, D., 1981. Phasic development in field crops. I.
937 Thermal response in the seedling phase, *Field Crops Res.* 3, 365-378.
- 938 Becker, H.C., Leon, J., 1988. Stability analysis in plant breeding, *Plant Breed.* 101, 1-23.
- 939 Boote, K.J., Jones, J.W., Batchelor, W.D., Nafziger, E.D., Myers, O., 2003. Genetic coefficients in
940 the CROPGRO-soybean model: links to field performance and genomics, *Agron. J.* 95, 32-51.
- 941 Borrell, A.K., Hammer, G.L., 2000. Nitrogen dynamics and the physiological basis of stay-green in
942 sorghum, *Crop Sci.* 40, 1295-1307.
- 943 Brisson, N., Gary, C., Justes, E., Roche, R., Mary, B., Ripoche, D., Zimmer, D., Sierra, J., Bertuzzi,
944 P., Burger, P., Bussi re, F., Cabidoche, Y. M., Cellier, P., Debaeke, P., Gaudill re, J. P., H nault, C.,
945 Maraux, F., Seguin, B. & Sinoquet, H., 2003. An overview of the crop model STICS, *Eur. J. Agron.*
946 18(3-4), 309-332
- 947 Brisson, N.; Launay, M.; Mary, B. & Beaudoin, N. ,eds, 2008. Conceptual Basis, Formalisations and
948 Parameterization of the STICS Crop Model, Editions Quae, Versailles, 304p.
- 949 Casadebaig, P., 2008. Analyse et mod lisation de l'interaction G notype - Environnement - Conduite
950 de culture : application au tournesol (*Helianthus annuus* L.), PhD thesis, Universit  de Toulouse.
- 951 Casadebaig, P., Debaeke, P., Lecoeur, J., 2008. Thresholds for leaf expansion and transpiration
952 response to soil water deficit in a range of sunflower genotypes, *Eur. J. Agron.* 28, 646-654.
- 953 CETIOM, 2004. Stades-rep res du tournesol, in CETIOM (Centre d'Etudes Techniques
954 Interprofessionnel des Ol oprot agineux M tropolitain), ed., Guide de l'exp rimentateur tournesol,
955 pp. 13-15.
- 956 Chapman, S., Cooper, M., Hammer, G., 2002. Using crop simulation to generate genotype by
957 environment interaction effects for sorghum in water-limited environments. *Aust. J. Agric. Res.* 53,
958 379-389.
- 959 Chapman, S., Hammer, G., Meinke, H., 1993. A sunflower simulation model: I. model development,
960 *Agron. J.* 85, 725-735.
- 961 Colson, J., Bouniols, A., Jones, J., 1995. Soybean reproductive development: adapting a model for
962 European cultivars, *Agron. J.* 87, 1129-1139.
- 963 Connor, D., Hall, A., (Eds), 1997. Sunflower physiology, *Sunflower Technology and Production.*

964 American Society of Agronomy, 35, 67-113.

965 Debaeke, P., Raffailac, D., 2006. Normalized SPAD index and Nitrogen Nutrition Index (NNI): two
966 indicators of plant N status for sunflower crop, in Proceedings of 9th ESA Congress, Varsovie
967 (Pologne), 4-7 Sept.2006, pp. 83-84.

968 Debaeke, P., Triboi, A.M., Vear, F., Lecoecur, J., 2004. Crop physiological determinants of yield in
969 old and modern sunflower hybrids, Proceedings of the 16th International Sunflower Conference,
970 Fargo (USA), pp. 267-273.

971 Denis, J., Vincourt, P., 1982. Panorama des méthodes statistiques d'analyse des interactions
972 génotype X milieu, *Agronomie* 2, 219-230.

973 Foucteau, V., El Daouk, M., Baril, C., 2001. Interpretation of genotype by environment interaction
974 in two sunflower experimental networks, *Theor. App. Genet.* 102, 327-334.

975 Granier, C., Tardieu, F., 1998. Is thermal time adequate for expressing the effects of temperature on
976 sunflower leaf development?, *Plant Cell Env.* 21, 695-703.

977 Guilioni, L., Radanielson, A., Christophe, A., Lecoecur, J., 2008. Root system and water extraction
978 variability for sunflower hybrids, Proceedings of the 17th International Sunflower Conference,
979 Cordoba (Spain).

980 Hammer, G. and Broad, I. J., 2003. Genotype and Environment Effects on Dynamics of Harvest
981 Index during Grain Filling in Sorghum, *Agronomy Journal* 95, 199-206.

982 Hammer, G., Chapman, S., van Oosterom, E., Podlich, W., 2005. Trait physiology and crop
983 modelling as a framework to link phenotypic complexity to underlying genetic systems, *Aust. J.*
984 *Agric. Res.* 56, 947-960.

985 Hammer, G., Cooper, M., Tardieu, F., Welch, S., Walsh, B., van Eeuwijk, F., Chapman, S., Podlich,
986 D., 2006. Models for navigating biological complexity in breeding improved crop plants, *Trends*
987 *Plant Sci.* 11, 587-593.

988 Horie, T., 1977. Simulation of sunflower growth: I. Formulation and parameterisation of dry matter
989 production, leaf photosynthesis, respiration and partitioning of photosynthesis, *Bull. Natl. Inst.*
990 *Agric. Sci. (Japan.)*, Ser. A, 24, 45-70.

991 Jeuffroy, M., Barbottin, A., Jones, J., Lecoecur, J., 2006. Crop models with genotype parameters. In
992 Wallach, D., Makowski, D. and Jones, J.W., ed. Working with dynamic crop models: Evaluation,
993 analysis, parameterization and applications, Elsevier, pp. 281-307.

994 Jones, H., 2004. Water use efficiency in plant biology, Blackwell Pub., Oxford.

995 Jones, J.W., Keating, B.A., Porter, C.H., 2001. Approaches to modular model development, *Agric.*
996 *Syst.* 70, 421-443.

997 Keating, B.A., Wafula, B.M., 1992. Modelling the fully expanded area of maize leaves, *Field Crops*

998 Res. 29, 163-176.

999 Kiniry, J., Blanchet, R., Williams, J., Texier, V., Jones, C., Cabelguenne, M., 1992. Sunflower
1000 simulation using the EPIC and ALMANAC models, *Field Crops Res.* 30, 403-423.

1001 Kobayashi, K., Salam, M., 2000. Comparing simulated and measured values using mean squared
1002 deviation and its components, *Agron. J.* 92, 345-352.

1003 Lecoeur, J., Sinclair, T.R., 1996. Field pea transpiration and leaf growth in response to soil water
1004 deficit, *Crop Sci.* 36, 331-335.

1005 Lecoeur, J., Poiré-Lassus, R., Christophe, A., Casadebaig, P., Debaeke, Ph., Vear, F. & Guilioni, L.,
1006 2009. SUNFLO: A joint phenotyping and modelling approach to analyse and predict the differences
1007 in yield potential of sunflower genotypes. Submitted to *Functional Plant Biology* FP09186

1008 Lemaire, G., Meynard, J., 1997. Use of the nitrogen nutrition index for the analysis of agronomical
1009 data, Springer-Verlag, Berlin, pp. 45-55.

1010 Leon, A. J., Lee, M., Andrade, F.H., 2001. Quantitative trait loci for growing degree days to
1011 flowering and photoperiod response in Sunflower (*Helianthus annuus L.*), *Theor.Appl.Genet.* 102,
1012 497-503.

1013 Limaux, F., Recous, S., Meynard, J.M., Guckert, A., 1999. Relationship between rate of crop growth
1014 at date of fertiliser N application and fate of fertiliser N applied to winter wheat, *Plant Soil* 214, 49-
1015 59.

1016 Lizaso, J.I., Batchelor, W.D., Westgate, M.E., 2003. A leaf area model to simulate cultivar-specific
1017 expansion and senescence of maize leaves, *Field Crops Res.* 80, 1-17.

1018 Mary, B., Beaudoin, N., Justes, E., Machet, J., 1999. Calculation of nitrogen mineralization and
1019 leaching in fallow soil using a simple dynamic model, *Eur. J. Soil Sci.* 50, 549-566.

1020 Mavromatis, T., Boote, K., Jones, J., Irmak, A., Shinde, D., Hoogenboom, G., 2001. Developing
1021 genetic coefficients for crop simulation models with data from crop performance trials, *Crop Sci.* 41,
1022 40-51.

1023 Merrien, A., Grandin, G., 1990. Comportement hydrique du tournesol, In : Blanchet, R., Merrien,
1024 A., ed., *Le tournesol et l'eau : adaptation à la sécheresse, réponse à l'irrigation*, Points science du
1025 CETIOM, Paris, 75-90.

1026 Messina, C., Boote, K., Loffler, C., Jones, J., Vallejos, C., 2006. Model-assisted genetic
1027 improvement of crops, In: *Working with dynamic crop models: Evaluation, analysis,*
1028 *parameterization, and applications*, 309-335.

1029 Miflin, B., 2000. Crop improvement in the 21st century, *J. Exp. Bot.* 51, 1-8.

1030 Monsi, M., Saeki, T., 2005. On the factor light in plant communities and its importance for matter
1031 production, *Ann. Bot.* 95, 549-567.

1032 Monteith, J.L., 1977. Climate and the efficiency of crop production in Britain', *Phil. Trans. Royal*
1033 *Soc. London, Series B, Biological Sciences* 281, 277-294.

1034 Moot, D.; Jamieson, P.; Henderson, A.; Ford, M. & Porter, J., 1996. Rate of change in harvest index
1035 during grain-filling of wheat, *Journal of Agricultural Science* 126, 387-395.

1036 Passioura, J.B., 1996. Simulation models: science, snake oil, education, or engineering?, *Agron. J.*
1037 88, 690-694.

1038 Penning de Vries, F., Jansen, D., ten Berge, H., Bakema, A., 1989. Simulation of ecophysiological
1039 processes of growth in several annual crops, *Simulation Monographs*.

1040 Pereyra-Irujo, G.A., Aguirrezabal, L.A., 2007. Sunflower yield and oil quality interactions and
1041 variability: Analysis through a simple simulation model, *Agric. For. Met.* 143, 252-265.

1042 Pereyra-Irujo, G.A., Izquierdo N.G, Covi M., Nolasco S.M, Quiroz F., Aguirrezabal L.A.N, 2009.
1043 Variability in sunflower oil quality for biodiesel production: A simulation study. *Biomass &*
1044 *Bioenergy* 33, 459-468

1045 Picq, G., 1988. Rayonnement solaire intercepté par les couverts des cultures de tournesol, facteur
1046 explicatif de la production, *Proc. 12th Int. Sunflower Conf., ISA, Novi-Sad (Yugoslavia)*, 81-85.

1047 R Development Core Team, 2006. R: A language and environment for statistical computing, R
1048 Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

1049 Rawson, H., Hindmarsh, J., 1983. Light, leaf expansion and seed yield in sunflower, *Aust. J. Plant*
1050 *Physiol.* 10, 25-30.

1051 Rey, H., 2003. Utilisation de la modélisation 3D pour l'analyse et la simulation du développement et
1052 de la croissance végétative d'une plante de tournesol en conditions environnementales fluctuantes
1053 (température et rayonnement)., PhD thesis, ENSA Montpellier.

1054 Rey, H., Dauzat, J., Chenu, K., Barczi, J.F., Dosio, G.A.A., Lecoeur, J., 2008. Using a 3-D virtual
1055 sunflower to simulate light capture at organ, plant and plot levels: contribution of organ interception,
1056 impact of heliotropism and analysis of genotypic differences, *Ann. Bot.* 101, 1139-1152.

1057 Sadras, V., Hall, A., 1988. Quantification of temperature, photoperiod and population effects on
1058 plant leaf area in sunflower crops, *Field Crops Res.* 18, 185-196.

1059 Sadras, V.O., Echarte, L., Andrade, F.H., 2000. Profiles of leaf senescence during reproductive
1060 growth of sunflower and maize, *Ann. Bot.* 85, 187-195.

1061 Sarr, B., Lecoeur, J., Clouvel, P., 2004. Irrigation scheduling of confectionery groundnut (*Arachis*
1062 *hypogaea* L.) in Senegal using a simple water balance model, *Agric. Water Manag.* 67, 201-220.

1063 Sinclair, T.R., 2005. Theoretical analysis of soil and plant traits influencing daily plant water flux on
1064 drying soils, *Agron. J.* 97, 1148-1152.

1065 Sinclair, T.R. & Muchow, R., 1995. Effect of nitrogen supply on maize yield. I: modelling
1066 physiological responses. *Agronomy Journal*, 87, 632-641.

1067 Sinclair, T.R., Purcell, L.C. & Sneller, C.H., 2004. Crop transformation and the challenge to increase
1068 yield potential. *Trends in Plant Science* 9, 70-75.

1069 Sinclair, T.R., Seligman, N., 2000. Criteria for publishing papers on crop modelling, *Field Crops*
1070 *Res.* 68, 165-172.

1071 Triboi, E. & Triboi-Blondel, A.-M., 2002. Productivity and grain or seed composition: a new
1072 approach to an old problem--invited paper, *European Journal of Agronomy* 16, 163-186.

1073 Slafer, G.A., 2003. Genetic basis of yield as viewed from a crop physiologist's perspective,
1074 *Ann.appl.Biol* 142, 117-128.

1075 Steer, B., Milroy, S., Kamona, R., 1993. A model to simulate the development, growth and yield of
1076 irrigated sunflower, *Field Crops Res.* 32, 83-89.

1077 Stewart, D., Dwyer, L., 1994. A model of expansion and senescence of individual leaves of field-
1078 grown maize (*Zea mays* L.), *Can. J. Plant Sci.* 74, 37-42.

1079 Stockle, C. & Debaeke, P., 1997. Modeling crop nitrogen requirements: a critical analysis, *European*
1080 *Journal of Agronomy.* 7, 161-169.

1081 Tao, Y.Z., Henzell, R.G., Jordan, D.R., Butler, D.G., Kelly, A.M., McIntyre, C.L., 2000.
1082 Identification of genomic regions associated with stay green in sorghum by testing RILs in multiple
1083 environments, *Theor.Appl.Genet.* 100, 1225-1232.

1084 Valé, M., Mary, B., Justes, E., 2007. Irrigation practices may affect denitrification more than
1085 nitrogen mineralization in warm climatic conditions, *Biol.Fertility Soils* 43, 641-651.

1086 Vear, F., Bony, H., Joubert, G., Tourvieille De Labrouhe, D., Pauchet, I. , Pinochet, X., 2003. 30
1087 years of sunflower breeding in France, *Oleagineux, Corps gras, Lipides* 10, 66-73.

1088 de la Vega, A.J., Chapman, S.C., 2001. Genotype by environment interaction and indirect selection
1089 for yield in sunflower: II. Three-mode principal component analysis of oil and biomass yield across
1090 environments in Argentina. *Field Crops Res.* 72, 39-50.

1091 Villalobos, F., Hall, A., Ritchie, J., Orgaz, F., 1996. OILCROP-SUN: a development, growth and
1092 yield model of the sunflower crop. *Agron. J.* 88, 403-415.

1093 Yin, X., Struik, P.C., Kropff, M. J., 2004. Role of crop physiology in predicting gene-to-phenotype
1094 relationships, *Trends Plant Sci.* 9, 426-432.

1095

Table 1 number	site	dep	year	sowing	harvest	genotype number	fert. (kg.ha-1)	irr. (mm)	AWC (mm)	precipitation (mm)	deficit (mm)	mean temp. (°C)	PAR sum (MJ.m-2)	label
01	Montpellier	34	2001	5-May	19-Sep	5	140	305	225	131	-553	21.4	3344	MTP.01
02	Montpellier	34	2002	15-May	13-Sep	5	120	125	225	290	-317	21.3	2950	MTP.02
03	Montpellier	34	2002	15-May	13-Sep	5	60	40	225	290	-317	21.3	2950	MTP.02
04	Auzeville	31	2004	9-Apr	7-Sep	2	120	55	225	236	-449	19.1	3312	AUZ.04
05	Auzeville	31	2003	16-Apr	23-Aug	13	60	133	225	85	-604	21.9	3053	AUZ.03
06	Auzeville	31	2002	16-Apr	12-Sep	8	25	20	225	296	-329	18.5	3071	AUZ.02
07	Auzeville	31	2002	16-Apr	12-Sep	8	70	20	225	296	-329	18.5	3071	AUZ.02
08	Auzeville	31	2002	16-Apr	12-Sep	8	70	60	225	296	-329	18.5	3071	AUZ.02
09	Auzeville	31	2005	12-Apr	15-Sep	2	0	125	225	237	-492	19.7	3365	AUZ.05
10	Auzeville	31	2005	12-Apr	15-Sep	2	40	125	225	237	-492	19.7	3365	AUZ.05
11	Auzeville	31	2005	12-Apr	15-Sep	2	80	125	225	237	-492	19.7	3365	AUZ.05
12	Auzeville	31	2005	12-Apr	15-Sep	2	120	125	225	237	-492	19.7	3365	AUZ.05
13	Auzeville	31	2005	12-Apr	15-Sep	2	80	125	225	237	-492	19.7	3365	AUZ.05
14	Auzeville	31	2005	12-Apr	15-Sep	2	160	125	225	237	-492	19.7	3365	AUZ.05
15	Auzeville	31	2006	19-Apr	12-Sep	2	0	219	225	115	-648	20.9	3356	AUZ.06
16	Auzeville	31	2006	19-Apr	12-Sep	2	50	219	225	115	-648	20.9	3356	AUZ.06
17	Auzeville	31	2006	19-Apr	12-Sep	2	100	219	225	115	-648	20.9	3356	AUZ.06
18	Auzeville	31	2006	19-Apr	12-Sep	2	150	219	225	115	-648	20.9	3356	AUZ.06
19	Auzeville	31	2006	19-Apr	4-Sep	2	0	50	225	115	-606	20.6	3201	AUZ.06
20	Auzeville	31	2006	19-Apr	4-Sep	2	50	50	225	115	-606	20.6	3201	AUZ.06
21	Auzeville	31	2006	19-Apr	4-Sep	2	100	50	225	115	-606	20.6	3201	AUZ.06
22	Auzeville	31	2006	19-Apr	4-Sep	2	150	50	225	115	-606	20.6	3201	AUZ.06
23	Le Magneraud	17	2005	18-Apr	8-Sep	2	60	35	70	184	-455	18.6	3233	MGN.05
24	Le Magneraud	17	1997	28-Mar	18-Sep	2	78	0	120	387	-207	17.2	3134	MGN.97
25	Le Magneraud	17	1997	18-Apr	18-Sep	2	78	0	120	387	-135	18.0	2728	MGN.97
26	Le Magneraud	17	1997	15-May	18-Sep	2	78	0	120	295	-144	19.1	2252	MGN.97
27	Le Magneraud	17	1997	28-Mar	18-Sep	2	78	100	120	387	-207	17.2	3134	MGN.97
28	Le Magneraud	17	1997	18-Apr	18-Sep	2	78	100	120	387	-135	18.0	2728	MGN.97
29	Le Magneraud	17	1997	15-May	18-Sep	2	78	100	120	295	-144	19.1	2252	MGN.97
30	Le Magneraud	17	1998	26-Mar	11-Sep	2	70	0	120	433	-114	16.6	2946	MGN.98
31	Le Magneraud	17	1998	22-Apr	11-Sep	2	70	0	120	205	-298	17.9	2619	MGN.98
32	Le Magneraud	17	1998	13-May	11-Sep	2	70	0	120	157	-287	18.6	2279	MGN.98
33	Le Magneraud	17	1998	26-Mar	17-Sep	2	70	195	120	433	-114	16.6	2946	MGN.98
34	Le Magneraud	17	1998	22-Apr	17-Sep	2	70	195	120	205	-298	17.9	2619	MGN.98
35	Le Magneraud	17	1998	13-May	17-Sep	2	70	225	120	157	-287	18.6	2279	MGN.98
36	Le Magneraud	17	1999	31-Mar	8-Sep	2	70	0	120	364	-149	17.6	2808	MGN.99
37	Le Magneraud	17	1999	22-Apr	8-Sep	2	70	0	120	286	-190	18.6	2532	MGN.99
38	Le Magneraud	17	1999	25-May	24-Sep	2	70	0	120	361	-61	19.6	2200	MGN.99
39	Le Magneraud	17	1999	31-Mar	24-Sep	2	70	110	120	502	-40	17.6	2983	MGN.99
40	Le Magneraud	17	1999	22-Apr	24-Sep	2	70	110	120	424	-81	18.6	2707	MGN.99
41	Auzeville	31	2001	17-Apr	11-Sep	20	51	0	225	293	-355	19.0	3220	AUZ.01
42	Auzeville	31	2000	22-Apr	29-Aug	20	121	81	225	292	-316	20.0	2864	AUZ.00
43	Levroux-Trégonce	36	2000	3-May	12-Sep	20	60	50	100	242	-302	17.8	2313	LEV.00
44	Lusignan	86	2000	5-May	2-Oct	20	60	0	270	287	-286	17.7	2950	LUS.00
45	Chaunay	86	2000	6-May	5-Oct	20	40	0	195	313	-254	18.3	2980	CAY.00
46	Levignac/Saves	31	2000	13-May	18-Sep	20	20	30	150	277	-370	20.7	2559	SAV.00
47	Montech	82	2000	22-Apr	11-Sep	20	60	0	270	286	-353	19.8	2863	MON.00
48	Oucques	41	2000	11-Apr	8-Sep	20	60	0	150	339	-180	16.7	2798	OUC.00
49	Gaillac	81	2000	25-Apr	1-Sep	20	40	60	120	253	-233	19.8	2407	GAL.00
50	Blois	41	2000	19-May	7-Oct	20	60	0	195	281	-226	17.0	2392	BGY.00
51	Longré	16	2000	23-May	24-Oct	20	60	0	105	356	-224	18.1	2702	LON.00
52	Puy Laroque	82	2001	12-May	21-Sep	20	60	35	270	261	-315	19.9	2742	PLR.01
53	Lempdes	63	2001	18-Apr	7-Sep	20	80	0	225	242	-333	17.4	2724	LEM.01
54	Gaillac	81	2001	25-Apr	5-Sep	20	40	60	120	328	-201	19.7	2664	GAL.01
55	Ruffec	16	2001	15-May	21-Sep	20	60	0	50	185	-351	18.5	2782	RUF.01
56	Fleurance	32	2001	2-Apr	27-Aug	20	60	0	105	312	-225	17.7	2899	FLE.01

Table 2	Type	Description	Unit
SIRF	climate	Sum of intercepted radiation after flowering	MJ
TTF	climate	Temperature Sum after Flowering	°C.d
SDV	stress	Stress Days during Vegetative development	d
SDF	stress	Stress Days during Flowering	d
SDM	stress	Stress Days during Maturity	d
TRF	stress	Transpired water after flowering	mm
NNIF	stress	Nitrogen nutrition Index at flowering	NA
TDMF	canopy	Biomass at flowering	g.m-2
LAD	canopy	Leaf Area Duration	d
IRP	genotype	Potential Harvest Index	NA
OCP	genotype	Potential Oil Content	%
TDF	genotype	Thermal Date of flowering	°C.d
k	genotype	Light Extinction Coefficient	NA

Harvest Index

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.370e-02	6.996e-02	1.339	0.182276
TDMF	-1.552e-04	6.376e-05	-2.434	0.015982 *
SDV	-2.828e-03	1.335e-03	-2.118	0.035650 *
SDF	-2.557e-03	1.174e-03	-2.178	0.030813 *
SDM	-1.940e-03	4.995e-04	-3.884	0.000148 ***
TRF	-3.907e-04	1.696e-04	-2.304	0.022464 *
TTF	1.274e-04	3.190e-05	3.992	9.80e-05 ***
IRP	8.189e-01	1.540e-01	5.317	3.34e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1				
Adjusted R-squared: 0.3036				

Oil Content

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	65.320223	19.856358	3.290	0.001204 **
SIRF	-0.244884	0.038339	-6.387	1.37e-09 ***
TTF	-0.009100	0.002769	-3.286	0.001218 **
SDM	-0.441561	0.052941	-8.341	1.79e-14 ***
LAD	-0.215173	0.031512	-6.828	1.24e-10 ***
NNIF	9.646419	3.357145	2.873	0.004544 **
TDF	-0.049909	0.012870	-3.878	0.000147 ***
k	30.973272	5.965681	5.192	5.53e-07 ***
OCP	1.266167	0.139389	9.084	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1				
Adjusted R-squared: 0.5749				

Table 3 Variables	RMSE	RRMSE (%)	SB (%)	SDSD (%)	LCS (%)
Sowing - Flowering (days)	2.7	3.5	0	0	100
Absorbed Nitrogen (kg/ha)	23.0	22.7	4	8	88
Leaf Area Index	0.6	31.0	1	0	99
Interception Efficiency	0.1	17.5	6	0	94
Dry Biomass (t/ha)	1.1	18.1	0	4	95
Yield (t/ha)	0.5	15.0	22	5	73
Grain Oil Content (%)	4.3	8.4	27	0	73

Table 4		Df	SS	MS	F	Pr(>F)	% MS
Observed	E	15	79.0	5.27	74.2	***	69.0
	G	19	43.7	2.30	32.4	***	30.1
	GxE	285	20.2	0.07			0.9
Simulated	E	15	50.2	3.35	693.9	***	71.1
	G	19	25.7	1.36	280.9	***	28.8
	GxE	285	1.4	0.01			0.1

Figure 1 Environment and Management

Crop Growth

Genotype

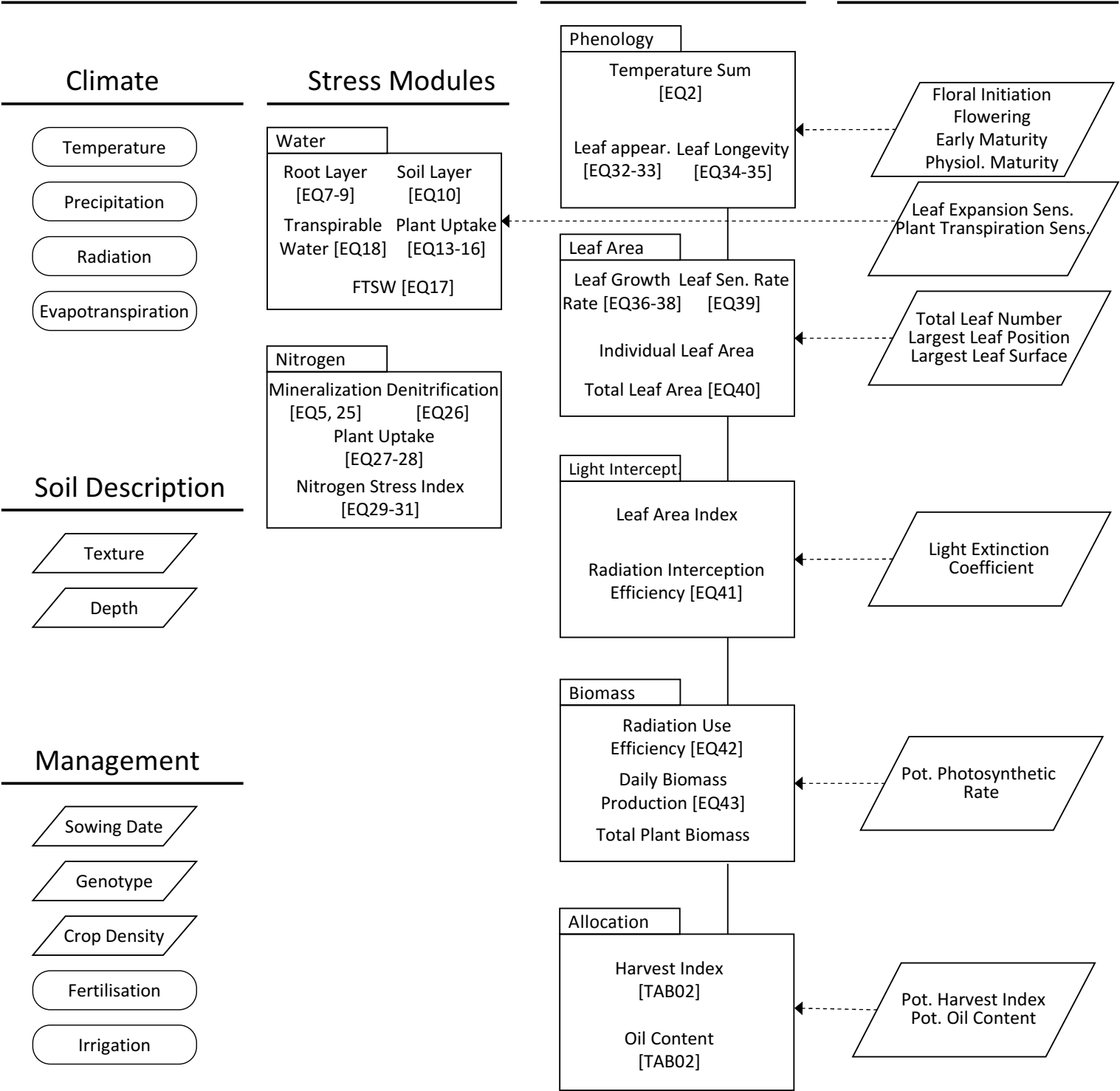


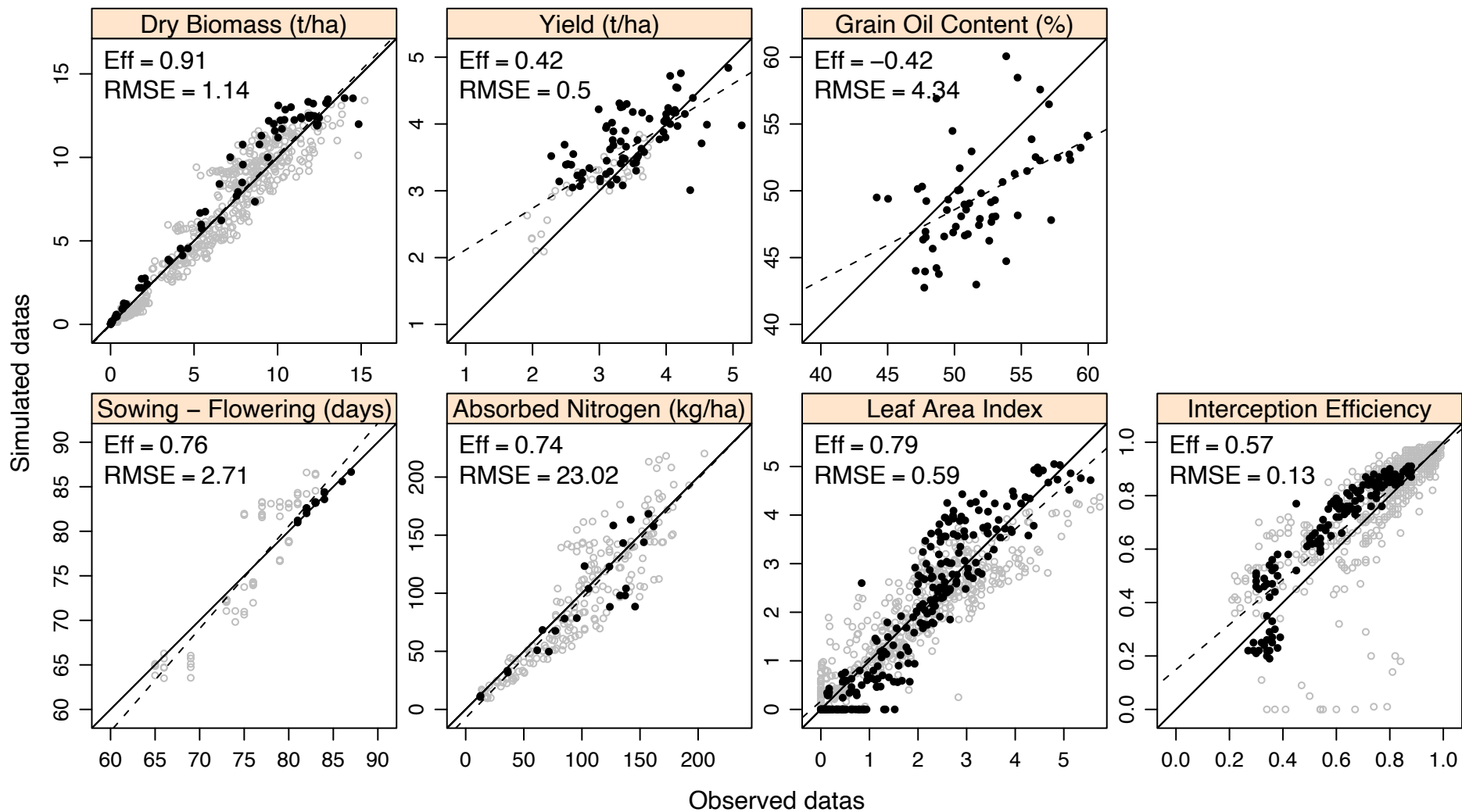
Figure 2

Figure 3

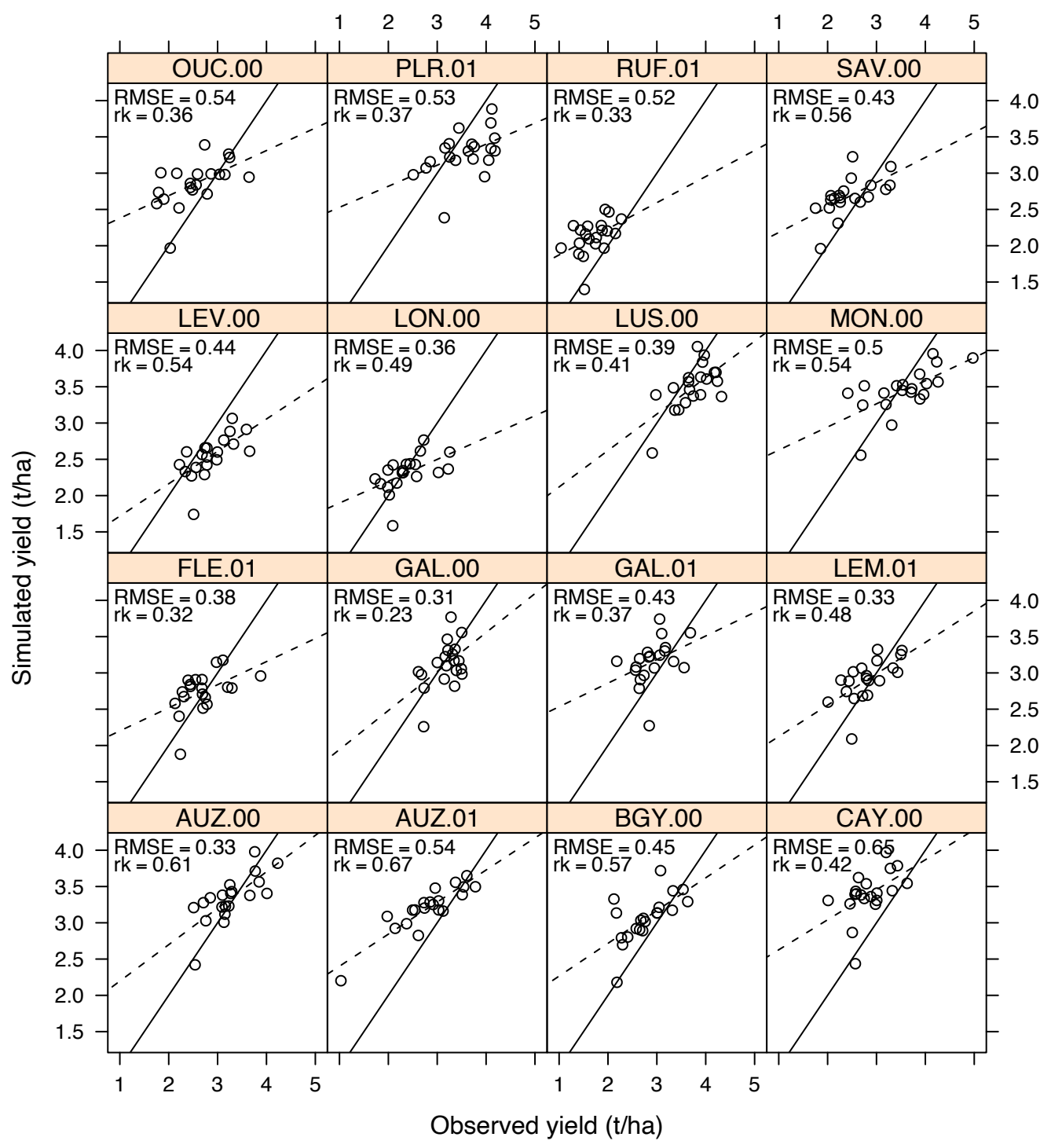


Figure 4

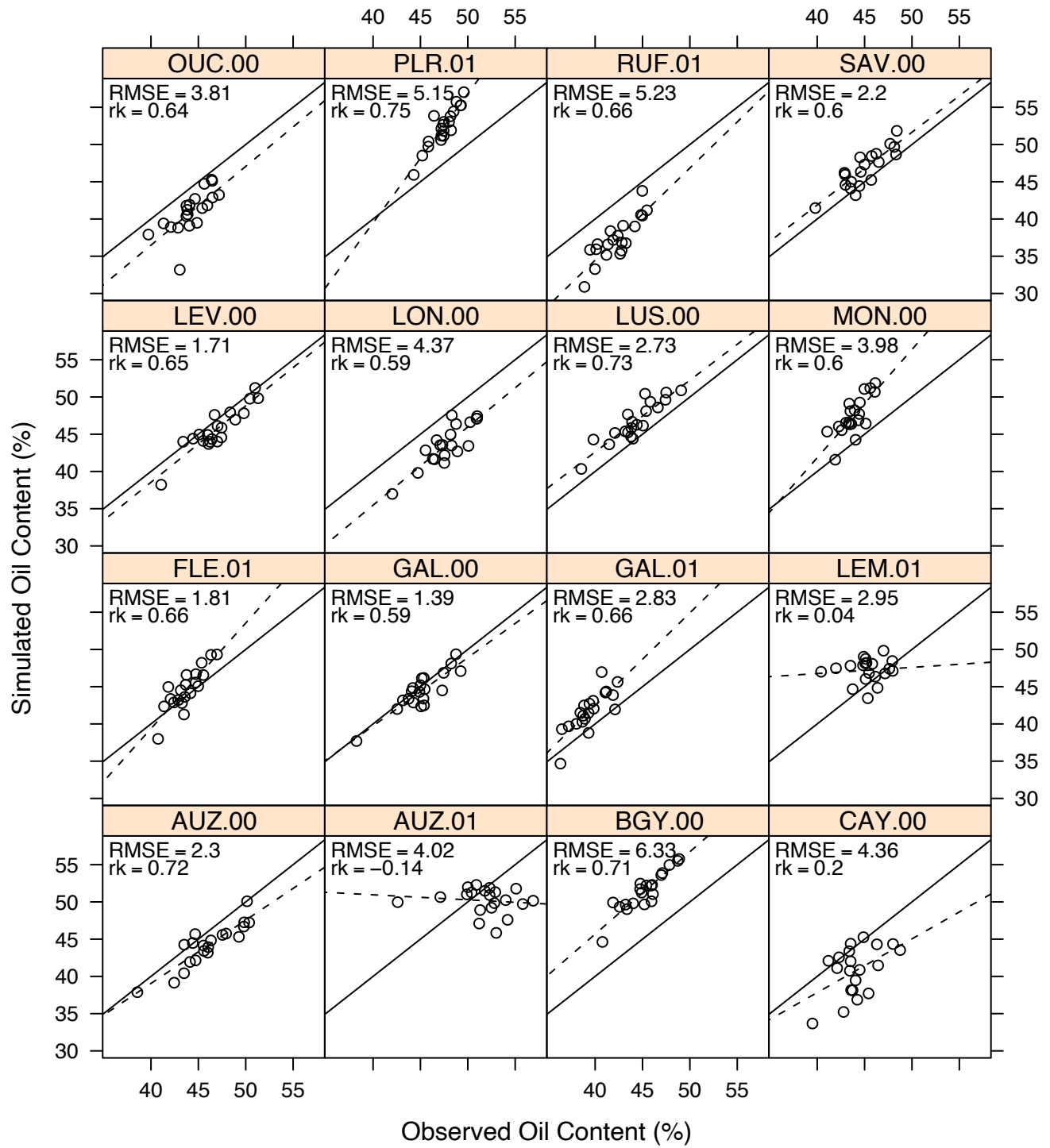


Figure 5

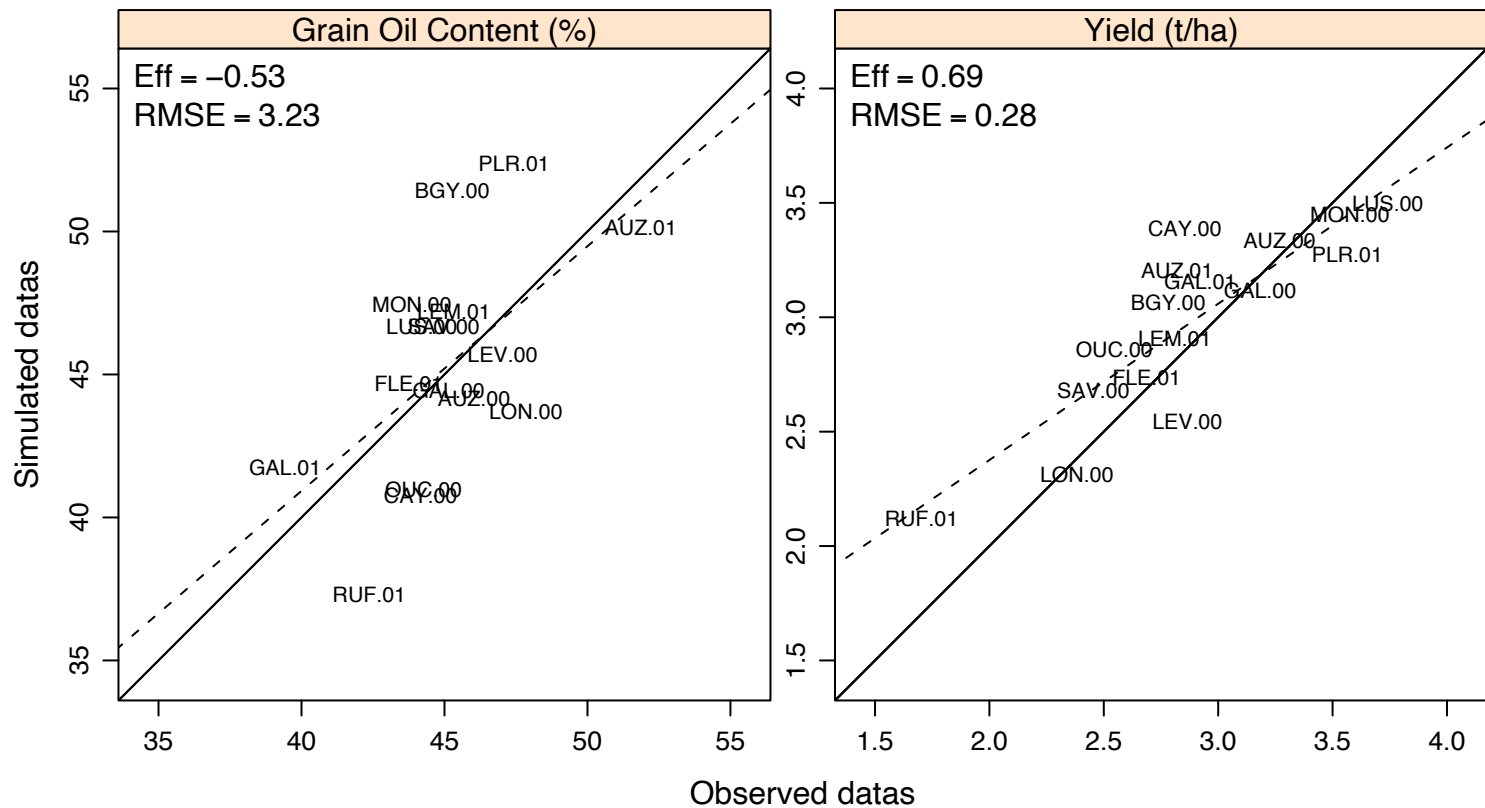


Figure 6

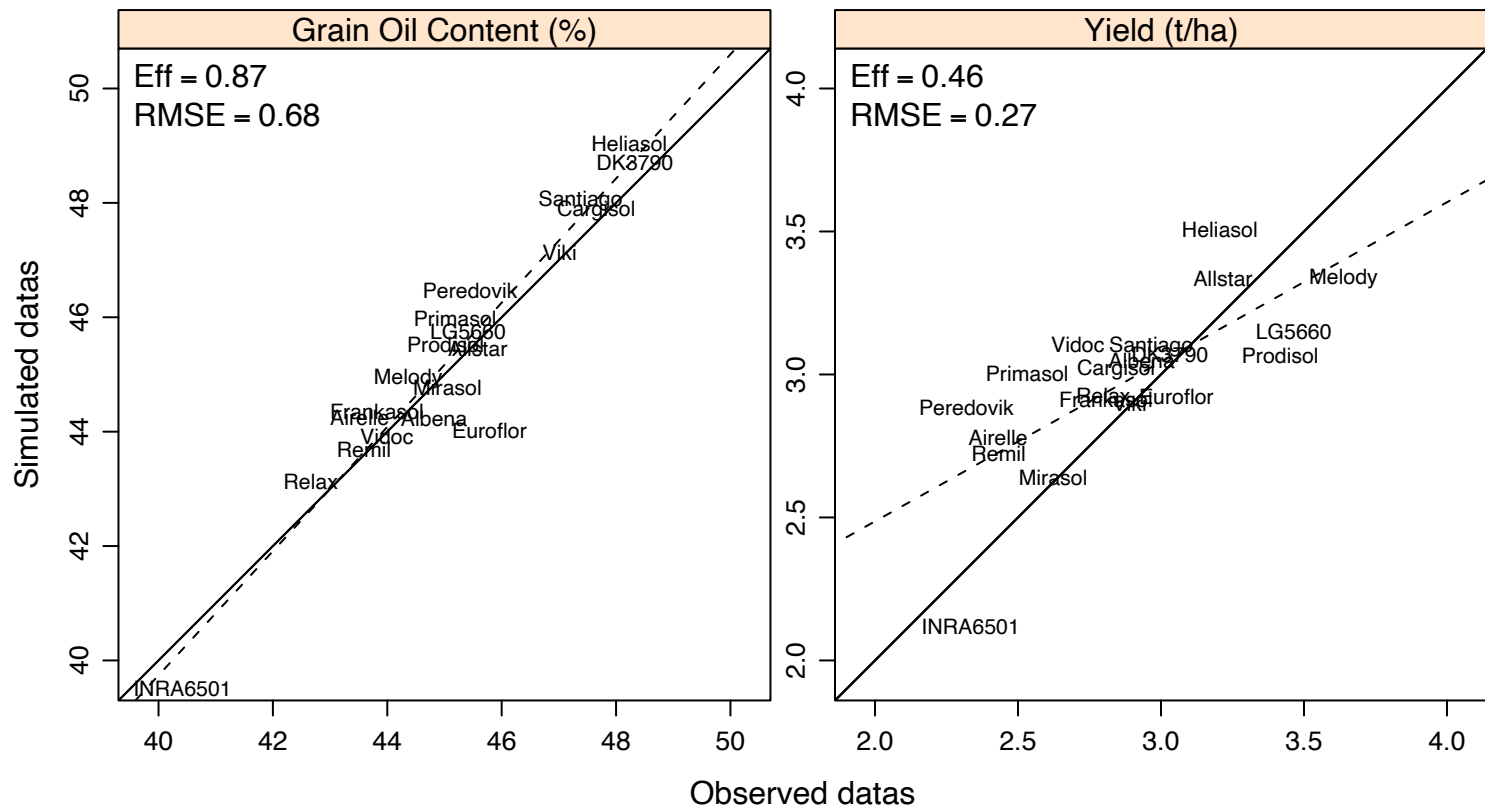


Figure 7

