



HAL
open science

Data-driven projections suggest large opportunities to improve Europe's soybean self-sufficiency under climate change

Nicolas Guilpart, Toshichika Iizumi, David Makowski

► **To cite this version:**

Nicolas Guilpart, Toshichika Iizumi, David Makowski. Data-driven projections suggest large opportunities to improve Europe's soybean self-sufficiency under climate change. *Nature Food*, 2022, 3, pp.255-265. 10.1038/s43016-022-00481-3 . hal-03626485

HAL Id: hal-03626485

<https://agroparistech.hal.science/hal-03626485v1>

Submitted on 31 Mar 2022

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.

1 Data-driven projections suggest large opportunities 2 to improve Europe's soybean self-sufficiency under 3 climate change

4 Nicolas Guilpart ^{1*}, Toshichika Iizumi ², David Makowski ³

5 ¹ Université Paris-Saclay, AgroParisTech, INRAE, UMR Agronomie, 78850, Thiverval-Grignon, France

6 ² Institute for Agro-Environmental Sciences, National Agriculture and Food Research Organization
7 (NARO), 3-1-3 Kannondai, Tsukuba, Ibaraki 305-8604, Japan

8 ³ Université Paris-Saclay, AgroParisTech, INRAE, UMR MIA-Paris, 75231, Paris, France

9 * Corresponding author: nicolas.guilpart@agroparistech.fr

10 **Abstract**

11 The rapid expansion of soybean-growing areas across Europe raises questions about the
12 suitability of agro-climatic conditions for soybean production. Here, using data-driven
13 relationships between climate and soybean yield derived from machine-learning, we made yield
14 projections under current and future climate with moderate (RCP 4.5) to intense (RCP 8.5)
15 warming, up to the 2050s and 2090s time horizons. The selected model showed high R^2 (higher
16 than 0.9) and low RMSE (0.35 t ha^{-1}) between observed and predicted yields based on cross-
17 validation. Our results suggest that a self-sufficiency level of 50% (100%) would be achievable in
18 Europe under historical and future climate if 4-5% (9-11%) of the current European cropland is
19 dedicated to soybean production. The findings could help farmers, extension services,
20 policymakers and agribusiness to reorganize the production area distribution. The
21 environmental benefits and side effects, as well as the impacts of soybean expansion on land-
22 use change, would need further research.

23 **Main text**

24 *Introduction*

25 The satisfaction of European* soybean demand is highly dependent on imports. Currently,
26 Europe imports about 58 Mt yr⁻¹ of soybean which accounts for nearly 90% of the domestic
27 consumption¹ (average over 2009-2013; Table S1). This large share of soybean imports in
28 Europe takes its roots in the post-World War II international trade agreements between Europe
29 and the USA that allowed tax-free entry of protein imports into Europe². Price support for
30 cereals cultivated within the European Economic Community led to a strong growth of cereal
31 production in Europe at the expense of grain legumes³. This political context explains why the
32 extent of legume production area has been limited in Europe, despite the increasing demand.
33 Only 1.7% of European cropland area was used for soybean production in 2016¹. However, it is
34 well-documented that legume (including soybean) production and consumption have numerous
35 benefits. First, it increases yield of the subsequent crop and reduces occurrence of weeds and
36 pathogens⁴⁻⁶ (agronomic benefit). Second, it reduces use of nitrogen (N) fertilizer due to
37 symbiotic N fixation and associated reductions in greenhouse gases emissions and energy use^{5,7}
38 (environmental benefit). Lastly, legume consumption contributes to reducing risks associated
39 with chronic diseases, such as cardiovascular diseases, diabetes, cancer, obesity and gut health⁸
40 (human health benefit, see^{9,10} for soybean). On the other hand, increases in legume-producing
41 areas may lead to side effects. For example, in comparison to cereals, soybean production
42 contributes less to soil carbon sequestration^{11,12}. It may also lead to an increased reliance on

* « Europe » refers to the FAO category which includes Russia

43 pesticides (e.g. pea¹³) or irrigation (e.g. soybean¹⁴). Despite these side effects, the overall
44 benefits of increasing the share of legume crops in European agricultural systems are still
45 expected to be positive⁸.

46 Among commonly cultivated grain legumes, soybean stands out as the crop species
47 experiencing the fastest expansion rate in Europe with an increase of more than four-fold from
48 1.2 Mha in 2004 to 5 Mha in 2016 (Figure S1) in response to a rising demand for locally-
49 produced, non-GM soybean^{15,16}. This expansion is expected to continue in the next decade but
50 at a slower pace¹⁷. In this context, a Europe-wide assessment on the agro-climatic suitability of
51 soybean production areas under current and future climate is of strategic importance. Building
52 on two recently published global datasets including historical soybean yield^{18,19} and
53 retrospective meteorological forcing²⁰, we developed data-driven relationships between climate
54 and soybean yield to estimate soybean suitable areas over Europe. Several machine learning
55 algorithms were trained and tested at the global scale (Random Forest, Artificial Neural
56 Networks, Generalized Additive Model, and Multiple Linear Regression) to predict soybean yield
57 as a function of monthly climate inputs (solar radiation, minimum and maximum temperature,
58 rainfall, and vapour pressure) calculated over the growing season (April to October). A large
59 share of the training data was taken from major soybean-producing countries (Argentina, Brazil,
60 Canada, China, India, Italy and the United States), and zero-yield data points were randomly
61 sampled in climate zones known to be unsuitable for soybean production (e.g. deserts and
62 arctic areas) and added to the dataset so that they represented about 20% of the final dataset.
63 The most accurate algorithm was selected after running a cross-validation procedure assessing
64 model transferability in time and space²¹. The selected algorithm (Random Forest) was then run

65 for the entire Europe to assess potential distribution of soybean suitable area in rainfed
66 conditions under current and future climate. Projections of soybean suitability in Europe were
67 performed using 16 climate change scenarios consisting of bias-corrected data produced by
68 eight Global Climate Models of the Coupled Model Intercomparison phase 5 (CMIP5)²² and two
69 Representative Concentration Pathways (RCPs; 4.5 and 8.5 W m⁻²)²³ in the 2050s and the 2090s.
70 The projections assume a growing season from April to October and no irrigation, although
71 soybean is often irrigated in Europe¹⁴. The no irrigation assumption prevents us from making
72 any hypothesis about available water for irrigation, which is a complex issue especially under
73 climate change²⁴. We therefore acknowledge that our yield projections are probably a bit
74 conservative from that point of view. Day length, soil type and atmospheric CO₂ concentration
75 are other factors not accounted for in our model, justification of these choices and implications
76 for the results are discussed at the end of the paper.

77 *Results*

78 **Model fitting and selection**

79 Among algorithms tested in this study, Random Forest appears to be the most accurate in terms
80 of root-mean-squared error of prediction (RMSEP) and Nash–Sutcliffe model efficiency
81 coefficient (MEF) (Table 1, Figure S2). It achieves the lowest prediction error (RMSEP = 0.35 t ha⁻¹)
82 and the highest efficiency (MEF = 0.93), as estimated with an unstratified cross validation
83 procedure. It also displays the best transferability in time (RMSEP = 0.45 t ha⁻¹ when applied to
84 years different from those used for training) and space (RMSEP = 0.43 t ha⁻¹ when applied in
85 locations distant by 500 km – or 5 grid-cells – from those used for training). Our results reveal
86 that transferability in space decreases with increasing distance between training and test

87 datasets for all models, with a threshold of 1000 km above which the performance of the
88 selected algorithm deteriorates markedly (Table 1). However, despite the limited number of
89 grid-cells located in Europe in the training dataset, most European cropping areas are within
90 1000 km of these grid-cells (Figure S3). This provides a good level of confidence in yield
91 projections from the selected algorithm, except in the north-east (Ukraine, Belarus, Russia,
92 Latvia, Estonia, Lithuania), west (Ireland), the south-west and south-east (south-western Spain,
93 Portugal, Turkey), where projections should be interpreted with more caution.

94 Projections of soybean yield in Europe

95 The projections of the Random Forest algorithm – which assume no irrigation and a fixed
96 growing period from April to October – suggest high suitability for soybean under historical and
97 future climate (Figure 1). Under historical climate (Figure 1A), about 106 Mha show projected
98 yield equal or higher than 2 t ha⁻¹ (Figure 2B), while in 2016 the soybean production area in
99 Europe was only 5 Mha with 2 t ha⁻¹ of average yield¹. Therefore, soybean suitable area appears
100 to be much larger than current harvested area in Europe, which suggests that soybean
101 production is not limited by climate conditions. Our projections indicate an overall positive
102 effect of climate change on soybean yield, with a projected increase of median soybean yield
103 from 1.2 t ha⁻¹ under historical climate to 1.6 t ha⁻¹ (2050s – RCP 4.5) and 1.8 t ha⁻¹ (2090s – RCP
104 8.5), even without effects of elevated CO₂ concentration (Figure 2A). Importantly, the increase
105 in the extent of low-yielding suitable areas (+27% to +42% for areas with projected yield
106 ≥ 1.5 t ha⁻¹ relative to historical climate) was associated with a decrease in the extent of high-
107 yielding suitable areas (-65% to -100% for areas with projected yield ≥ 2.5 t ha⁻¹ relative to
108 historical climate) (Figure 2B, Table S2). The decrease in medium-yielding suitable areas was

109 substantial under RCP 8.5 (-21% to -58% relative to historical climate) compared to RCP 4.5 (-4%
110 to -10%). These changes reflected losses in the South (e.g. Spain, Italy) and gains in the North
111 and the East (e.g. Russia, Ukraine, Poland, and Belarus). The northward and eastward shifts of
112 higher-yielding suitable area and the decrease of suitable area in the South of Europe would
113 become noticeable by the middle of this century (Figure 1B,D) and further intensify by the end
114 of this century, in particular under RCP 8.5 (Figure 1C,E). Although the projections made by the
115 end of the century are probably more uncertain, we highlight that these projections do not
116 involve any extrapolation of the model beyond the range of training data (historical growing
117 season climate). Indeed, only 0.03% of the data samples in the future climate scenarios fall out
118 of the range of training climate data (Figure S4, Table S3). Moreover, current available evidence
119 from farmer's fields and on-station field experiments (Figure S5-A) as well as from available
120 estimates of current soybean harvested area in Europe (Figure S5-B) confirms that soybean can
121 be grown at high latitude in Europe of 55°N to 57.5°N (corresponding to the northern part of
122 Latvia).

123 Comparison with process-based crop models

124 Previous studies relied on dynamic process-based crop models to estimate the impact of current
125 and future climate on soybean yield at the global scale^{25,26}, although none of them focused in
126 Europe specifically. We compared our projections of soybean yield (obtained with the RF
127 algorithm) to soybean yield values simulated by the Agricultural Model Intercomparison and
128 Improvement Project (AgMIP), which is based on state-of-the-art global process-based crop
129 models (see Methods for details). Although a detailed comparison of projected yield levels
130 between the two approaches is hampered by different underlying assumptions (i.e. RF

131 simulates actual yields while AgMIP simulates water- and nitrogen-limited yields), the results of
132 our comparative analysis indicate a good level of agreement between the conclusions obtained
133 from RF and AgMIP. First, outputs from AgMIP confirm that a large share of the European area
134 is expected to be suitable under historical climate conditions (Figure S6). Second, RF yield
135 projections under historical climate fall within the range of yields simulated by the individual
136 AgMIP models in most part of Europe (Figure S7 and S8), except northern Germany, Belgium,
137 Czech Republic, Slovakia and Belarus where RF predictions are lower; and south-western France,
138 northern Spain, southern Danube region, southern Ukraine and Russia and northern Turkey
139 where RF predictions are higher. Regions with higher RF predictions than AgMIP correspond
140 mainly to current major soybean producing regions in Europe, where the highest yield levels in
141 Europe are reported (Figure S9). In these high-yielding areas, the data-driven approach (RF)
142 seems to better simulate the high yield values currently observed there than the process-based
143 models (AgMIP). Regions with lower RF predictions than AgMIP are located in the north of
144 Europe where soybean is not currently grown by farmers (Figure S5). As no or very few
145 observed yield data is available in these areas, further research is needed to determine which of
146 the AgMIP or the RF predictions are the most accurate. However, this comparison suggests that
147 the use of RF leads to conservative results in these areas that could possibly underestimate
148 soybean production but are unlikely to overestimate it compared to AgMIP. Third, similarly to
149 RF, AgMIP simulations suggest that a shift of suitable areas toward the north-east is most likely
150 with climate change (Figures S10 to S13). According to AgMIP, this shift would become
151 noticeable by mid-century with little differences between RCPs, and intensify by the end of the

152 century especially under RCP 8.5, which is consistent with RF projections (Figure 1). All these
153 conclusions hold with and without effect of CO₂ fertilization included in the simulations.

154 Climate drivers of projected yield changes

155 To identify climate drivers of projected shifts in soybean suitability, we performed a linear
156 discriminant analysis (LDA) to find out combinations of climate variables that best discriminate
157 three types of yield response – yield decrease (by at least -0.3 t ha⁻¹), yield increase (by at least
158 +0.3 t ha⁻¹), and a marginal change (projected yield change between -0.3 and +0.3 t ha⁻¹) – when
159 comparing RCP 4.5 in the 2050s to historical climate (Figure S14). The +/- 0.3 t ha⁻¹ threshold
160 was chosen to be higher than the observed interannual variability of soybean yield in Europe,
161 which is 0.2 t ha⁻¹ (standard deviation) over the 2000-2014 time period¹. We also performed the
162 same analysis with RCP 8.5 in the 2050s, but we present results for RCP 4.5 because conclusions
163 are similar. The LDA showed an overall accuracy of 89%, and was able to discriminate between
164 grid-cells experiencing yield increase or yield decrease (Figure 3A, Table S5). This analysis
165 reveals the key role of temperature (both minimum and maximum) in driving projected yield
166 changes. Indeed, climate variables showing the highest contributions to the first two linear
167 discriminants are mostly temperature variables (Figure 3B-C, Figure S15). Our results suggest
168 that projected yield decrease in the South of Europe is mainly associated with detrimental
169 warming effects during the reproductive period in warmer regions. Maximum temperatures in
170 months 4 and 5 of the growing season (i.e. July-August) reach 31.3°C and 30.9°C, respectively,
171 under RCP 4.5 in the 2050s (Table 2), which exceeds the optimum of 28.5-30°C for pollen
172 germination (see Table S7 and references therein). These results are consistent with findings
173 reported by ^{27,28} for soybean in the US, who reported significant negative effects of

174 temperatures higher than 30°C on yield. Conversely, projected yield increase in the North and
175 East of Europe are mainly associated with positive warming effects in colder regions, where an
176 increase of temperature is expected to have a positive effect on soybean yield because
177 temperatures get closer to the optimum for a number of physiological processes in soybean
178 (Table 2, Table S7). A detailed analysis of the Random Forest algorithm using partial dependence
179 plots relating temperature variables to simulated soybean yield (Figure S16) confirms that
180 model outputs are very consistent with the current knowledge on soybean physiology (Table
181 S7). Indeed, several temperature thresholds established from field experiments are captured in
182 the partial dependence plots: the minimum temperature of 4°C for germination (Figure S16-A),
183 the minimum temperature of 10°C and optimum temperature of 30°C for pollen germination
184 (Figure S16-B-D-E), the maximum temperature of 40°C for crop development pre- and post-
185 anthesis (Figure S16-B-C). Together these results suggest that the soybean yield projections
186 presented in this paper are in line with the current understanding of soybean physiology.

187 Area requirements for soybean self-sufficiency in Europe

188 Our projections of soybean suitable area suggest untapped opportunities to increase soybean
189 production in Europe. We estimated the soybean production area required to reach a self-
190 sufficiency level of 50% and 100% based on yield projections presented in Figure 1. A three-step
191 procedure was followed. First, we assumed that soybean could only be grown on current
192 cropland²⁹. Under this assumption, soybean cannot be grown in place of permanent pastures, in
193 line with the Common Agricultural Policy of the European Union aiming at their protection³⁰.
194 Second, we considered four scenarios for the increase of soybean frequency in crop sequences.
195 In these scenarios, soybean is grown one year in three, four, five, or six years, which correspond

196 to 33%, 25%, 20%, and 16% cropland area in a grid-cell under soybean, respectively. These
197 scenarios are consistent with observed and recommended soybean frequencies in crop
198 sequences in Europe. Indeed, a 1-in-3 year or 1-in-4 year soybean cultivation is often
199 recommended to limit the risk of disease occurrence³¹ (especially those caused by two fungal
200 pathogens *Sclerotinia sclerotiorum* and *Rhizoctonia solani*^{32,33}), although higher frequencies are
201 observed in Europe^{31,34} and other countries^{35,36}. Third we assumed that soybean is grown
202 preferably in high-yielding grid-cells. This assumption could be tested more thoroughly by
203 comparing the profitability of soybean to other crop species currently grown in these areas.
204 However, these economic considerations are beyond the scope of this study as they depend on
205 market dynamics and the evolution of public subsidies which are subject to considerable
206 uncertainty. Therefore, soybean areas were allocated to grid-cells ranked in decreasing order of
207 projected yield values until the cumulated production (calculated as the product of area and
208 yield) reached 50% and 100% of current annual soybean consumption of Europe (58 Mt in
209 average over 2009-2013¹). Results suggest that a self-sufficiency level of 50% (100%) would be
210 achievable in Europe under historical and future climate whatever the frequency of soybean in
211 crop sequences, if 4 to 5% (9 to 11%) of the current European cropland is dedicated to soybean
212 production (Figure 4, Figure S17). For the self-sufficiency level of 50%, this share corresponds to
213 11 to 14.5 Mha or about 2 to 3 times larger than the current European soybean area (5 Mha in
214 2016¹). For the self-sufficiency level of 100%, the corresponding values are 24.5 to 32.4 Mha or
215 about 5 to 6 times larger than the current area (Figure 5A). Harvested area maps indicate that
216 four crops currently dominate the area needed to achieve 50% and 100% soybean self-

217 sufficiency (Figure 4): wheat is the main crop with 36 to 43% of the area, followed by maize (14-
218 31%), barley (10-21%), and sunflower (7-15%) (Table S8).

219 Potential nitrogen fertilizer savings from soybean expansion

220 An expansion of soybean area may have some environmental benefits, in particular by reducing
221 nitrogen (N) fertilizer applications. Soybean is an N₂-fixing crop which is usually fertilized at very
222 low rates or even not fertilized at all with nitrogen, saving N-fertilizer use compared to other
223 crops. Assuming soybean is not fertilized with nitrogen in Europe, and using published global
224 maps of crop-specific N-fertilizer rate^{37,38} (Figures S18 to S20), we estimated N-fertilizer savings
225 from the replacement of fertilized crops (e.g. wheat) by unfertilized soybean. Results show that
226 the extra soybean area needed to reach 50% (100%) self-sufficiency would reduce total N-
227 fertilizer use in Europe by 4 to 7% (13 to 17%) (Figure 5C). This estimate is likely to be
228 conservative because additional N-fertilizer savings could be expected from reducing N-fertilizer
229 rates applied to non-legume crops following soybean in the crop sequence by about 20 kgN ha⁻¹,
230 as commonly recommended by agronomists for cereals in relation with the high N content of
231 legume residuals^{4,5}. However, an accurate estimation of this positive effect of soybean as a
232 previous crop for cereals would require assumptions about which crops would preferentially be
233 replaced by soybean, which is out of scope of this study but deserves further research.

234

235 *Discussion*

236 Large opportunities to increase soybean production in Europe

237 Our study shows that soybean suitable area estimated from agro-climatic conditions is much
238 larger than current soybean harvested area in Europe, even under projected climate change. It
239 also suggests that current and future climate would allow Europe to grow enough soybean to
240 reach a self-sufficiency level of 50%, i.e. five time greater than the current level of 10%.
241 Moreover, our projections suggest that achieving 100% self-sufficiency is possible, at least from
242 an agro-climatic point of view. Although the 100% self-sufficiency scenario might not be a
243 realistic target for a number of reasons discussed below, this nevertheless highlights the large
244 opportunities to increase soybean production in Europe. These results have concrete
245 implications for Europe. First, soybean production doesn't appear to be limited by climate only.
246 Socio-economic factors are currently limiting the development of soybean as well, like low
247 market competitiveness compared to other crops or imported soybean, lack of value chains
248 development, public subsidies in favor of cereals, difficulty to account for non-market
249 environmental benefits of soybean, low information dissemination on best management
250 practices, as previously mentioned by other studies ^{2,3}. Second, our results show that a shift of
251 soybean suitable areas from the south of Europe towards the north-east of Europe is projected
252 under climate change by the middle of this century, according to the moderate and intense
253 climate change scenarios considered (RCPs 4.5 and 8.5). Therefore, the relative profitability of
254 soybean production in the different countries within Europe might change accordingly. By
255 highlighting regions with high projected soybean yield under both current and future climate,
256 our findings could help policymakers and agribusiness to reorganize the production area

257 distribution. This could also be of interest for breeders, who recently started making efforts to
258 create soybean varieties specifically adapted to European conditions, especially high latitudes³⁹.
259 In this study, moderate and intense levels of warming were considered. These two emission
260 scenarios were selected to consider both strong (RCP8.5) and moderate (RCP 4.5) impacts of
261 climate change.

262 On the environmental impacts of soybean expansion in Europe

263 Based on a simple assumption that non-fertilized soybean would replace N-fertilized crops, we
264 estimate that the expansion of the soybean area needed to reach 50% and 100% self-sufficiency
265 and the resulting decrease of crop areas receiving high amounts of N-fertilizer would reduce
266 total N-fertilizer use in Europe by 4 to 7% and 13 to 17%, respectively. The excess use of N-
267 fertilizers is widely recognized to be associated with negative impacts on soil, air, and water
268 quality, climate change (through greenhouse gases emissions from manufacture and field
269 application), and biodiversity conservation⁴⁰. Therefore, our results suggest that the reduction
270 in N-fertilizer use resulting from a soybean expansion in Europe could have positive
271 environmental effects. However, further research is necessary to better quantify the possible
272 benefits and side effects. Increasing soybean area might also help control pests, diseases, and
273 weeds in European agricultural systems through a diversification of cereal-based intensive
274 cropping systems⁶. For example, it is now well established that diversifying crop sequences help
275 control weeds, especially when the diversity of sowing periods (e.g. autumn, spring) in the crop
276 sequence increases⁴¹. An expansion of the soybean area also questions agricultural water
277 management. Indeed, an increase in soybean acreage will have an impact on water demand,
278 and this impact can be negative or positive depending on which crop is replaced by soybean. As

279 soybean is a summer crop, its cultivation in replacement for a winter crop (e.g. wheat) is
280 expected to increase water demand during summer and negatively impact water resources,
281 especially in Southern Europe^{14,24}. On the other hand, if soybean replaces another irrigated
282 summer crop such as maize, the demand for summer water should decrease because the
283 amount of water applied is generally higher on irrigated corn than on irrigated soybeans⁴².

284 Soybean expansion and land use change in Europe and abroad

285 Projected shifts in soybean suitable area and possible expansion of soybean area in Europe
286 could have important implications in terms of land use, both in Europe and in major soybean
287 producing countries. For example, if an additional 9 Mha of soybean is grown in place of wheat
288 (this corresponds to area requirements for 50% self-sufficiency), the European wheat
289 production area (60 Mha, average 2013-2017¹) would be reduced by 15% with strong
290 consequences on wheat production. Identifying which crop would be more likely replaced by
291 soybean remains an open question, but our results indicate that four crops currently dominate
292 the area needed to achieve 50% and 100% soybean self-sufficiency: wheat, maize, barley, and
293 sunflower (Table S8). An expansion of soybean area into previously uncultivated land could have
294 negative impacts for the environment through associated GHGs emissions and potential loss of
295 biodiversity. A large increase in European soybean production would also have potential
296 impacts in other countries. As previously mentioned, Europe currently imports about 90% of its
297 domestic soybean consumption. A large share of these imports come from South America,
298 particularly Argentina, where a link has been established between the deforestation of the Gran
299 Chaco dry tropical forest – a biodiversity hotspot⁴³ – and global demand for soybeans⁴⁴. Given
300 current soybean yield level in Argentina (2.9 t ha⁻¹, average 2013-2017¹), the additional soybean

301 production needed for Europe to achieve 50% self-sufficiency would represent about 8 Mha of
302 soybean area in Argentina, or more than 40% of the national soybean area in Argentina. Overall,
303 these results suggest that an expansion of the European soybean area could help prevent
304 deforestation in biodiversity hotspots. However, we acknowledge that land use dynamics are
305 difficult to anticipate because of their complexity⁴⁵, and international trade needs be considered
306 through economic modeling.

307 On the use of machine learning to model soybean yield

308 In recent years, machine learning techniques have been successfully applied to predict yields of
309 a variety of crops in different world regions^{46–49}. Consistently with those previous studies, the
310 results presented here demonstrate the good predictive ability of the Random Forest algorithm
311 when applied to soybean, with a RMSEP of 0.35 t ha⁻¹. Two important results reinforce the
312 reliability of the conclusions drawn from our model. First, the model behavior is very consistent
313 with current knowledge on soybean physiology, as shown by partial dependence plots relating
314 temperature variables to yield (Figure S16, Table S7). This is remarkable as no information was *a*
315 *priori* included in the model on this point. Second, our projections do not involve any
316 extrapolation of the model beyond the range of data used for training (Figure S4, Table S3).
317 Our machine learning algorithm (RF) has also several advantages compared to standard
318 parametric statistical models. RF does not make any prior assumption on the relationship
319 between soybean yield and climate inputs. RF is able to handle nonlinear effects and complex
320 interactions between climate inputs which would have been difficult to include in standard
321 statistical models. Moreover, with RF, the yield response to climate is data-driven and does not
322 rely on pre-specified equations. The good performance of our model suggests that the

323 combined use of large global climate and yield data sets with machine learning techniques is a
324 promising approach to studying the impact of climate on agricultural production. One key point
325 underlying this is probably the wide range of climate conditions captured in training data.
326 The assessment of transferability in space revealed that our model predictive ability decreased
327 markedly when the distance between training and test grid-cells was higher than 1000 km
328 (Table 1). The global soybean yield dataset we used in this study²⁰ contains only a few grid-cells
329 in Europe, which are located in the north of Italy (Figure S21). This low amount of grid-cells
330 located in Europe is consistent with the limited (although expanding) extent of soybean
331 production area in Europe, which was especially true around the year 2000 that corresponds to
332 the time period depicted by the dataset we used¹⁸. These few grid-cells located in Europe are of
333 great value as they allow capturing local features of soybean yield – climate relationships in our
334 model. However, increasing the number of observed soybean yield data in Europe appears as a
335 great opportunity to improve yield predictions from climate inputs. To this end, databases
336 containing large amounts of experimental data such as the one recently published by Cernay et
337 al.⁵⁰ could be used in the future to update our projections.

338 Potential effects of some factors not included in the model

339 Our predictive model presents some limitations due to the fact that several factors are
340 neglected, in particular day length and maturity group, soil type, atmospheric CO₂
341 concentration, and shifting growing season due to climate change. Soybean is known to be a
342 short-day plant that needs day length to stay below a given threshold to flower at a maximum
343 rate: if day length exceeds this threshold, flowering is delayed and maturity might not be
344 reached before the end the growing season⁵¹. This critical day length depends on the maturity

345 group: soybean cultivars of high maturity groups have a low critical day length while cultivars of
346 low maturity groups have a high one³⁹. Therefore, high maturity groups are usually cultivated at
347 low latitudes and low maturity groups at high latitudes. However, day length has not been
348 included in the predictors of our model because it is a function of latitude, and thus there is a
349 risk of confounding effects associated with environmental or socio-economic variables
350 correlated with latitude (e.g. GDP per capita, access to fertilizer and modern varieties⁵²).
351 Likewise, maturity group was not taken into account directly because, to our knowledge, no
352 global dataset of soybean maturity groups is available. We believe this approach is likely to have
353 a small impact on our results for three reasons. First, model residuals show no association with
354 latitude (Figure S27 A). Importantly, residuals were not larger and did not reveal any bias at high
355 latitudes where day length might be an issue. Second, although short day lengths may be an
356 issue at high latitudes, there is already some evidence from farmer's fields and field
357 experiments that soybean could be grown at high latitude (up to 55-57 °N) in Europe (Figure S5-
358 A). Third, our model doesn't project high yields at latitudes higher than 55-57 °N. Nevertheless,
359 we acknowledge that further research will benefit from exploring soybean cultivation at high
360 latitudes with process-based crop models including the effect of day length on soybean
361 physiology⁵³.
362 Soil type is known to have impacts on crop growth and yield in multiple ways. However, no
363 reliable historical soil dataset is currently available at the global scale for key soil characteristic
364 relevant to crop growth^{54,55}. We therefore decided not to include soil type in our model.
365 Nevertheless, we note that some analyses based on global soil datasets suggest no important
366 limitation to root growth in European soils⁵⁶, thus making our yield projections more likely to be

367 pessimistic than optimistic. Additionally, published maps of European soil types⁵⁷ highlight the
368 existence of specific soil types in the North of Europe (e.g. leptosols in Norway, and podzols in
369 Scandinavia), and the North-East of Europe (e.g. albeluvisols in Russia), but these areas are not
370 identified as high-yielding soybean areas in our projections.

371 In line with previous work, we do not consider the effect of an increase in atmospheric CO₂
372 concentration on soybean yields because this effect is still very uncertain due to many complex
373 interaction mechanisms, and is still widely discussed in the research community⁵⁸⁻⁶³. However,
374 if an increase in atmospheric CO₂ concentration positively affects soybean yield as suggested by
375 the most up-to-date quantitative synthesis of available experimental and modeling studies
376 (which reports an estimated global average yield increase of soybean of +11% for an increase in
377 atmospheric CO₂ concentration of +100ppm)⁶⁴, it further supports our findings that large
378 opportunities exist to improve Europe soybean self-sufficiency under climate change. And
379 although it is likely that an increase in atmospheric CO₂ concentration will impact absolute
380 yields, there is no evidence that it will change the relative yields of the different geographical
381 regions, and thus the ranking of the grid-cells considered here.

382 Our projections assume a fixed growing season from April to October. A shift in the growing
383 season (sowing date and cultivars with different maturity or length of growing cycle) might offer
384 opportunities for crop adaptation to climate change. But investigation of these effects is not
385 straightforward with the RF algorithm we used for projections, and this raises a number of
386 methodological questions that are out of the scope of this paper. For example, a possible caveat
387 of statistical models including RF is that the extent to which a shortened crop duration and
388 associated yield losses under increased temperature conditions is accounted for is unclear. The

389 fixed growing season assumed in this study may be a reason for relatively optimistic soybean
390 yield projections. However, in Northern and Eastern Europe, the positive impact of climate
391 change on yields can be interpreted as the result of a decrease in cold stress, which would
392 compensate for the negative impact of reduced crop growth duration. Here again, further
393 research will benefit from the use of process-based crop models able to capture these
394 processes, to make comparisons with the results presented in this paper.

395 *Methods*

396 Soybean yield, irrigated fraction, and climate data

397 Soybean yields used in this paper are from the global dataset of historical yields updated version
398 ^{18,19}. This includes grid-wise soybean yields worldwide with the grid size of 1.125 degree, which
399 covers the period 1981-2010. Yield values reported in this dataset result from the combination
400 of national-scale yield statistics from the FAO, global crop calendars and harvested areas, and
401 satellite-derived net primary production values. Therefore, the grid-cell yields are estimated
402 values resulting from the combination of several sources of information. In this dataset, the
403 soybean harvested areas are those of the year c.a. 2000, and are kept constant (Figure S21). The
404 geographical coverage of soybean harvested area found in¹⁸ is relatively limited compared to
405 other datasets³⁸ because in some parts of the world the crop calendar used to generate grid-cell
406 yield estimates⁶⁵ is missing. Regarding historical climate data, the global retrospective
407 meteorological forcing dataset tailored for agricultural application (GRASP) was used²⁰. This
408 dataset contains monthly average of five climatic variables relevant in explaining crop growth
409 and yield: daily maximum and minimum air temperatures at 2m, daily precipitation, daily solar
410 radiation, and daily vapor pressure. These variables are available for the period 1961–2010 at
411 the same spatial resolution as yield data, i.e. a grid size of 1.125 degree. Other meteorological
412 forcing datasets are available⁶⁶, but uncertainties associated with different datasets are small at
413 monthly time scale. The SPAM2005 v3.2 dataset (available at <http://mapspam.info/>) was used
414 to retrieve irrigated soybean fraction in each grid-cell⁶⁷. This dataset provides the irrigated and
415 rainfed harvested area for a set of crops (including soybean) at the global scale for around the
416 year 2005, at a spatial resolution of 5 arc min (~0.08 degree). These data were regridded to the

417 spatial resolution of the yield data (1.125 degree) using the *projectRaster()* function of the *raster*
418 R package with argument *method* set to “*bilinear*”.

419 Data preprocessing

420 We focus on the major soybean producers representing 91% of the global soybean harvested
421 area, namely Argentina, Brazil, Canada, China, India, Italy, and USA (Figure S21). Information
422 regarding yield data and crop calendars (sowing and harvest dates) – needed to define the
423 growing seasons – may be considered more reliable for major soybean producers than for minor
424 players. Moreover, previous studies suggest that soybean actual yield is close to the estimated
425 yield potential in at least some areas within these countries, e.g. in USA ⁶⁸, Argentina ⁶⁹, and
426 Brazil ⁷⁰. This is of importance because climate effects on soybean yield are easier to detect
427 when non-climatic factors (e.g. sub-optimal management) are not limiting yield. We remove all
428 grid-cells with soybean areas lower than 1% of grid-cell area, a threshold below which we
429 consider soybean production to be too marginal to be included in the analysis. To avoid any
430 confusion with technological progress, soybean yield data are detrended in order to remove the
431 increasing trends of soybean yield time series due to improved cultivars and technological
432 progress ⁷¹. For all grid-cells, yield time series are detrended using a cubic smoothing spline $f(t)$
433 and each yield data is then expressed relatively to the expected yield value in 2010 as $Yd(t) =$
434 $f(2010) + A(t)$, where $f(2010)$ is the smoothing spline yield estimate for the year 2010 (the most
435 recent year available in the yield dataset) and $A(t)$ is the yield anomaly $A(t) = yield(t) - f(t)$.
436 Histograms of soybean yields before and after detrending are shown in Figure S22. The soybean
437 growing season is defined country-by-country according to the crop calendars provided by the
438 Agricultural Market Information System (available at: <http://www.amis-outlook.org/amis->

439 [about/calendars/soybeancal/en/](#)). Based on this source of information, the soybean growing
440 season is considered to range from April to October in China, USA, and Italy, from November to
441 May in Argentina and Brazil, from May to November in Canada, and from June to December in
442 India.

443 Adding zero yield data

444 In order to take into account climate conditions preventing soybean cultivation and leading to
445 zero yields, the yield dataset was expanded by adding grid-cells located in climate zones known
446 to be environmentally unsuitable for crop production, like deserts and arctic areas. Six climate
447 zones from the last version of the Köppen-Geiger climate classification (available at
448 <http://koepfen-geiger.vu-wien.ac.at/present.htm> ⁷²) were selected, and 67 grid-cells were
449 selected at random in each selected climate zone, so that added zero yield values represented
450 20% of the final dataset. The six selected climate zones are described in Table S9, and a map
451 showing locations of added grid-cells is available in Figure S23. The final dataset includes 30,337
452 yield data values. This procedure allows us to significantly increase the range of environmental
453 conditions captured in our dataset, which has been shown to have a strong impact on the
454 performances of such models⁷³.

455 Modeling soybean yield

456 Detrended soybean yield data is related to 35 climate variables defined at a monthly time step
457 over the seven months of the soybean growing season, plus the fraction of irrigated area, i.e. a
458 total of 36 variables. The 35 climate variables are monthly mean daily minimum and maximum
459 temperatures (*Tmin* and *Tmax*, degree Celsius), monthly total precipitation (*rain*, mm month⁻¹),
460 monthly mean daily total solar radiation (*solar*, MJ m⁻² day⁻¹), monthly mean air vapor pressure

461 (VP, hPa). Four different approaches are used to predict yield from the 36 input variables:
462 Artificial Neural Network (ANN), Random Forests (RF), Generalized Additive Model (GAM), and
463 Multiple Linear Regression Model (MLR). All these algorithms were fitted using the *R* software
464 v3.4.0. For ANN we used the *neuralnet()* function of the *neuralnet* package⁷⁴, with one 10-
465 neurons hidden layer and default values for other parameters. RF was fitted with the *ranger()*
466 function of the *ranger* package⁷⁵, with a number of trees set to 500 and default values for other
467 parameters. MLR was fitted with the *glm()* function of *R*, and GAM was fitted with the *gam()*
468 function of the *gam* package⁷⁶.

469 Assessing model transferability in time and space

470 The model predictive ability is first assessed using a bootstrap approach with 25 out-of-bag
471 samples generated by bootstrap, using the *train()* function of the *caret* R package. However,
472 recent articles have highlighted the importance of rigorous cross-validation strategies to ensure
473 that the predictive capacity of a given algorithm is evaluated on data as independent as possible
474 from the data used to train that algorithm²¹. Here, we run two cross-validation strategies to
475 assess transferability of the above algorithms in time and space. Transferability in time was
476 assessed by splitting the dataset into two periods in order to assess the ability of each algorithm
477 to predict a period of time different from the one used for the training: 1981-1995, and 1996-
478 2010. In a first step, each algorithm was fitted on 1981-1995 to predict 1996-2010. In a second
479 step, each algorithm was fitted on 1996-2010 to predict 1981-1995. Transferability in space was
480 assessed using a five-step procedure implemented for each algorithm in turn: (i) select a grid-
481 cell at random in the yield database (excluding the zero yield cells), (ii) define 7 buffer zones of
482 different sizes (radius) around the selected grid-cell (radius values are 100 km, 500 km, 1000

483 km, 1500 km, 2000 km, 2500 km, and 3000 km), (iii) for each buffer zone, remove grid-cells
484 within the buffer zone and fit the algorithm on the rest of the dataset (including added zero
485 yield grid-cells) – note that to avoid any confusion with the size of the training dataset, the
486 training dataset was composed of 700 grid-cells selected at random outside the buffer zone, (iv)
487 predict the 30 years of yield for the grid-cell selected at step (i) considering each buffer zone in
488 turn, (v) compute average error of prediction over years for the selected grid-cell considering
489 each buffer zone in turn. This procedure is repeated over 10 grid-cells selected at random in
490 each country in order to estimate transferability in space for various degree of spatial proximity
491 between the training and test datasets. Assessing transferability in space is key here because of
492 the low number of grid-cells located in Europe in the historical soybean yield dataset (Figure
493 S21). Therefore, making projections of soybean yield in Europe will necessarily imply some
494 degree of transferability in space of the algorithm. In both cases (transferability in space and
495 time), predictive ability was measured by computing the root mean square error of prediction
496 (RMSEP, t ha⁻¹), and Nash–Sutcliffe model efficiency (MEF, unitless). An efficiency of 1
497 corresponds to a perfect match of modeled to observed data, an efficiency of 0 indicates that
498 predictions are as accurate as the mean of observed data, whereas an efficiency lower than zero
499 occurs when the observed mean is a better predictor than the tested algorithm. The algorithm
500 showing best transferability in time and space (i.e. the lowest RMSEP and highest efficiency)
501 among ANN, RF, GAM, and MLR is used for projections of soybean yield in Europe under current
502 and future climate.

503 Yield projections in Europe under current and future climate

504 We used 16 climate change scenarios consisting of bias-corrected data of eight Global
505 Circulation Models (GCM; GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR, MIROC5, MIROC-ESM,
506 MIROC-ESM-CHEM, MRI-CGCM3, and NorESM1-M, used in the Coupled Model Intercomparison
507 phase 5 (CMIP5)²² and two Representative Concentration Pathways (RCPs; 4.5 and 8.5 W m⁻²)²³.
508 Details on the bias-correction method used here is available in ⁷⁷. Although daily data are
509 available in the bias-corrected GCM outputs, we computed and used monthly data in our
510 analysis. We consider three time periods for projections: 1981-2010 (historical), 2050-2059, and
511 2090-2099. We present the median predicted soybean yield over the eight GCMs. Soybean
512 growing season used for prediction is April to October. All projections assumed irrigated fraction
513 equals to zero. Projections are shown only on agricultural area (cropland plus pasture), in the
514 year 2000²⁹ (Figure S24).

515 Comparison with process-based crop models

516 The soybean yield projections performed with the RF model were compared to the outputs of
517 the Agricultural Model Intercomparison and Improvement Project (AgMIP). AgMIP is a major
518 international effort of coordinated agricultural modeling aiming at the improvement of crop
519 models for assessing impacts of climate change and variability on agriculture⁷⁸. Crop models
520 used in AgMIP are process-based while our RF model is data-driven and make use of a machine
521 learning algorithm with no explicit representation of crop physiological processes. This major
522 difference between the two approaches makes the comparison of their outputs very relevant to
523 assessing the robustness of our results. To perform this comparison, we used two datasets from
524 AgMIP: (i) the Global Gridded Crop Model Intercomparison phase 1 dataset²⁶ for historical

525 yields, hereafter referred to as the “GGCMI phase 1 dataset”, and (ii) the AgMIP global fast-track
526 climate impact assessment dataset⁷⁹ for future yields under different climate change scenarios,
527 hereafter referred to as the “AgMIP fast-track dataset”. The AgMIP fast-track dataset was
528 obtained using the on-line tool described by Villoria et al.⁸⁰. Both datasets provide annual
529 simulated yields for a number of crops including soybean, simulated with different Global
530 Gridded Crop Models. In the GGCMI phase 1 dataset (historical period), 14 Global Gridded Crop
531 Models, up to 11 weather datasets, 3 management harmonization levels, and purely rainfed and
532 fully irrigated conditions were used in the simulations. In the AgMIP fast-track dataset (future
533 scenarios), 7 Global Gridded Crop Models, 4 RCPs, 5 Global Circulation Models, and two options
534 for the effect of CO₂ fertilization (with and without) were used in the simulations. The
535 comparison with the outputs of our RF model was performed using a subset of those
536 simulations. Selected simulations were chosen to maximize consistency with the input variables
537 used in our RF model, and to rely as much as possible on the same crop models over the
538 historical period and future scenarios considered. The resulting subset of selected simulations is
539 presented in Table S10 for the GGCMI phase 1 dataset, and in Table S11 for the AgMIP fast-track
540 dataset. Then, the median simulated yield over years, crop models, and (when relevant) GCMs
541 was computed and mapped, in the same way as described above for the RF model outputs.

542 Analysis of climate drivers of projected yield changes

543 A Linear Discriminant Analysis (LDA) was performed to identify combinations of climate
544 variables that best discriminate between three groups of grid-cells. These groups of grid-cells
545 are defined as Group 1: yield decrease (projected yield change < - 0.3 t ha⁻¹), Group 2: yield
546 increase (projected yield change > +0.3 t ha⁻¹), and Group 3: marginal change (yield change

547 between -0.3 and $+0.3 \text{ t ha}^{-1}$) (see Figure S14 for a map of the geographical repartition of these
548 three groups in Europe). The 0.3 t ha^{-1} threshold was chosen to be higher than the observed
549 interannual variability of soybean yield in Europe, which is 0.2 t ha^{-1} (standard deviation) over
550 the 2000-2014 time period¹. The LDA was performed with the function *lda()* of the *MASS R*
551 package, with default settings.

552 Area requirements for soybean self-sufficiency in Europe

553 In average over 2009-2013, Europe domestic supply of soybean was composed of 32 Mt of
554 soybean cake, and 18 Mt of soybean grains (Table S1). Assuming a conversion factor of 0.8
555 between soybean grains and soybean cake, this is equivalent to a total domestic supply of 58 Mt
556 of soybean grain. This value was kept constant in future scenarios because (i) the European
557 soybean domestic supply is relatively stable since the 2000s (Figure S25 A), and (ii) there is little
558 effect of GDP per capita on soybean domestic supply when GDP per capita exceeds 20 000 US\$
559 per year (this value that has been exceeded since about 2005) (Figure S25 B). We estimate the
560 soybean production area required to reach a self-sufficiency level of 50% and 100% based on
561 yield projections presented in Figure 1. A three-step procedure was followed. First, we assumed
562 that soybean could only be grown on current cropland²⁹ (Figure S24 A). Under this assumption,
563 soybean cannot be grown in place of permanent pastures, in line with the Common Agricultural
564 Policy of the European Union aiming at their protection³⁰. Second, we considered four scenarios
565 for the increase of soybean frequency in crop sequences. In these scenarios, soybean is grown
566 in one year in three, four, five, or six years, which correspond to 33%, 25%, 20%, and 16%
567 cropland area in a grid-cell under soybean, respectively. Third we assumed that soybean is
568 grown preferably in high-yielding grid-cells. Based on this assumption, soybean areas were

569 allocated to grid-cells ranked in decreasing order of projected yield values until the cumulated
570 production (calculated as the product of area and yield) reached 50% and 100% of current
571 annual soybean consumption of Europe.

572 Potential N-fertilizer savings from soybean expansion

573 Soybean is an N₂-fixing crop which is usually not fertilized with nitrogen. Recent data collected
574 in a farm survey in France (588 farmers) indicate that only 5% of soybean fields received
575 mineral-N fertilization (average rate 45 kgN ha⁻¹) in 2016 (www.terresinovia.fr, unpublished
576 data). Thus N-fertilizer savings are generated when soybean replaces an N-fertilized crop like
577 wheat. To estimate potential N-fertilizer savings if soybean production area expanded enough
578 to reach self-sufficiency, we used published global maps of crop-specific N-fertilizer rate and
579 harvested area for wheat, barley, maize, potato, rapeseed, sugarbeet, and sunflower (Figure
580 S18 and S19)^{37,38}. These maps report data for around the year 2000, but represent the most
581 detailed spatially-explicit crop-specific dataset on N-fertilization to date. Crop specific N-
582 fertilizer rate were area-weighted to generate a unique map representing average N-fertilizer
583 rate of major arable crops in Europe (Figure S20). Then, this map was combined with maps of
584 soybean production area required to reach self-sufficiency under the different climate scenarios
585 (Figure 4, and Figure S17) to calculate potential N-savings of soybean area expansion. Finally,
586 calculated potential N-fertilizer savings were compared to total agricultural N use in Europe that
587 is 13.7 Mt N (average in 2009-2013¹).

588 Data availability

589 The soybean yield projections generated during this study have been deposited in the Zenodo
590 repository (DOI: [10.5281/zenodo.6136215](https://doi.org/10.5281/zenodo.6136215)) (ref⁸¹)

591 Code availability

592 R code to reproduce key results of this paper is available at:

593 https://github.com/nguilpart/soybean_yield_projections_europe

594 **Acknowledgements**

595 This work was supported by the CLAND convergence institute (16-CONV-0003) funded by the
596 French National Research Agency (ANR), by the ACCAF INRA meta-program (COMPROMISE
597 project, COMPROMISE_MP-P10177), and by the LegValue project funded by the European
598 Union's Horizon 2020 research and innovation programme under grant agreement N°727672.
599 T.I. was partly supported by the Environment Research and Technology Development Fund (S-
600 14) of the Environmental Restoration and Conservation Agency of Japan and Grant-in-Aid for
601 Scientific Research (16KT0036, 17K07984 and 18H02317) of JSPS.

602 **Author contributions statement**

603 N.G and D.M. designed research and performed the analysis. T.I. supplied yield and climate
604 data. N.G. wrote the manuscript, with substantial contributions from all co-authors. D.M.
605 initiated research.

606 **Competing interests**

607 The authors declare no competing interests.

608

609

610 **Tables**

611 **Table 1. Performances assessment of the tested machine learning algorithms.** Soybean yield was expressed as a function 5 climate variables (daily
 612 minimum and maximum temperatures, rainfall, solar radiation, and vapour pressure) calculated monthly for a 7-month growing season, plus the
 613 irrigated fraction in a grid-cell, making a total of 36 predictors. The whole dataset contains 30,337 yield observations from 1981 to 2010. The
 614 model predictive ability is first assessed using 25 out-of-bag samples generated by bootstrap. Then, transferability in time is assessed by fitting the
 615 algorithms on a period of time different than the predicted period, and transferability in space is assessed by ensuring a minimum spatial distance
 616 between training and test datasets, with seven minimal distances ranging from 100 km to 3000 km considered successively. RMSEP: root mean
 617 square error of prediction ($t\ ha^{-1}$). RMSEP values for transferability in space are the median over 70 RMSEP values (10 grid-cells * 7 countries). MEF:
 618 Nash–Sutcliffe model efficiency. An efficiency of 1 corresponds to a perfect match of modeled to observed data, an efficiency of 0 indicates that
 619 model predictions are as accurate as the mean of observed data, whereas an efficiency lower than zero occurs when the observed mean is a better
 620 predictor than the model.

	Random Forest		Artificial Neural Network		Generalized Additive model		Multiple Linear Regression	
	<i>RMSEP</i>	<i>MEF</i>	<i>RMSEP</i>	<i>MEF</i>	<i>RMSEP</i>	<i>MEF</i>	<i>RMSEP</i>	<i>MEF</i>
Random sampling with replacement (bootstrap)	0.35	0.93	0.53	0.53	0.55	0.82	0.83	0.60
Transferability in time								
1/ Training period: 1981-1995 Predicted period: 1996-2010	0.44	0.88	0.68	0.72	0.60	0.78	0.84	0.57
2/ Training period: 1996-2010 Predicted period: 1981-1995	0.46	0.87	0.62	0.76	0.61	0.77	0.83	0.59
Transferability in space								
100 km	0.25	0.57	0.43	0.05	0.42	0.28	0.55	-0.46
500 km	0.43	0.26	0.76	-1.26	0.57	-0.08	0.69	-0.87
1000 km	0.72	-0.40	0.79	-1.34	0.80	-0.65	0.82	-1.41
1500 km	0.89	-0.90	0.98	-1.94	0.93	-1.22	0.90	-1.73
2000 km	1.00	-1.33	0.91	-1.73	0.97	-1.37	0.92	-1.77
2500 km	0.98	-1.31	0.98	-1.94	0.97	-1.40	0.95	-2.08
3000 km	1.02	-1.42	0.55	-0.68	0.98	-1.61	0.98	-2.25

621

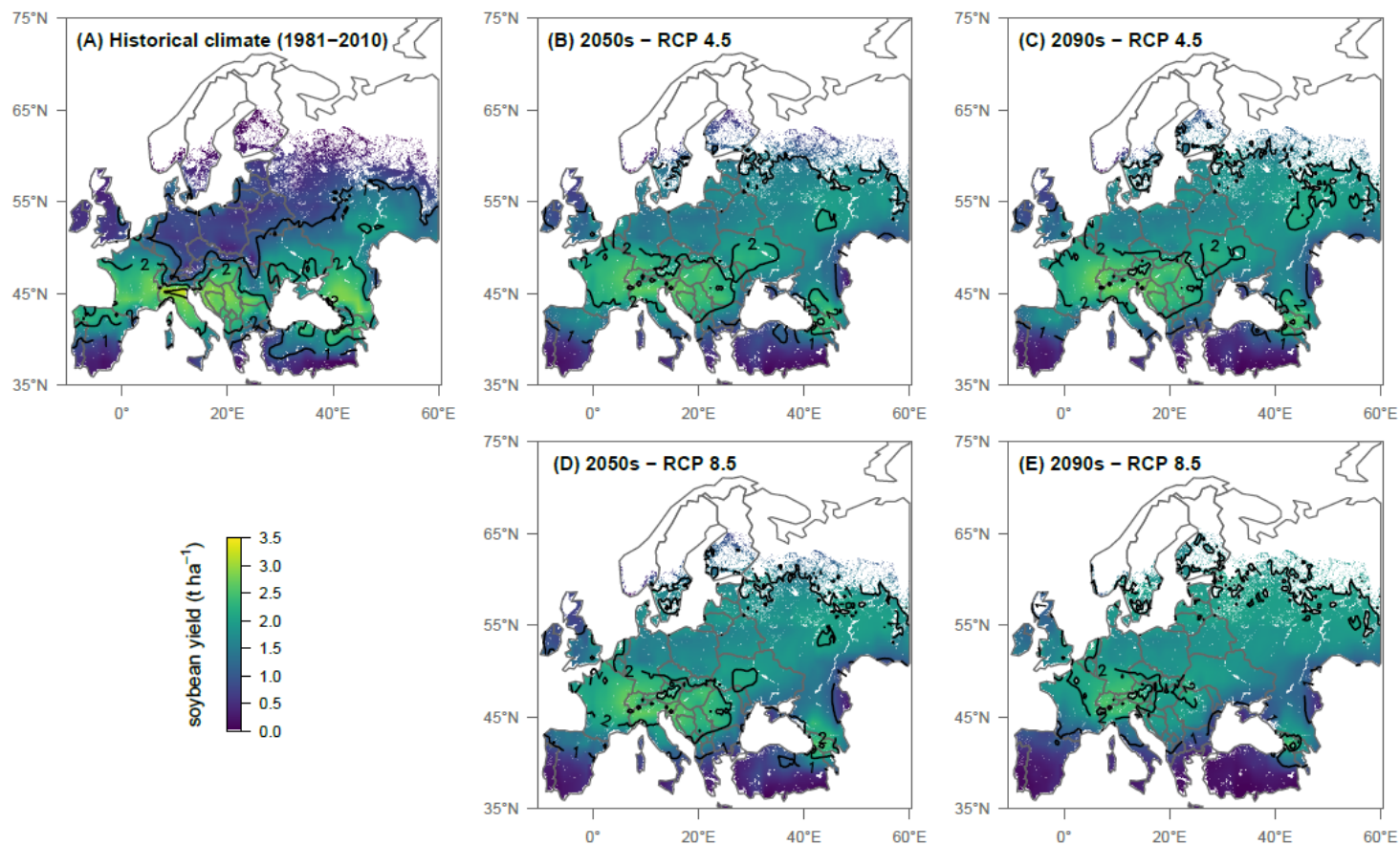
622 Table 2. Temperature changes associated with a decrease (Group 1) or a decrease (Group 2) in
 623 projected soybean yield under RCP 4.5 by mid-century relative to historical climate. Reported
 624 temperature values represent mean values for each group of grid-cells. Group 1: yield decrease
 625 (projected yield change < - 0.3 t ha⁻¹), Group 2: yield increase (projected yield change > + 0.3 t ha⁻¹).
 626 Groups are the same than those in the Linear Discriminant Analysis presented in Figure 3, but
 627 for clarity, only Group 1 and Group 2 are presented here (see Table S6 for a description of all
 628 groups and other climate variable means by group). The GRASP dataset²⁰ is used for historical
 629 climate, and the median over height Global Circulation Models²² is shown for RCP 4.5 by mid-
 630 century. Yield projections are performed with the Random Forest algorithm presented in Table 1
 631 and Figure S2.
 632

Climate variable	Month*	Group 1 yield decrease		Group 2 yield increase	
		Historical climate	2050s (RCP 4.5)	Historical climate	2050s (RCP 4.5)
Tmax (°C)	1	15,3	16,5	10,6	13,1
	2	21,1	22,6	17,6	19,6
	3	25,5	27,9	21,3	23,8
	4	28,6	31,3	23,4	25,9
	5	28,0	30,9	21,7	24,2
	6	23,4	26,1	16,5	18,9
	7	16,8	18,7	10,1	12,1
Tmin (°C)	1	4,9	5,8	1,7	3,2
	2	9,7	11,1	6,9	8,8
	3	13,3	15,8	10,5	13,2
	4	15,9	18,3	13,0	15,4
	5	15,3	18,0	11,7	14,0
	6	11,5	13,8	7,9	9,7
	7	7,0	8,3	3,3	4,9

633 * month of the soybean growing season (April to October in Europe)

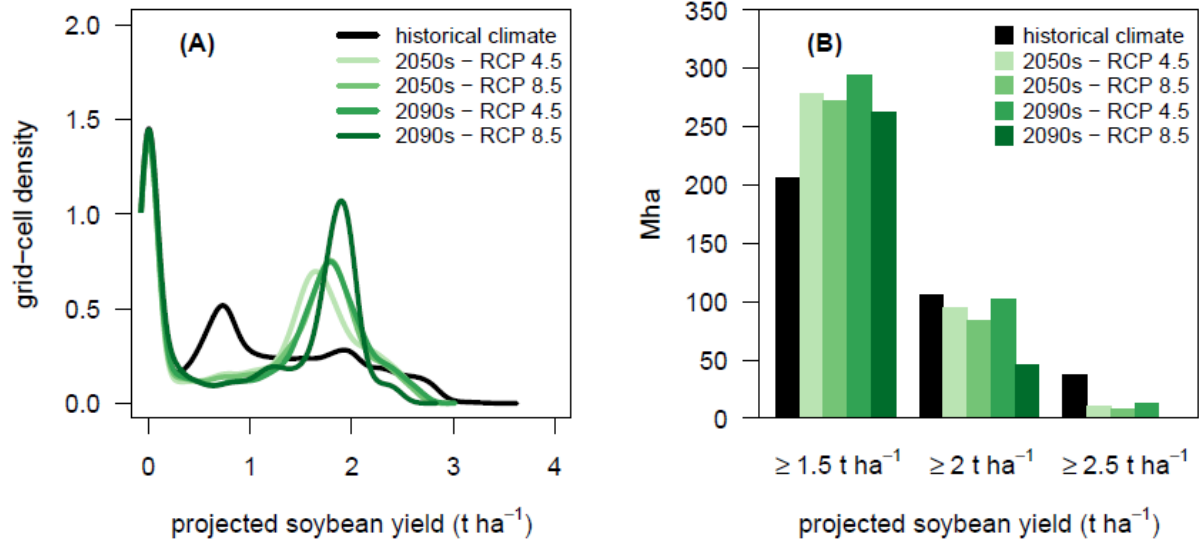
634

635 **Figures legends/captions**



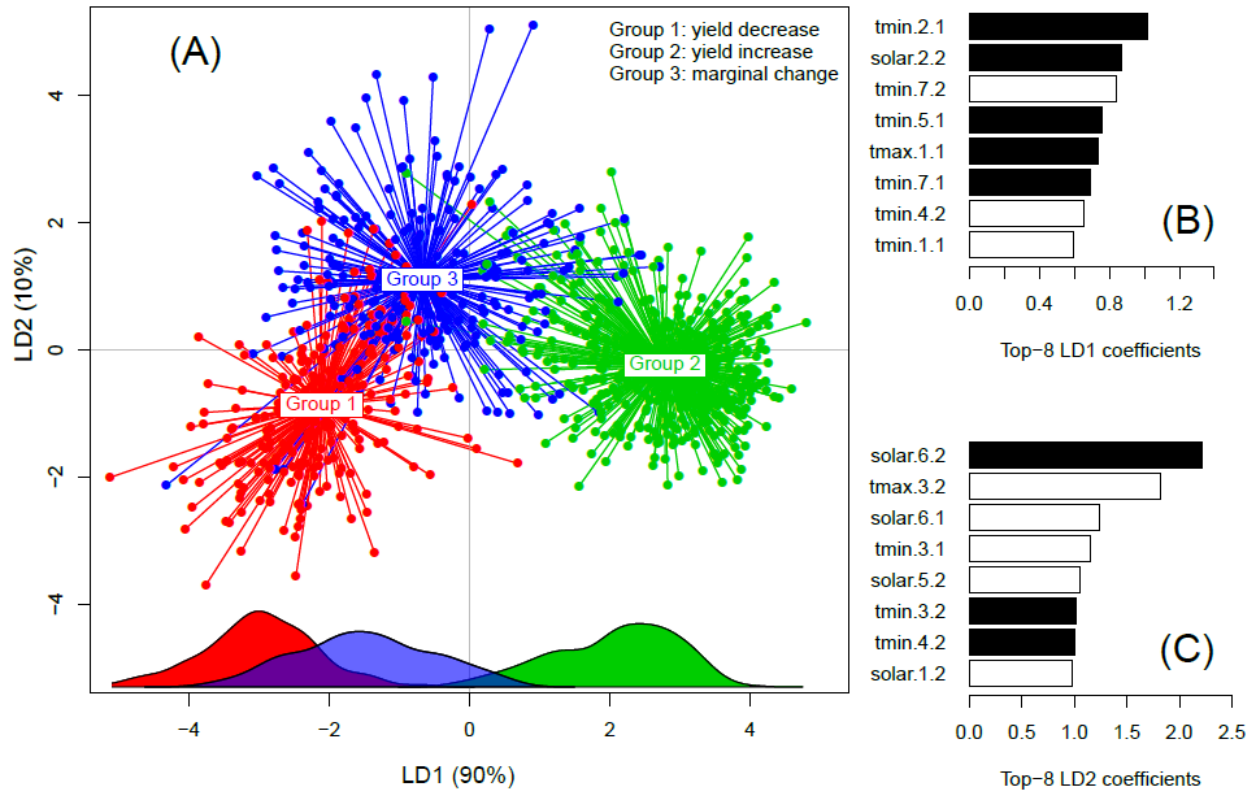
636

637 **Figure 1. Projected soybean yield in Europe under historical and future climate.** Projected soybean yield (A) under historical climate (1981-2010), (B)
638 by mid-century (2050-2059) under RCP 4.5, (C) by the end of the century (2090-2099) under RCP 4.5, (D) by mid-century (2050-2059) under RCP
639 8.5, (E) by the end of the century (2090-2099) under RCP 8.5. Maps show median projected yield using a Random Forest algorithm run with the
640 GRASP dataset²⁰ for historical climate (1981-2010), and over the eight Global Circulation Models²² considered in this study for future climate
641 scenarios. Projections are shown only on agricultural area (cropland plus pasture), in the year 2000²⁹.
642



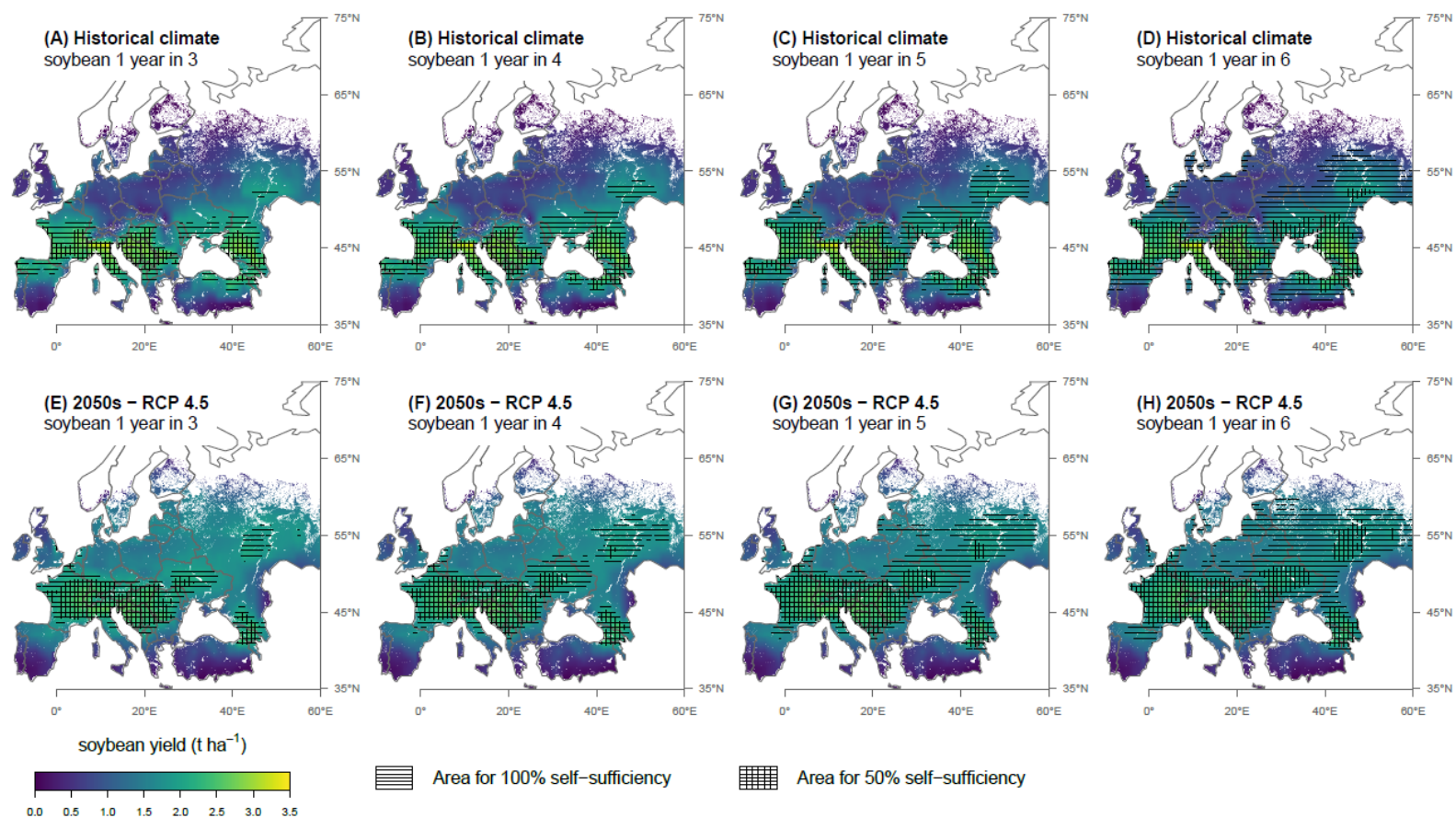
643
644

645 **Figure 2. Effect of climate change on projected soybean yield in Europe.** (A) Probability density functions of
 646 projected soybean yield under historical climate and different future climate scenarios. (B) Extent of
 647 European agricultural area for which projected soybean yield is higher or equal to a given yield threshold.
 648 Soybean yield projections were performed with a Random Forest algorithm. Projections for historical
 649 climate used the GRASP meteorological dataset²⁰ from 1981 to 2010. For future climate, eight Global
 650 Circulation Models²² were used and median projected yield over the height model was calculated.
 651



652
653

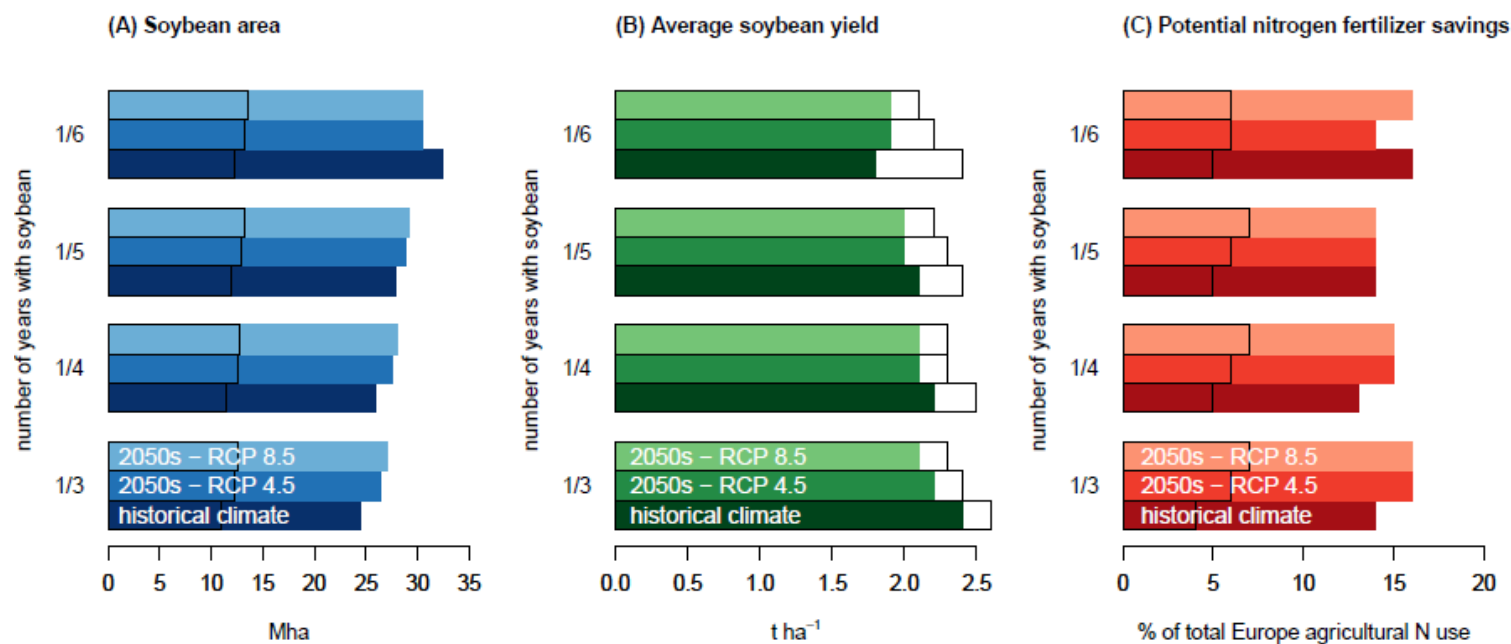
654 **Figure 3. Analysis of climate drivers of projected yield changes by 2050s under RCP 4.5 relative to historical**
 655 **climate. Panel (A)** shows the linear discriminant analysis (LDA) performed on climate variables for three
 656 groups of grid-cells defined by predicted soybean yield change between RCP 4.5 by 2050s and historical
 657 climate. Groups of grid-cells are defined as Group 1: yield decrease (projected yield change $< -0.3 \text{ t ha}^{-1}$),
 658 Group 2: yield increase (projected yield change $> 0.3 \text{ t ha}^{-1}$), and Group 3: marginal change (yield change
 659 between -0.3 and $+0.3 \text{ t ha}^{-1}$). The plot axes are the two main LDA discriminant functions. Density along
 660 the x-axis is shown for each group (same color code applies). See Figure S14 for a map showing the
 661 projected yield changes in Europe between RCP 4.5 by 2050s and historical climate. **Panels (B) and (C)**
 662 show climate variables contributions to linear discriminant 1 and 2, respectively (the higher the value, the
 663 higher the contribution of the corresponding climate input). White bars indicate a positive contribution,
 664 and black bars indicate a negative contribution. To improve clarity, only the eight climate variables
 665 contributing most to each linear discriminant are shown (see Figure S15 for the contributions of all climate
 666 variables). Suffixes to climate variables names indicate, first, the month of the soybean growing season,
 667 and second, the time period (“1” standing for historical climate, and “2” standing for the 2050s under RCP
 668 4.5). For example, “tmin.2.1” means “monthly average daily minimum temperature in the second month
 669 of the growing season under historical climate”
 670 .



671
672

673 **Figure 4. Area requirements for 50% and 100% soybean self-sufficiency in Europe under historical climate (A-D) and by 2050s under RCP 4.5 (E-H).**
 674 Based on soybean yield projections presented in Figure 1 and assuming various levels of soybean frequency in crop sequences (one year out for
 675 three, four, five and six years), soybean areas were allocated to grid-cells ranked in decreasing order of projected yield values until the cumulated
 676 production (calculated as the product of area and yield) reached 50% (light blue) and 100% (dark blue) of the current annual soybean consumption
 677 of Europe (58 Mt, average 2009-2013). We assume that soybean can only be grown on current cropland²⁹, which excludes permanent pastures in
 678 line with the Common Agricultural Policy of the European Union aiming at their protection³⁰. Background colors indicate projected soybean yield in
 679 $t\ ha^{-1}$ as in Figure 1.

680



681

682 Figure 5. (A) Soybean production area required to reach 50% and 100% soybean self-sufficiency, and associated (B) average soybean yield, and (C)
 683 potential N-fertilizer savings, under historical climate (1981–2010) and by 2050s under RCP 4.5 and 8.5. Color-filled bars indicate values
 684 corresponding to 100% soybean self-sufficiency, and empty bars with a black border indicate values corresponding to 50% soybean self-sufficiency.
 685 Four levels of soybean frequency in crop sequences are evaluated in which soybean is grown in one year in three, four, five, or six years, which
 686 correspond to 33%, 25%, 20%, and 16% cropland area in a grid-cell under soybean, respectively. Potential nitrogen fertilizer savings are calculated
 687 based on the assumption that soybean is not fertilized with nitrogen and that it will replace N-fertilized crops.
 688

689 **References**

- 690 1. Food and Agriculture Organization of the United Nations. FAOSTAT Statistics Database.
691 (2019). Available at: <http://www.fao.org/faostat/en/#data>.
- 692 2. Magrini, M. B. *et al.* Why are grain-legumes rarely present in cropping systems despite
693 their environmental and nutritional benefits? Analyzing lock-in in the French agrifood
694 system. *Ecol. Econ.* **126**, 152–162 (2016).
- 695 3. Zander, P. *et al.* Grain legume decline and potential recovery in European agriculture: a
696 review. *Agron. Sustain. Dev.* **36**, (2016).
- 697 4. Cernay, C., Makowski, D. & Pelzer, E. Preceding cultivation of grain legumes increases
698 cereal yields under low nitrogen input conditions. *Environ. Chem. Lett.* **16**, 631–636
699 (2018).
- 700 5. Nemecek, T. *et al.* Environmental impacts of introducing grain legumes into European
701 crop rotations. *Eur. J. Agron.* **28**, 380–393 (2008).
- 702 6. Gaba, S. *et al.* Multiple cropping systems as drivers for providing multiple ecosystem
703 services : from concepts to design. *Agron. Sustain. Dev.* (2014). doi:10.1007/s13593-014-
704 0272-z
- 705 7. Jensen, E. S. *et al.* Legumes for mitigation of climate change and the provision of
706 feedstock for biofuels and biorefineries. A review. *Agronomy for Sustainable Development*
707 **32**, (2012).
- 708 8. Foyer, C. H. *et al.* Neglecting legumes has compromised human health and sustainable
709 food production. *Nat. Plants* **2**, 16112 (2016).
- 710 9. Messina, M., Rogero, M. M., Fisberg, M. & Waitzberg, D. Health impact of childhood and
711 adolescent soy consumption. *Nutr. Rev.* **75**, 500–515 (2017).
- 712 10. Jayachandran, M. & Xu, B. An insight into the health benefits of fermented soy products.
713 *Food Chem.* **271**, 362–371 (2019).
- 714 11. Dold, C. *et al.* Long-term carbon uptake of agro-ecosystems in the Midwest. *Agric. For.*
715 *Meteorol.* **232**, 128–140 (2017).
- 716 12. Gilmanov, T. G. *et al.* Productivity and Carbon Dioxide Exchange of Leguminous Crops :
717 Estimates from Flux Tower Measurements. *Agron. Journal* **106**, 545–559 (2014).
- 718 13. Urruty, N., Deveaud, T., Guyomard, H. & Boiffin, J. Impacts of agricultural land use
719 changes on pesticide use in French agriculture. *Eur. J. Agron.* **80**, 113–123 (2016).
- 720 14. Rüdelsheim, P. L. J. & Smets, G. *Baseline information on agricultural practices in the EU*
721 *Soybean (Glycine max (L.) Merr.)*. (2012).
- 722 15. Martin, N. Domestic soybean to compensate the European protein deficit: illusion or real
723 market opportunity? *Oilseeds Fats Crop. Lipids* **22**, (2015).
- 724 16. Krön, M. & Bittner, U. Danube Soya – Improving European GM-free soya supply for food
725 and feed. *Oilseeds Fats Crop. Lipids* **22**, (2015).

- 726 17. OECD/FAO. *OECD-FAO Agricultural Outlook 2019-2028*. (2019).
- 727 18. Iizumi, T. *et al.* Historical changes in global yields: Major cereal and legume crops from
728 1982 to 2006. *Glob. Ecol. Biogeogr.* **23**, 346–357 (2014).
- 729 19. Iizumi, T. *et al.* Uncertainties of potentials and recent changes in global yields of major
730 crops resulting from census- and satellite-based yield datasets at multiple resolutions.
731 *PLoS One* **13**, e0203809 (2018).
- 732 20. Iizumi, T., Okada, M. & Yokozawa, M. A meteorological forcing data set for global crop
733 modeling: Development, evaluation, and intercomparison. *J. Geophys. Res. Atmos. Res.*
734 **119**, 363–384 (2014).
- 735 21. Roberts, D. R. *et al.* Cross-validation strategies for data with temporal, spatial,
736 hierarchical, or phylogenetic structure. *Ecography (Cop.)*. **40**, 913–929 (2017).
- 737 22. Taylor, K. e., Stouffer, R. J. & Meehl, G. A. An Overview of CMIP5 and experiment design.
738 *Am. Meteorol. Soc.* **93**, 485–498 (2012).
- 739 23. Van Vuuren, D. P. *et al.* The representative concentration pathways : an overview. *Clim.*
740 *Change* **109**, 5–31 (2011).
- 741 24. Koutroulis, A. G. *et al.* Freshwater vulnerability under high end climate change. A pan-
742 European assessment. *Sci. Total Environ.* **614**, 271–286 (2018).
- 743 25. Rosenzweig, C. *et al.* Assessing agricultural risks of climate change in the 21st century in a
744 global gridded crop model intercomparison. **111**, 4–9 (2014).
- 745 26. Müller, C. *et al.* The Global Gridded Crop Model Intercomparison phase 1 simulation
746 dataset. *Sci. Data* **6**, 1–22 (2019).
- 747 27. Schlenker, W. & Roberts, M. J. Nonlinear temperature effects indicate severe damages to
748 U.S. crop yields under climate change. *Proc. Natl. Acad. Sci.* **106**, 15594–15598 (2009).
- 749 28. Mourtzinis, S. *et al.* Climate-induced reduction in US-wide soybean yields underpinned by
750 region- and in-season specific responses. *Nat. Plants* **1**, 14026 (2015).
- 751 29. Ramankutty, N., Evan, A. T., Monfreda, C. & Foley, J. A. Farming the planet : 1. Geographic
752 distribution of global agricultural lands in the year 2000. *Global Biogeochem. Cycles* **22**,
753 1–19 (2008).
- 754 30. Commission, E. Sustainable land use (greening). Available at:
755 [https://ec.europa.eu/info/food-farming-fisheries/key-policies/common-agricultural-](https://ec.europa.eu/info/food-farming-fisheries/key-policies/common-agricultural-policy/income-support/greening_en)
756 [policy/income-support/greening_en](https://ec.europa.eu/info/food-farming-fisheries/key-policies/common-agricultural-policy/income-support/greening_en). (Accessed: 27th February 2020)
- 757 31. Đorđević, V., Malidža, G., Vidić, M., Milovac, Ž. & Šeremešić, S. *Best practice manual for*
758 *soya bean cultivation in the Danube region*. (Danube Soya, 2016).
- 759 32. Hartman, G. L., West, E. D. & Herman, T. K. Crops that feed the World 2. Soybean-
760 worldwide production, use, and constraints caused by pathogens and pests. *Food Secur.*
761 **3**, 5–17 (2011).
- 762 33. Pannecoque, J. *et al.* Screening for soybean varieties suited to Belgian growing
763 conditions based on maturity, yield components and resistance to Sclerotinia

- 764 sclerotiorum and Rhizoctonia solani anastomosis group 2-2IIIB. *J. Agric. Sci.* 1–8 (2018).
765 doi:10.1017/S0021859618000333
- 766 34. Réseau biologique des Chambres d’agriculture de France, A. *Grandes cultures biologiques*
767 *- Les clés de la réussite.* (2017).
- 768 35. Grassini, P., Specht, J. E., Tollenaar, M., Ciampitti, I. & Cassman, K. G. High-yield maize-
769 soybean cropping systems in the US Corn Belt. in *Crop physiology. Applications for genetic*
770 *improvement and agronomy* 15–44 (2014).
- 771 36. Salembier, C., Elverdin, J. H. & Meynard, J. Tracking on-farm innovations to unearth
772 alternatives to the dominant soybean-based system in the Argentinean Pampa. *Agron.*
773 *Sustain. Dev.* 1–10 (2016). doi:10.1007/s13593-015-0343-9
- 774 37. Mueller, N. D. *et al.* Closing yield gaps through nutrient and water management. *Nature*
775 **490**, 254–257 (2012).
- 776 38. Monfreda, C., Ramankutty, N. & Foley, J. A. Farming the planet : 2. Geographic
777 distribution of crop areas , yields , physiological types , and net primary production in the
778 year 2000. *Global Biogeochem. Cycles* **22**, 1–19 (2008).
- 779 39. Kurasch, A. K. *et al.* Identification of mega-environments in Europe and effect of allelic
780 variation at maturity E loci on adaptation of European soybean. **0000**, 1–14 (2017).
- 781 40. Houlton, B. Z. *et al.* A World of Cobenefits: Solving the Global Nitrogen Challenge. *Earth’s*
782 *Futur.* **7**, 865–872 (2019).
- 783 41. Weisberger, D., Nichols, V. & Liebman, M. Does diversifying crop rotations suppress
784 weeds? A meta-analysis. *PLoS One* **14**, 1–12 (2019).
- 785 42. Gibson, K. E. B., Gibson, J. P. & Grassini, P. Benchmarking irrigation water use in producer
786 fields in the US central Great Plains. *Environ. Res. Lett.* **14**, 054009 (2019).
- 787 43. Nori, J. *et al.* Protected areas and spatial conservation priorities for endemic vertebrates
788 of the Gran Chaco, one of the most threatened ecoregions of the world. *Divers. Distrib.*
789 **22**, 1212–1219 (2016).
- 790 44. Fehlenberg, V. *et al.* The role of soybean production as an underlying driver of
791 deforestation in. *Glob. Environ. Chang.* **45**, 24–34 (2017).
- 792 45. Meyfroidt, P. *et al.* Middle-range theories of land system change. *Glob. Environ. Chang.*
793 **53**, 52–67 (2018).
- 794 46. Delerce, S. *et al.* Assessing weather-yield relationships in rice at local scale using data
795 mining approaches. *PLoS One* **11**, (2016).
- 796 47. Everingham, Y., Sexton, J., Skocaj, D. & Inman-Bamber, G. Accurate prediction of
797 sugarcane yield using a random forest algorithm. *Agron. Sustain. Dev.* **36**, (2016).
- 798 48. Jeong, J. H. *et al.* Random Forests for Global and Regional Crop Yield Predictions. *PLoS*
799 *One* **11**, e0156571 (2016).
- 800 49. Partridge, T. F. *et al.* Mid-20th century warming hole boosts US maize yields. *Environ. Res.*
801 *Lett.* **14**, 114008 (2019).

- 802 50. Cernay, C., Pelzer, E. & Makowski, D. A global experimental dataset for assessing grain
803 legume production. *Sci. data* **3**, 160084 (2016).
- 804 51. Setiyono, T. D. *et al.* Understanding and modeling the effect of temperature and
805 daylength on soybean phenology under high-yield conditions. *F. Crop. Res.* **100**, 257–271
806 (2007).
- 807 52. Hafner, S. Trends in maize, rice, and wheat yields for 188 nations over the past 40 years:
808 A prevalence of linear growth. *Agric. Ecosyst. Environ.* **97**, 275–283 (2003).
- 809 53. Schoving, C. *et al.* Combining Simple Phenotyping and Photothermal Algorithm for the
810 Prediction of Soybean Phenology: Application to a Range of Common Cultivars Grown in
811 Europe. *Front. Plant Sci.* **10**, 1–14 (2020).
- 812 54. Folberth, C. *et al.* Uncertainty in soil data can outweigh climate impact signals in global
813 crop yield simulations. *Nat. Commun.* **7**, 1–13 (2016).
- 814 55. Guilpart, N. *et al.* Rooting for food security in Sub-Saharan Africa. *Environ. Res. Lett.* **12**,
815 114036 (2017).
- 816 56. Shangguan, W., Hengl, T., Mendes de Jesus, J., Yuan, H. & Da, Y. Mapping the global
817 depth to bedrock for land surface modeling. *J. Adv. Model. Earth Syst.* **9**, 65–88 (2017).
- 818 57. Jones, A., Montanarella, L. & Jones, R. *Soil atlas of Europe*. (2005).
- 819 58. Cober, E. R. & Morrison, M. J. Soybean Yield and Seed Composition Changes in Response
820 to Increasing Atmospheric CO₂ Concentration in Short-Season Canada. *Plants* **8**, 250
821 (2019).
- 822 59. Thomey, M. L., Slattery, R. A., Bernacchi, C. J., Köhler, I. H. & Ort, D. R. Yield response of
823 field-grown soybean exposed to heat waves under current and elevated [CO₂]. *Glob.*
824 *Chang. Biol.* **25**, 4352–4368 (2019).
- 825 60. Vera, U. M. R., Bernacchi, C. J., Siebers, M. H. & Ort, D. R. Canopy warming accelerates
826 development in soybean and maize, offsetting the delay in soybean reproductive
827 development by elevated CO₂ concentrations. *Plant Cell Environ.* **41**, 2806–2820 (2018).
- 828 61. Li, Y. *et al.* Elevated CO₂ Increases Nitrogen Fixation at the Reproductive Phase
829 Contributing to Various Yield Responses of Soybean Cultivars. *Front. Plant Sci.* **8**, 1546
830 (2017).
- 831 62. Gray, S. B. *et al.* Intensifying drought eliminates the expected benefits of elevated carbon
832 dioxide for soybean. *Nat. Plants* **2**, 16132 (2016).
- 833 63. Xu, G. *et al.* Soybean grown under elevated CO₂ benefits more under low temperature
834 than high temperature stress : Varying response of photosynthetic limitations, leaf
835 metabolites, growth, and seed yield. *J. Plant Physiol.* **205**, 20–32 (2016).
- 836 64. Makowski, D., Marajo-Petizon, E., Durand, J. L. & Ben-Ari, T. Quantitative synthesis of
837 temperature, CO₂, rainfall, and adaptation effects on global crop yields. *Eur. J. Agron.*
838 **115**, (2020).
- 839 65. Sacks, W. J., Deryng, D., Foley, J. A. & Ramankutty, N. Crop planting dates: an analysis of
840 global patterns. *Glob. Ecol. Biogeogr.* **19**, 607–620 (2010).

- 841 66. Ruane, A. C., Goldberg, R. & Chryssanthacopoulos, J. Climate forcing datasets for
842 agricultural modeling: Merged products for gap-filling and historical climate series
843 estimation. *Agric. For. Meteorol.* **200**, 233–248 (2015).
- 844 67. You, L., Wood, S., Wood-Sichra, U. & Wu, W. Generating global crop distribution maps:
845 From census to grid. *Agric. Syst.* **127**, 53–60 (2014).
- 846 68. Grassini, P. *et al.* Soybean yield gaps and water productivity in the western U.S. Corn Belt.
847 *F. Crop. Res.* **179**, 150–163 (2015).
- 848 69. Merlos, F. A. *et al.* Potential for crop production increase in Argentina through closure of
849 existing yield gaps. *F. Crop. Res.* **184**, 145–154 (2015).
- 850 70. Sentelhas, P. C. *et al.* The soybean yield gap in Brazil - Magnitude, causes and possible
851 solutions for sustainable production. *J. Agric. Sci.* **153**, 1394–1411 (2015).
- 852 71. Grassini, P., Eskridge, K. M. & Cassman, K. G. Distinguishing between yield advances and
853 yield plateaus in historical crop production trends. *Nat. Commun.* **4**, 2918 (2013).
- 854 72. Rubel, F., Brugger, K., Haslinger, K. & Auer, I. The climate of the European Alps: Shift of
855 very high resolution Köppen-Geiger climate zones 1800-2100. *Meteorol. Zeitschrift* **26**,
856 115–125 (2017).
- 857 73. Dupin, M. *et al.* Effects of the training dataset characteristics on the performance of nine
858 species distribution models: Application to *Diabrotica virgifera virgifera*. *PLoS One* **6**,
859 (2011).
- 860 74. Günther, F. & Fritsch, S. neuralnet: Training of neural networks. *R J.* **2**, 30–38 (2010).
- 861 75. Wright, M. N. & Ziegler, A. ranger: A Fast Implementation of Random Forests for High
862 Dimensional Data in C++ and R. *J. Stat. Software* **77**, 1–17 (2017).
- 863 76. Hastie, T. gam: Generalized Additive Models, R Package, version 0.98. *R Found. Stat.*
864 *Comput. Vienna, Austria.* (2013).
- 865 77. Minamikawa, K., Fumoto, T., Iizumi, T., Cha-un, N. & Pimple, U. Prediction of future
866 methane emission from irrigated rice paddies in central Thailand under different water
867 management practices. *Sci. Total Environ.* **566–567**, 641–651 (2016).
- 868 78. Rosenzweig, C. *et al.* The Agricultural Model Intercomparison and Improvement Project
869 (AgMIP): Protocols and pilot studies. *Agric. For. Meteorol.* **170**, 166–182 (2013).
- 870 79. Rosenzweig, C. *et al.* Assessing agricultural risks of climate change in the 21st century in a
871 global gridded crop model intercomparison. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 3268–3273
872 (2014).
- 873 80. Villoria, N. B. *et al.* Rapid aggregation of global gridded crop model outputs to facilitate
874 cross-disciplinary analysis of climate change impacts in agriculture. *Environ. Model. Softw.*
875 **75**, 193–201 (2016).
- 876 81. Guilpart, N., Iizumi, T. & Makowski, D. Soybean yield projections in Europe under
877 historical (1981-2010) and future climate (2050-2059 and 2090-2099 for RCP4.5 and
878 RCP8.5). [Dataset] Zenodo (2022). doi:<https://doi.org/10.5281/zenodo.6136216>