

## Rothamsted Repository Download

### A - Papers appearing in refereed journals

Coleman, K., Whitmore, A. P., Hassall, K. L., Shield, I. F., Semenov, M. A., Dobermann, A., Bourhis, Y., Eskandary, A. and Milne, A. E. 2021. The potential for soybean to diversify the production of plant-based protein in the UK. *Science of the Total Environment*. p. 144903.  
<https://doi.org/10.1016/j.scitotenv.2020.144903>

The publisher's version can be accessed at:

- <https://doi.org/10.1016/j.scitotenv.2020.144903>

The output can be accessed at: <https://repository.rothamsted.ac.uk/item/9832w/the-potential-for-soybean-to-diversify-the-production-of-plant-based-protein-in-the-uk>.

© 28 January 2021, Please contact [library@rothamsted.ac.uk](mailto:library@rothamsted.ac.uk) for copyright queries.

1 **The potential for soybean to diversify the production of plant-based**  
2 **protein in the UK**

3 Kevin Coleman<sup>a\*</sup>, Andrew P. Whitmore<sup>a</sup>, Kirsty L. Hassall<sup>b</sup>, Ian Shield<sup>a</sup>, Mikhail A.  
4 Semenov<sup>c</sup>, Achim Dobermann<sup>d</sup>, Yoann Bourhis<sup>a</sup>, Aryena Eskandary<sup>a</sup>, Alice E. Milne<sup>a</sup>

5 <sup>a</sup> *Sustainable Agriculture Sciences department, Rothamsted Research, Harpenden,*  
6 *Hertfordshire, AL5 2JQ, UK*

7 <sup>b</sup> *Computational and Analytical Sciences department, Rothamsted Research,*  
8 *Harpenden, Hertfordshire, AL5 2JQ, UK*

9 <sup>c</sup> *Plant Sciences department, Rothamsted Research, Harpenden, Hertfordshire, AL5*  
10 *2JQ, UK*

11 <sup>d</sup>*Directorate, Rothamsted Research, Harpenden, Hertfordshire. AL5 2JQ. UK.*  
12 *Present address: International Fertilizer Association, 75116 Paris France*

13

14 \*Corresponding author

15 E-mail address [kevin.coleman@rothamsted.ac.uk](mailto:kevin.coleman@rothamsted.ac.uk)

16

17 HIGHLIGHTS

- 18 • Data on novel soybean varieties was used to calibrate and validate the Rothamsted  
19 Landscape Model
- 20 • Simulations were run for 26 UK sites using current and future climate scenarios
- 21 • Under current climate early-maturing varieties matured in the south of the UK
- 22 • Under climate change soybean is predicted to mature as far north as Scotland
- 23 • No meaningful increases in yield are predicted under climate change

24

25

26

27 **ABSTRACT**

28 Soybean (*Glycine max*) offers an important source of plant-based protein. Currently  
29 much of Europe's soybean is imported, but there are strong economic and agronomic  
30 arguments for boosting local production. Soybean is grown in central and eastern Europe but  
31 is less favoured in the North due to climate. We conducted field trials across three seasons  
32 and two sites in the UK to test the viability of early-maturing soybean varieties and used the  
33 data from these trials to calibrate and validate the Rothamsted Landscape Model. Once  
34 validated, the model was used to predict the probability soybean would mature and the  
35 associated yield for 26 sites across the UK based on weather data under current, near-future  
36 (2041-60) and far-future (2081-2100) climate. Two representative concentration pathways, a  
37 midrange mitigation scenario (RCP4.5) and a high emission scenario (RCP8.5) were also  
38 explored. Our analysis revealed that under current climate early maturing varieties will mature  
39 in the south of the UK, but the probability of failure increases with latitude. Of the 26 sites  
40 considered, only at one did soybean mature for every realisation. Predicted expected yields  
41 ranged between 1.39 t ha<sup>-1</sup> and 1.95 t ha<sup>-1</sup> across sites. Under climate change these varieties  
42 are likely to mature as far north as southern Scotland. With greater levels of CO<sub>2</sub>, yield is  
43 predicted to increase by as much as 0.5 t ha<sup>-1</sup> at some sites in the far future, but this is  
44 tempered by other effects of climate change meaning that for most sites no meaningful  
45 increase in yield is expected. We conclude that soybean is likely to be a viable crop in the UK  
46 and for similar climates at similar latitudes in Northern Europe in the future but that for yields  
47 to be economically attractive for local markets, varieties must be chosen to align with the  
48 growing season.

49 **Keywords:**

50 Rothamsted Landscape model, soil processes, nutrient flow, soya bean, agriculture, future  
51 climate

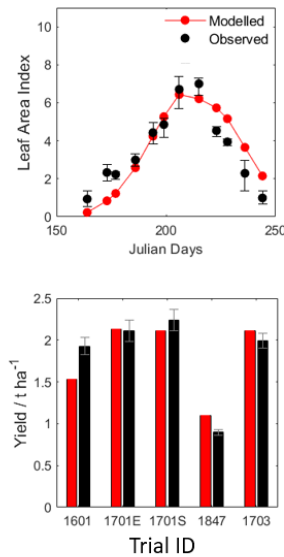
52

53 GRAPHICAL ABSTRACT

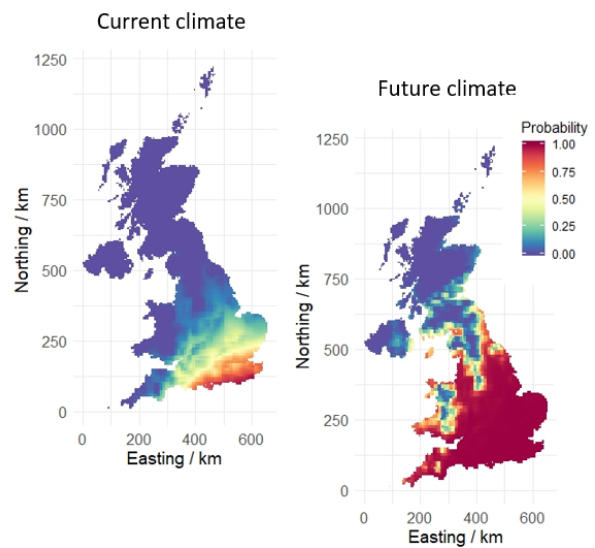
Field experiments using early maturing varieties



Model developed and validated



The probability of soybean maturing under current and future climate scenarios



54

55 **1. Introduction**

56 In 2019 the Eat-Lancet commission published a report that established clear scientific  
 57 targets to guide transformation to a healthier more sustainable food system (Willett et al.,  
 58 2019). At the top of the list of strategies to achieve this urgently needed change is a call to  
 59 increase the consumption of plant-based foods and substantially reduce the consumption of  
 60 animal source foods. This accords with the research of others who have quantified the relative  
 61 inefficiencies of meat-based food compared with plant-based (Reijnders and Soret, 2003;  
 62 Sabate and Soret, 2014; Springmann et al., 2016). Tessari et al. (2016) countered the  
 63 argument that plant-based proteins were less environmentally damaging than animal-based  
 64 proteins by comparing production based on the delivery of essential amino acids. They  
 65 demonstrate that animal production has a similar environmental impact to plant production on  
 66 an essential amino acid basis, with the exception for soybean (*Glycine max*), which has a  
 67 significantly smaller impact.

68 Globally, soybean is an important source of plant-based protein, with a percentage of  
 69 crude protein larger than many other legumes or pulses in commercial production (Cheng et  
 70 al., 2019). Total soya consumption in the UK is estimated to be 3.8 million tonnes, including

71 soya beans and meal, but also 0.7 million tonnes imported as soya embedded in other product  
72 (efeca, 2018). Currently much of Europe's soybean is imported from the United States and  
73 South America (European Commission, 2019), with only modest amounts of it grown in  
74 Europe itself, particularly in the southeastern and eastern regions of the European continent,  
75 as locally produced, non-GM soybean for feed and oil, or for premium markets such as organic  
76 food and fresh vegetables (IDH and IUCN NL, 2019). The EU non-GM soy market accounts  
77 for around 15% of the total feedgrade market and growing consumer concerns over  
78 environmental and animal welfare issues are expected to further segment the livestock feed  
79 market between conventional and premium feed. Hence, the search for alternative protein  
80 sources in Europe is driven by a desire to increase self-sufficiency in these market niches,  
81 which enable European soybean farmers to charge premiums of €80 to €120 per tonne of  
82 non-GM soybeans, with organic soy earning double this premium (Curtis et al., 2006).

83         Besides such economic incentives, there are other reasons for boosting more local  
84 production Direct consumption of soybean by humans is likely to rise due to shifts towards  
85 more plant-based diets (FarmingUK, 2018; Román et al., 2017; Tuorila and Hartmann, 2020).  
86 Moreover, European agriculture is in dire need for diversification and would greatly benefit  
87 from an economically viable, N-fixing legume that breaks the pest, competitor or disease  
88 cycles in the main cash crops that dominate current rotations. New agricultural policies in the  
89 EU as well as in the UK will likely stimulate agronomic measures that diversify cropping and/or  
90 benefit soil health and other ecosystem functions.

91         Soybean crops are grown in cold-temperate regions, such as the USA and Canada,  
92 as well as sub-tropical and tropical regions. Temperatures between 22 and 35°C are best  
93 suited for growth. If average temperatures fall below this then there is a delay in development  
94 lowering the chances of the crop reaching maturity. This is an issue for growing soybeans in  
95 Northern Europe. Despite this, soybean has been grown commercially in the UK since at least  
96 the late 1990's but take up has been limited because the available varieties were not well

97 suited to the UK climate and there were difficulties in harvesting. Recent advances in breeding  
98 mean that there are now more varieties that mature earlier (which is essential for the UK's  
99 colder, wetter climate) and have a canopy architecture that makes them easier to harvest.  
100 This means soybean could become a viable plant-based alternative source of protein for UK  
101 production systems.

102 As well as providing an alternative to animal-based protein (being relatively rich in the  
103 amino-acids Lysine and Methionine unlike most other legumes currently grown in Europe)  
104 there are several other benefits to growing soybeans in the UK. First, as a leguminous crop  
105 soybean can fix nitrogen reducing the need for fertilizer and increasing system-level N use  
106 efficiency. Second, with increasing resistance of weeds, slugs, insect-pests and diseases to  
107 chemical control agents, and the loss of active ingredients due to more stringent legislation,  
108 diverse crop rotation, including a spring sown protein crop such as soybean is becoming of  
109 increasing agronomic interest to UK farming. A key question facing farmers, however, is what  
110 is the likelihood that the crop will grow successfully, and can this crop be a profitable part of a  
111 diverse crop rotation now and in the future? Research trials can help answer these questions  
112 in part, but they are both expensive and time consuming and questions related to the effects  
113 of climate change become infeasible to test: therefore, we turn to models.

114 In this study, we set out to determine the spatial extent over which soybean is a viable  
115 crop in the UK based on the current climate, and to determine how this is likely to alter under  
116 climate change. For this we consider both the probability that early maturing varieties of  
117 soybean will mature, and the yield that could be expected. To achieve this we used data from  
118 field trials designed to test the viability of growing earlier maturing varieties of soybean in the  
119 UK to calibrate and validate the crop model in the Rothamsted Landscape Model (Coleman et  
120 al., 2017) for soybean. Once the model was validated, we used it with simulated weather data  
121 based on current and future climates for 26 sites across the UK to determine the probability  
122 that soybean crops would mature, and how this is affected by location and climate change.

123

## 124 **2. Methodology**

### 125 *2.1 Soybean trials*

126 Between 2016 and 2018, a total of six field trials were carried out at Rothamsted Research's  
127 experimental farms located in Harpenden, Hertfordshire, UK (51° 48' N, 0° 21' W), and Brooms  
128 Barn, near Bury St Edmunds Suffolk, UK (52° 16' N, 0° 34' E) to test the viability of early  
129 maturing soybean varieties under UK conditions. At each trial between two and twelve  
130 advanced breeding lines or varieties that had been developed in North America were grown  
131 in randomised replicated plot designs with variety as the treatment factor (S.I. Table S7). In  
132 2018 two European varieties were also tested at each site (full details are given in S.I. Table  
133 S8). The materials were chosen in consultation with breeders working in the more northern  
134 growing areas of North America, where the temperatures are lower, and the day-length is  
135 similar to that in the UK. The maturity groupings of each variety tested ranged between 000  
136 and 0 and are given in Table S8 (Song et al., 2019) For trial 1701 sowing time was also used  
137 as a treatments factor (see Table 1). No inorganic fertilizer was applied to the experiments,  
138 but the soybean seed was inoculated with *Bradyrhizobium japonicum* (Legume Technology,  
139 Nottinghamshire, UK). Standard herbicide and molluscicide programmes were applied to  
140 control weeds and slugs, and some bird protection was required. Little disease was detected.  
141 The soil at Rothamsted is described as silty clay loam (Batcombe series) by Avery and Catt  
142 (1991) and Aquic (or Typic) Paleudalf (Soil Survey Staff, 1999). The soil at Broom's Barn is a  
143 Sandy Loam belonging to the Moulton and Ashley Variant series. Both sites are research  
144 farms with closely monitored soil physical condition and nutritional status. As such, we found  
145 no notable nutrient deficiencies or soil physical impediments in the soil.

146 Soybean yields were measured on each of the six trials (see Table 1). The nitrogen  
147 (N) in the seed was measured in two trials (trial references 1702 and 1703). Leaf area index  
148 (LAI) was measured at two trials (trial references 1701 and 1702). To ensure we had both LAI  
149 and seed N measures in both the validation and calibration sets and to maximise site and

150 season diversity in both sets, we chose to use experiments 1601, 1701, 1703 and 1847 for  
 151 our calibration set and 1702, and 1848 for our validation set.

152 **Table 1**

153 Details of the six trials at Harpenden (H) and Brooms Barn (B). The trials used as our  
 154 validation set are marked by \*.

Trial ID	Year	Site	Field Name	Number of varieties grown	Sowing Dates	Seed Rates /seeds m <sup>2</sup>	Harvested†
1601	2016	H	Great Field 4	9	27 <sup>th</sup> April	45	22 <sup>nd</sup> September
1701	2017	H	Great Knott 3	2	3 <sup>rd</sup> and 28 <sup>th</sup> April	60	4 <sup>th</sup> October
1847	2018	H	Great Knott 3	6	25 <sup>th</sup> April	60	13 <sup>th</sup> November
1703	2017	B	Dun Holme	12	27 <sup>th</sup> April	60	17 <sup>th</sup> October
1702	2017*	H	Fosters	12	28 <sup>th</sup> April	60	4 <sup>th</sup> October
1848	2018*	B	Marl Pit	6	10 <sup>th</sup> May	60	19 <sup>th</sup> September

155 † Some trials were harvested over a number of days for practical reasons and the date given  
 156 is the earliest of the recorded dates.

157 *2.2 The soybean model*

158 The Rothamsted Landscape Model (Coleman et al., 2017) is a daily process-based  
 159 model that simulates soil processes (including soil organic matter, soil nutrient and water  
 160 dynamics), livestock production, crop growth and yield of cereals (wheat, barley, and oats),  
 161 oilseed rape, field beans, sugar beet, forage maize, potato, onions and grass. The crop model,  
 162 which is based on the LINTUL 5 model (Wolf, 2012), uses daily weather variables to predict  
 163 canopy development and resource accumulation. The weather data required to run the model  
 164 is minimum and maximum temperature, rainfall, solar radiation, vapour pressure and  
 165 windspeed. The model can be run as a point scale model or in a spatially explicit fashion with  
 166 adjacent pieces of land (fields or watercourses) linked to simulate spatial movement of water  
 167 and nutrients. The model components are based on well-established existing models such as  
 168 RothC (Coleman and Jenkinson, 2014), LINTUL (Wolf, 2012), SUCROS (van Laar et al.,  
 169 1997), and Century (Parton et al., 1994) as described in Coleman et al. (2017), and water  
 170 movement as described by Addiscott and Whitmore (1991) and Van Ittersum et al. (2003).



171 The crop model (which is based on LINTUL, Wolf (2012) is a generic plant growth  
172 model, which has a bespoke parameterisation for each crop modelled. It uses a light use  
173 efficiency (LUE, g dry matter MJ<sup>-1</sup>) based approach to calculate biomass production (Monteith,  
174 1990; Monteith and Moss, 1977). The rate of biomass ( $B_{crop}$ ) produced each day is given by

$$175 \quad \frac{dB_{crop}}{dt} = Q \epsilon W_{rf} N_{NI} P_{NI} F_{CO_2} \#(1)$$

176 where  $Q$  is the intercepted PAR (MJ PAR m<sup>-2</sup> surface area) which depends on the solar  
177 radiation and canopy leaf area,  $\epsilon$  is the crop specific LUE,  $W_{rf}$  is the transpiration reduction  
178 factor,  $N_{NI}$  and  $P_{NI}$  are nitrogen and phosphorus nutrition indices, which range from zero to  
179 one,  $F_{CO_2}$  is a CO<sub>2</sub> factor which allows dry matter production to change according to

$$180 \quad F_{CO_2} = 1.52 - 1.74 (0.9966^{CO_2}) \#(2)$$

181 where CO<sub>2</sub> is the atmospheric CO<sub>2</sub> in ppm. This function is based on that in Wolf (2012). The  
182 biomass formed is partitioned between roots, stem, leaves and storage organs based on the  
183 development stage ( $D$ ) which starts from zero at germination and finishes at a value of two  
184 which represents maturity (Boons-Prins et al., 1993; Wolf, 2012). Development stages  
185 accumulate as a function of photo-vernal-thermal time (as described in Wolf (2012) and Weir  
186 et al. (1984)).

187 The uptake of plant nutrients (N and P) is determined by the crop demand and the  
188 supply of these nutrients by soil. The total nutrient demand of the crop is the sum of the nutrient  
189 demand from its individual organs, i.e. roots, stems and leaves excluding storage organs, for  
190 which nutrient demand is met by translocation from the other organs. Note that in our version  
191 of this model, translocation from roots follows similar dynamics to that of stem and leaves to  
192 avoid cases where the stem and leaves become depleted of N whilst large amounts remain in  
193 the roots, in all cases the translocation rate was set to 1. Nutrient demand of the individual  
194 organs is calculated as the difference between maximum and actual organ nutrient contents.  
195 The maximum nutrient content is defined as a function of canopy development stage. For most  
196 crops including soybean, the total nutrient uptake of the crop takes place before anthesis. Sub-

197 optimal nutrient availability in the soil leads to nutrient stress in the crop. A detailed description  
198 of crop N dynamics is reported by Shibu et al. (2010). Further details for N and P are given in  
199 Coleman et al. (2017).

200 To model soybeans and their interaction with soil nutrient cycling, we included  
201 processes related to daily biological N fixation ( $N_{BNF}$ ). For this, we adopted the model  
202 described in Bouniols et al. (1991) and Williams et al. (1989). Biological N fixation ( $N_{BNF}$ ) is  
203 assumed to be

204

$$205 \quad N_{BNF} = \min [N_{Dem} f(D, w, N_{SMN}), B_{Max}] \#(3)$$

206

207 where  $B_{Max}$  is the maximum that  $N_{BNF}$  per day and assumed to take the value 6.0 (following  
208 LINTUL (Wolf, 2012)). The variable  $N_{Dem}$  is the total N demand of the crop,  $f(D, w, N_{SMN})$  is  
209 a function of crop development stage ( $D$ ), soil water ( $w$ ), and soil mineral-N content ( $N_{SMN}$ ),  
210 given by

$$211 \quad f(D, w, N_{SMN}) = g_{DVS}(D) \min [g_w(w), g_{SMN}(N_{SMN})] \#(4)$$

212

213 The functions  $g_{DVS}(D)$ ,  $g_w(w)$  and  $g_{SMN}(N_{SMN})$  are scaling factors;  $g_{DVS}(D)$  rises linearly  
214 from zero at  $D = 0.2$  to reach a maximum of one at  $D = 0.6$ . It then reduces linearly from a  
215 value of one at  $D = 1.2$  to zero at  $D = 1.6$ . Outside of the range  $D = (0.2, 1.6)$  it is zero  
216 (see Bouniols et al. (1991) noting that their development stages are scaled by a factor of 0.5  
217 compared with ours). The function  $g_w(w)$  is zero when  $w$  is less than 0.45 of the difference  
218 between field capacity and wilting point and rises linearly to a maximum of one at field capacity.  
219 The function  $g_{SMN}(N_{SMN})$  takes a value of 1 when the average  $N_{SMN}$  in the rooting depth of  
220 the soil is less than  $100 \text{ kg-N ha}^{-1} \text{ m}^{-1}$ , falling linearly to zero at  $300 \text{ kg-N ha}^{-1} \text{ m}^{-1}$ . We note  
221 that in the model our soil profile is assumed to be 1-m in depth (see Coleman et al., 2017).

222 The biological N fixed each day (BNF) is added to the N in the root, stem and leaves.  
223 The proportional split is based on the N already in each part of the plant. For example, the  
224 addition N partitioned to the leaves ( $\eta_{\text{leaf}}$ ) is given by

$$225 \quad \eta_{\text{leaf}} = N_{\text{BNF}} \frac{N_{\text{leaf}}}{N_{\text{leaf}} + N_{\text{stem}} + N_{\text{root}}} \#(5)$$

226  
227 where  $N_{\text{leaf}}$ ,  $N_{\text{stem}}$ , and  $N_{\text{root}}$  are the amounts of N in the leaf, stem and root prior to the daily  
228 addition of N from BNF. Santachiara et al. (2018) found no evidence to suggest that BNF  
229 constitutes a net extra energy cost to soybean crop in terms of growth or yield. Therefore,  
230 similar to Sinclair et al. (2003), we assume none in our model.

231

### 232 *2.3 Model parametrisation and calibration*

233 We used the soybean model parameter values reported in Wolf (2012) for our model.  
234 We noted, however, that the maximum N in the seed from trial 1703 experiments was larger  
235 than the value allowed by the existing parameterisation and so we increased the value of the  
236 parameter defining this from 5.6 % to 7.35%, which is the maximum seed N content of our  
237 calibration trial 1703. In addition, we expected the new varieties to have earlier flowering dates  
238 and a different canopy structure than those reported in Wolf (2012) and we noted that other  
239 LINTUL-based models of soybean (Corrêa, 2008) proposed smaller values for light use  
240 efficiency (LUE) and greater values for specific leaf area than those reported in the original  
241 LINTUL model. Therefore, we recalibrated the LUE, specific leaf area, anthesis and maturity  
242 parameters using the data from our experiments. Our aim was to minimise the root mean  
243 squared error (RMSE) between measured and modelled values of LAI and yield.

### 244 *2.4 Climate scenarios*

245 We ran the simulation model with weather data generated from current climate (1980-  
246 2010), near-future climate scenarios (2041-2060) and far-future climate scenarios (2081-  
247 2100) for 26 sites across the UK (see Fig. 1). The current climate was based on daily observed

248 weather data during 1981–2010. The summary statistics for temperature and precipitation at  
249 each site are listed in the S.I. Table S10. The future weather scenarios were based on climate  
250 projections from 18 global climate models (GCMs) from the multi-model ensemble used in  
251 IPCC Assessment Report 5 (AR5) (Taylor et al., 2012), two representative concentration  
252 pathways (RCPs), a midrange mitigation scenario (RCP4.5) and a high emission scenario  
253 (RCP8.5) (van Vuuren et al., 2011), and two future points in time (near 2041-60 and far 2081-  
254 2100 future). This resulted in four future climate sets which we refer to as (i) near-future-  
255 RCP4.5, (ii) near-future-RCP8.5, (iii) far-future-RCP4.5 and (iv) far-future-RCP8.5. To  
256 generate the local-scale future daily weather scenarios for each set, we used the LARS-WG  
257 weather generator (Semenov et al., 2010), a stochastic weather generator used in many  
258 recent European climate change impact and risk assessments (Trnka et al., 2015; Trnka et  
259 al., 2014; Vanuytrecht et al., 2014). For further details see (Semenov and Stratonovitch, 2015)  
260 and Harkness et al. (2020). Vapour pressure and windspeed, not generated by the LARS-WG,  
261 were estimated using methods described by the FAO (Allen et al., 1998).

262           Due to the coarse spatial and temporal resolution of GCMs and large uncertainties in  
263 the model outputs, it is not appropriate to use daily output from GCMs in combination with  
264 nonlinear process-based models when analysing impacts of changes in climatic variability and  
265 extreme weather events (Semenov et al., 2010). Therefore, for each of our 26 sites, we  
266 downscaled the climate projections from GCMs to local-scale daily climate scenarios by using  
267 LARS-WG, a stochastic weather generator (Semenov and Stratonovitch, 2010). LARS-WG  
268 downscales the projections from the GCMs to a local scale, incorporating changes in the mean  
269 climate, climatic variability and extreme events derived from the GCMs, by modifying the  
270 statistical distributions of the weather variables (Semenov, 2007).

271           For each [site] x [climate set] x [GCM], future synthetic daily weather data (300  
272 realisations of single weather years) were generated by the LARS-WG weather generator  
273 based on changes in distributions of climate variables derived from each GCM and emissions  
274 scenario. The CO<sub>2</sub> concentration for each climate sets listed in Table 2, along with the CO<sub>2</sub>

275 concentration assumed for the current climate set. To understand the relative effects of climate  
 276 change and increases in CO<sub>2</sub>, we also ran the model with current climate weather data and  
 277 the CO<sub>2</sub> concentration associated with far-future-RCP8.5 (i.e. 844). The model was run for  
 278 each year and the date of soybean maturity and yield were recorded. For soybean to be a  
 279 viable crop it must mature early enough to not disrupt the sowing of the next crop in the rotation  
 280 and also to avoid weather conditions unfavourable for drying the crop in the field, risking  
 281 difficult harvest conditions and expensive artificial drying of the crop. On the advice of our  
 282 agronomist (an author of this paper) we decided on a cut-off date of the 1<sup>st</sup> Oct with soybean  
 283 crops maturing before this date deemed viable. Based on this, the variables of interest in our  
 284 study are the probability that soybean will mature before 1<sup>st</sup> Oct and the yield. It should be  
 285 noted, however, that this is a conservative cut-off date, i.e. in many years harvest of soybean  
 286 and sowing of winter crops could still be feasible later in October.

287

288 **Table 2**

289 Concentrations of CO<sub>2</sub> (ppm) for current, RCP4.5 and RCP8.5. The current values are based  
 290 on measurements from 2017 and the future on those reported in Harkness et al. (2020).

	Current	RCP 4.5	RCP 8.5
2017	405		
2041 – 2060		487	541
2081 – 2100		533	844

291

292 *2.5 Statistical analysis*

293 For each Site by Climate combination, the probability of maturity was calculated as  
 294 the proportion of simulations (out of 300) that resulted in maturity before the 1<sup>st</sup> October.

295 Under future climate scenarios, this was averaged over the 18 GCMs,

$$P(\text{Maturity}) = \begin{cases} \frac{\# \text{ Mature}}{300}, & \text{if current scenario} \\ \frac{\frac{1}{18} \sum_{\text{GCM} = 1}^{18} \# \text{ Mature}_{\text{GCM}}}{300}, & \text{if future scenario} \end{cases} \quad \#(6)$$

297

298 For many Climate × Site combinations, the probability of maturity is estimated at the  
 299 boundaries of the [0, 1] interval. Thus, for consistency, confidence intervals for the  
 300 probability of maturity were obtained using the Clopper-Pearson (Clopper and Pearson,  
 301 1934) approach with interval defined by,

$$\text{Beta}\left(\frac{\alpha}{2}; x, n - x + 1\right), \text{Beta}\left(1 - \frac{\alpha}{2}; x + 1, n - x\right), \#(7)$$

303 where x is the numerator in Equation (6), n is the denominator and Beta (p,a,b) is the p<sup>th</sup>  
 304 quantile from a beta distribution with parameters a and b.

305 Yield was analysed only where maturity occurred, a total of 419 386 simulation runs. The  
 306 following linear model was fitted

$$\text{Yield}_i = (\text{Climate}_i \backslash (\text{RCP}_i * \text{Period}_i * \text{GCM}_i + \text{AtmCO2}_i)) * \text{Site}_i + \varepsilon_i \#(8)$$

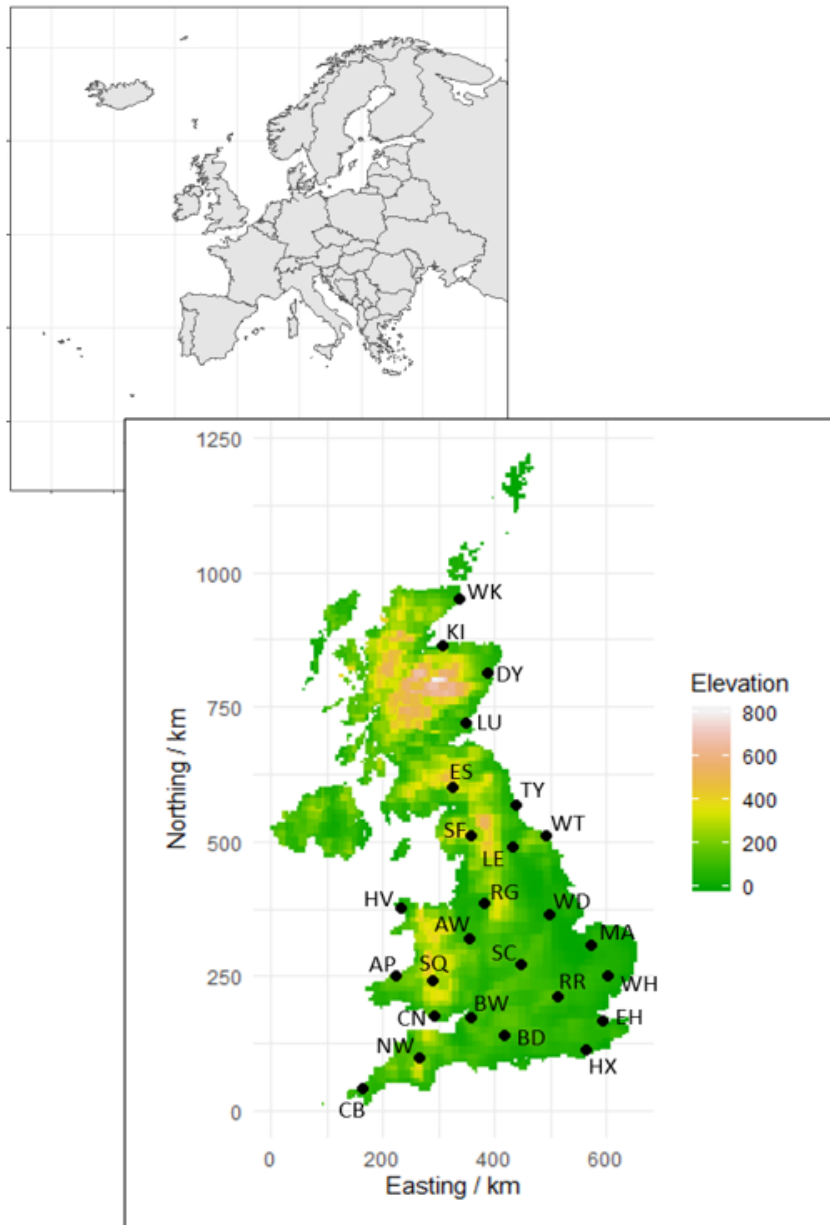
308 where  $\varepsilon_i$  are iid Normal random variables. The factor Climate has two levels; Climate and  
 309 Future. Levels of the factors RCP, Period and GCM vary only in Future climate scenarios,  
 310 whilst levels of AtmCO2 vary only in Current climate scenarios. High levels of imbalance in  
 311 the number maturing results in unequal numbers of yield observations across the different  
 312 factors. Consequently, results are analysed through both the marginal (respecting marginality)  
 313 and conditional F-statistics.

314 Clopper-Pearson intervals were calculated in the R software environment (RStudio Team,  
 315 2020). The linear model for Yield was fitted in Genstat 20<sup>th</sup> edition (VSN International, 2019).

316 Maturity is based on climatic data and day length, and so longitude, latitude and  
 317 elevation are plausible covariates to support spatial prediction. Therefore, to support spatial

318 prediction (mapping) of the probability of maturity we fitted linear models to the logit of the  
319 probability of soybean maturing with these covariates as explanatory variables, and then used  
320 these covariates to predict the probability of maturity across the UK.

321



322

323 **Fig. 1.** A map of the UK showing the location of the climate stations (black dots) that were  
324 used in the simulations. The map was produced using the R software. We use OSGB cartesian  
325 co-ordinates as measures of easterly and northerly distance. The location of the UK in Europe  
326 is shown the in inset pane.

327

### 328 **3. Results**

#### 329 *3.1 Soybean trials*

330           The soybean crop successfully matured in all field trials conducted. Yields ranged  
331 between 0.4 t ha<sup>-1</sup> in 2018 to 2.9 t ha<sup>-1</sup> in 2017 with an average of 1.7 t ha<sup>-1</sup> (Table 3 and  
332 Supplementary Information). In general, yields at Brooms Barn were greater than those at  
333 Harpenden. In all experiments, varietal performance differed significantly in terms of yield  
334 (Table S8).

335           Across the 2016 and 2017 variety trials the highest yielding cultivars gave moderate  
336 yields with the means over replicate plots having maximum of 2.72, 2.34 and 2.61 t ha<sup>-1</sup>. Yields  
337 in 2018 were substantially lower with a maximum of 1.08 t ha<sup>-1</sup> (Harpenden) and 1.61 t ha<sup>-1</sup>  
338 (Brooms Barn) primarily due to the exceptionally dry weather during the months of June and  
339 July (see Supplementary Information Tables S2–6) which affected the soil moisture. Despite  
340 the reasonable water holding capacity of the silty clay loam soil at Harpenden it is not unusual  
341 for later spring sown crops to suffer from drought as rooting fails to extend sufficiently rapidly  
342 to maintain water supply to the plant.

343           Analysis of trial 1701 showed significant differences in yield according to sowing date  
344 ( $F_{1,3}=24.15$ ,  $p=0.016$ ) with late drilling yielding an average of 0.18 t ha<sup>-1</sup> more. Given that  
345 variety was not accounted for in our model we calibrated our simulations to the mean values  
346 of yield and seed N across varieties for each site, season and sowing time (Table 3 and 4). A  
347 complete analysis of the trials data for all years is given in the Supplementary Information.

348           There was no consistent response of variety between seasons, and this is  
349 disappointing from the point of view of selecting well adapted genetics for UK agriculture. It  
350 was noted that the rhizobium applied to the seeds in 2016 was poor quality (the peat-based  
351 carrier had dried out) resulting in few root nodules and low seed nitrogen contents (data not  
352 presented). The seed of two of our varieties sown in 2016 was poor quality and this was



353 reflected in low plant counts (see Supplementary Information Table S8, varieties Canada 4  
354 and 6). Fresh seed was sown in 2017 and a new, liquid, formulation of rhizobium was applied.

355

356 **Table 3**

357 The summary statistics for soybean seed yield for each trial. The trials used as for validation  
358 are marked by \*.

Trial ID	Sowing time	Seed yield / t ha <sup>-1</sup> at 14% moisture content					
		Mean	Variance	Number of plots	Standard error	Min	Max
1601	Standard	1.929	0.292	27	0.104	0.860	2.822
1701	Early	2.113	0.0951	8	0.109	1.645	2.527
1701	Standard	2.235	0.182	8	0.151	1.786	2.805
1703	Standard	1.992	0.271	33	0.0907	0.491	2.888
1847	Standard	0.898	0.0390	30	0.0360	0.392	1.325
1702*	Standard	1.714	0.178	33	0.0734	0.808	2.490
1848*	Standard	1.283	0.0729	30	0.0493	0.639	1.696

359

360 **Table 4**

361 The summary statistics for soybean seed N. The trials used for validation are marked by \*.

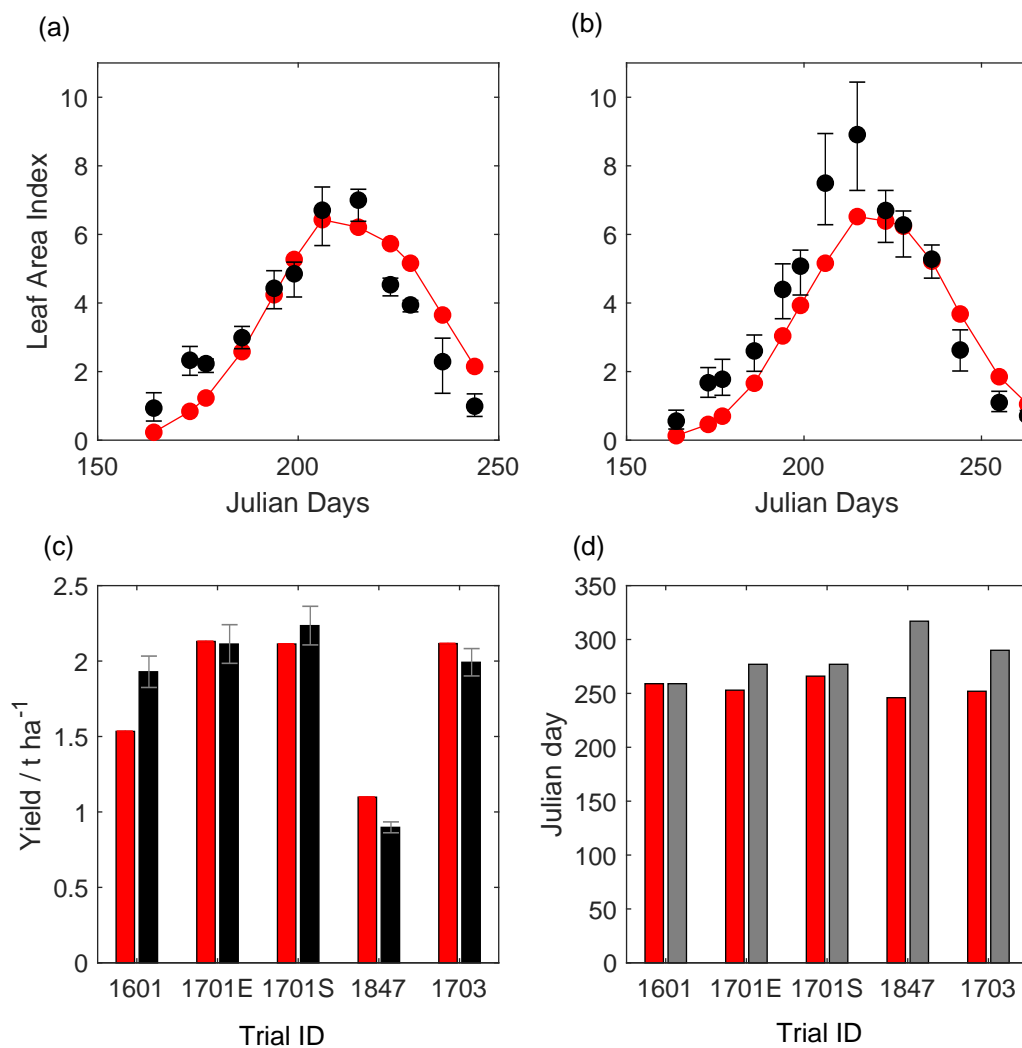
Trial ID	Seed N / %					
	Mean	Variance	Number of plots	Standard error	Min	Max
1702*	6.600	0.121	33	0.0607	5.912	7.1
1703	6.670	0.133	33	0.0608	6.013	7.345

362

363 *3.2 The soybean model calibration and validation*

364 The smallest RMSE between observed and predicted LAI and yield results (Fig. 2) when the  
365 LUE equals 1.6, specific leaf area equals 0.03, photo-vernal-thermal time for anthesis (DVS =

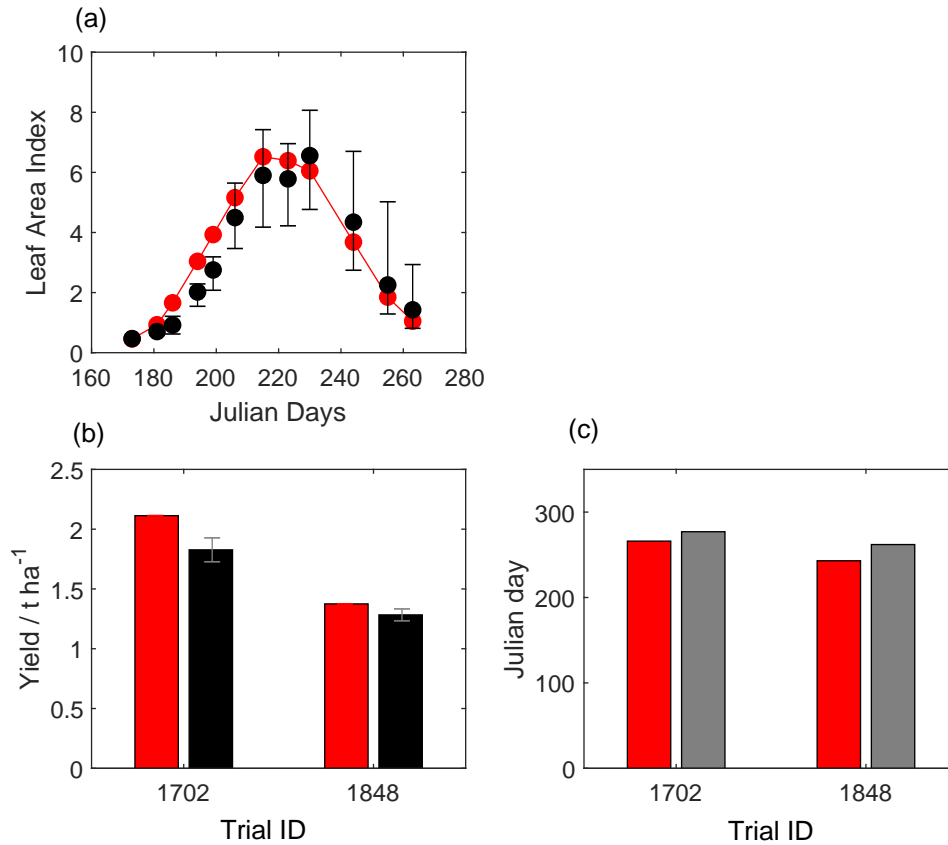
366 1) equals 745 and maturity is a further 400 units of photo-vernal-thermal. Validation sets  
 367 performed consistently well (Fig. 3). Modelled biological N fixation, crop N uptake and N in  
 368 the seed are shown in Table S9. In the model the low yields were clearly caused by water  
 369 stress and lower levels of biological fixation also resulting from the unfavourable soil moisture  
 370 conditions (Table S9, trials 1847 and 1848 and Table S4 and S6). Our biggest discrepancy in  
 371 predicted date of maturity was site 1847, where the observed crop was harvested much later  
 372 than predicted.



373  
 374 **Fig. 2.** Modelled (red) and measured (black) (a) leaf area index for early sown soybean in  
 375 experiment 1701. The bars show the range of observations from four replicates. (b) leaf area  
 376 index for standard sown soybean in experiment 1701. The error bars show the range of our

377 four observations. (c) mean yield across experiments with standard error bars. (d) the  
 378 modelled maturity date and measured harvest date (grey) which indicated an upper bound for  
 379 maturity.

380



381  
 382 **Fig. 3.** Modelled (red) and measured (black) (a) leaf area index with bars showing the range  
 383 of observations from six replicates. (b) mean yield across experiments with standard error  
 384 bars. (c) the modelled maturity date and measured harvest date (grey) which indicated an  
 385 upper bound for maturity.

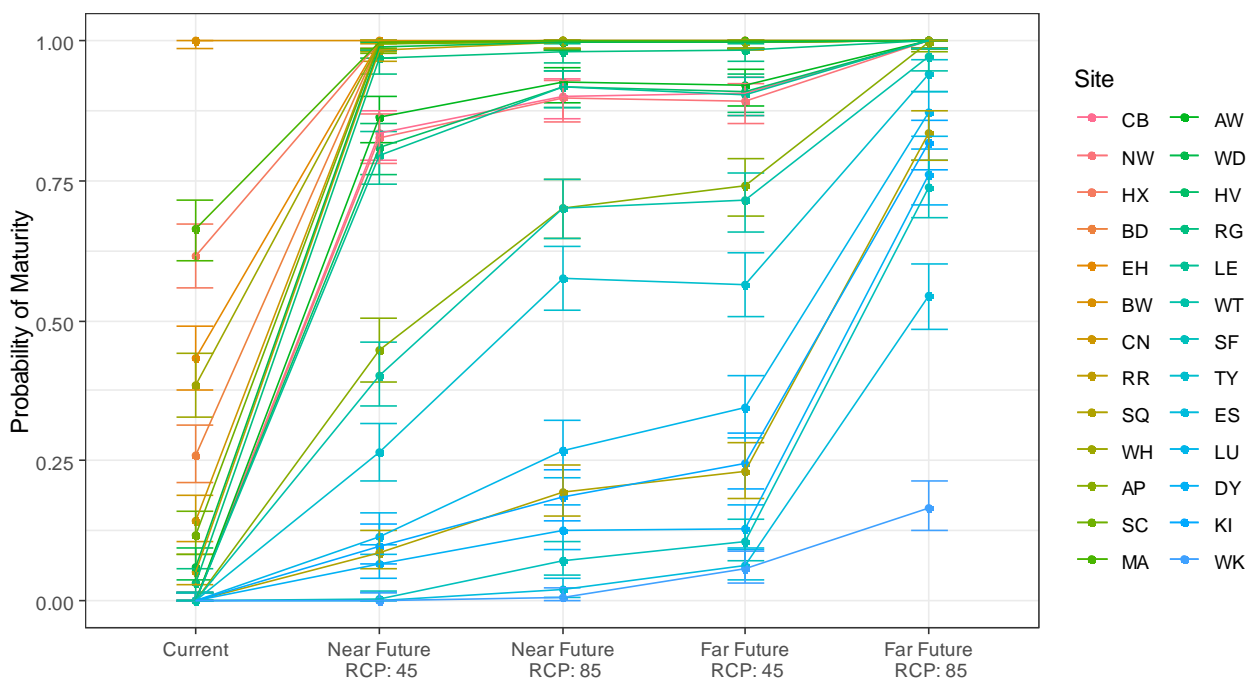
386

### 387 3.3 Scenario Results

388 Under the current climate scenario only a single site (BD) could guarantee maturity by 1<sup>st</sup>  
 389 October (95% CI: 0.988, 1.000), whilst under the most extreme climate scenario (far-future-  
 390 RCP8.5) 16 sites matured 100% of the time, and only a single site (WK) matured less than

391 half of the time (95% CI: 0.125, 0.213). In the more southerly sites, the greatest increase in  
 392 probability of maturity is seen in the near-future-RCP4.5, scenario (Fig. 4). Little difference is  
 393 observed between near-future-RCP8.5 versus far-future-RCP4.5 (Fig. 4). In the more  
 394 northerly sites, there is a general trend of increasing the probability of maturity from near-  
 395 future-RCP4.5 to far-future-RCP8.5 (Fig. 5). See Fig. S5 in the Supplementary Information for  
 396 the numbers of simulations that successfully matured under each climate scenario. The spatial  
 397 predictions illustrate clearly that the probability of maturity increases under future climate  
 398 predictions, particularly in the south (Fig. 5). See Supplementary Information (Table S11, Fig.  
 399 S6, S7) for the parameters of the spatial models and maps of predictions of the probability that  
 400 soybean crops will mature and associated errors of prediction.

401



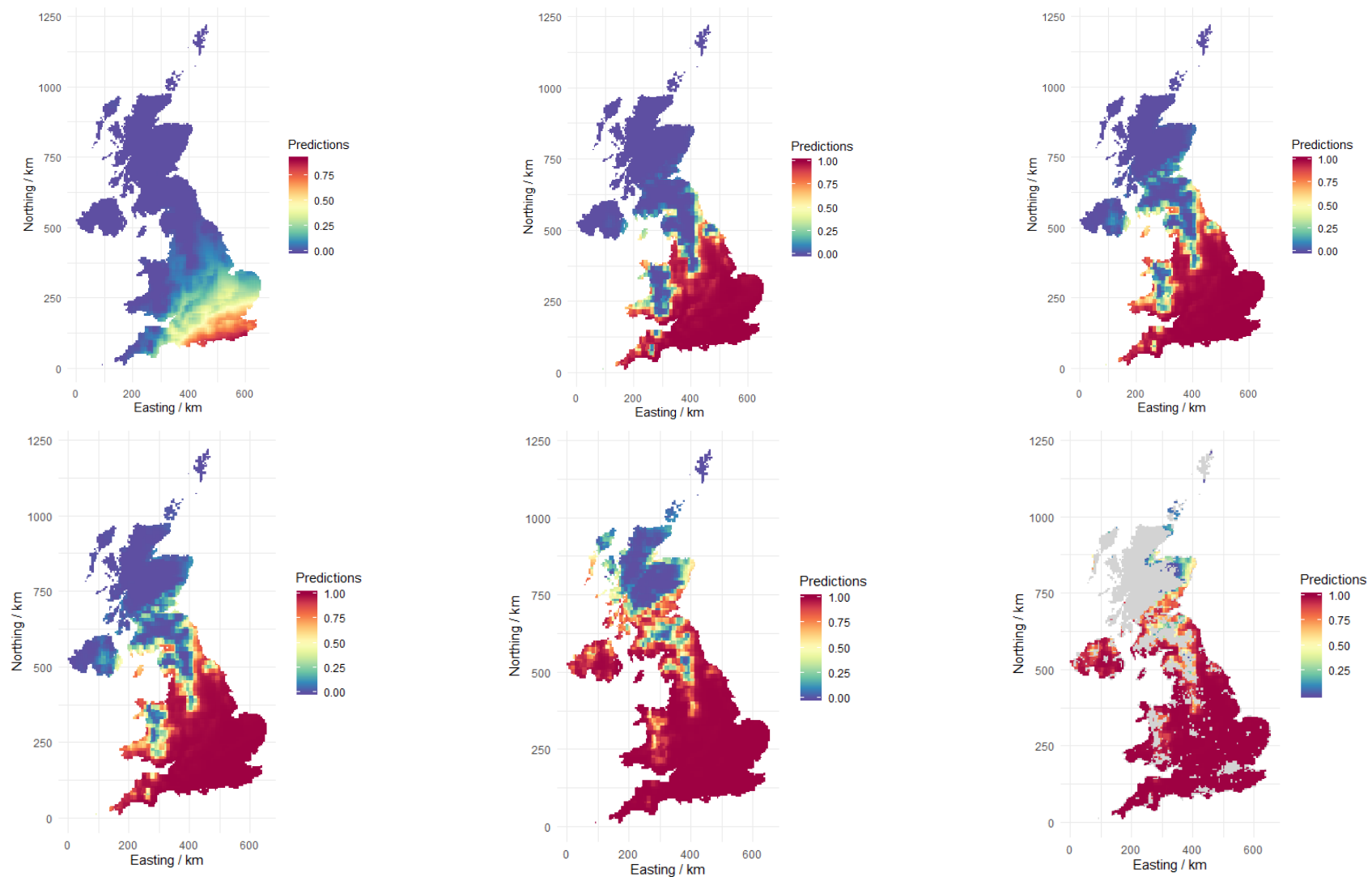
402

403 **Fig. 4.** Probability of maturity calculated for each climate scenario. Error bars are the 95%  
 404 Clopper-Pearson confidence intervals. Colour indicates the Site, with colour scale defined by  
 405 the order of latitude (Red=Southernmost site and Blue=Northernmost site).

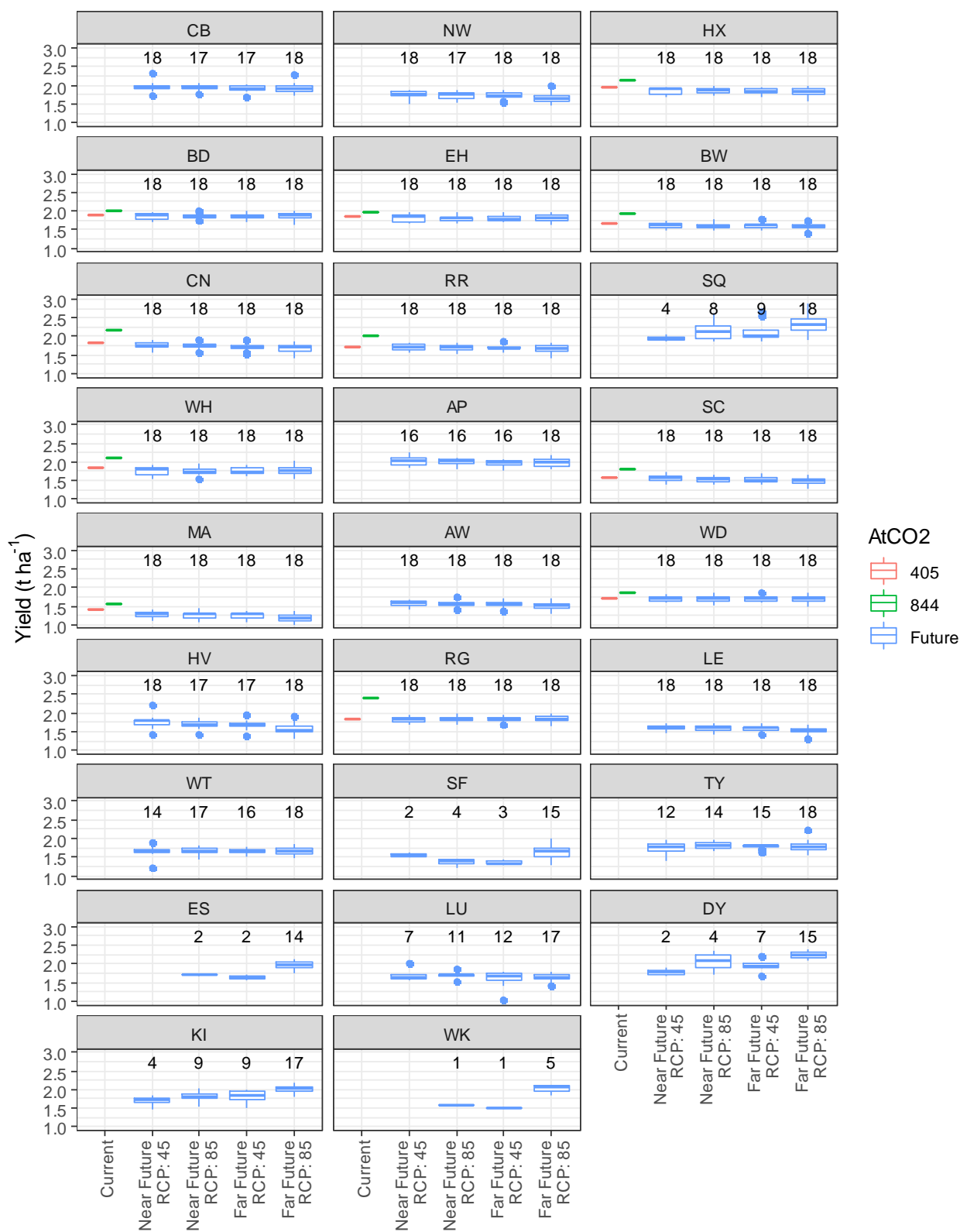
406 Although future climate scenarios predict an increase in the probability of soybean maturing,  
 407 the magnitude of the associated yields is less certain (Fig. 6, Supplementary Information Fig.  
 408 S5). Investigating the partition of variability between the different simulation scenarios (Table

409 5), location is the main factor for different yield predictions (marginal  $F_{25,417825} = 7261$ ) ranging  
410 from an average (over all climate scenarios) of 1.23 t ha<sup>-1</sup> (MA) to 2.16 t ha<sup>-1</sup> (SQ). It is clear  
411 that where maturity can already be reached under the current climate, substantive increases  
412 in yield are expected with increasing atmospheric CO<sub>2</sub> ( $F_{1,417825} = 339$ ). However, given the  
413 large variation observed from different GCMs (marginal  $F_{1,417825}=1334$ ), little overall effect can  
414 be observed in the 4 future climate scenarios on the predicted soybean yield. When  
415 accounting for site to site variation and the variation due to GCMs, future time period  
416 (conditional  $F = 903$ ) has a larger impact on yield predictions than RCP (conditional  $F = 274$ )  
417 overall. We note that there are levels of confounding between these variables and so caution  
418 against over interpretation. The yield trends over future climate scenarios are not consistent  
419 across sites (Fig. 7). There is a slight decrease in yield at the majority of the sites as RCP  
420 changes from 4.5 to 8.5 or when period changes from 2041–2060 to 2081–2100. However,  
421 there is a small subset of sites where the yield increases substantially. In general, those sites  
422 with large predicted increases in yield (SQ, ES, KI, SF, DY, WK) are also the sites with least  
423 probability of maturing.

424



**Fig. 5.** Predictions of the probability that soybean crops will mature for (a) current weather, (b) near-future-RCP4.5, (c) near-future-RCP8.5, (d) far-future-RCP4.5, (e) far-future-RCP8.5, (f) far-future-RCP8.5 with areas where crops are not currently grown masked out (grey).

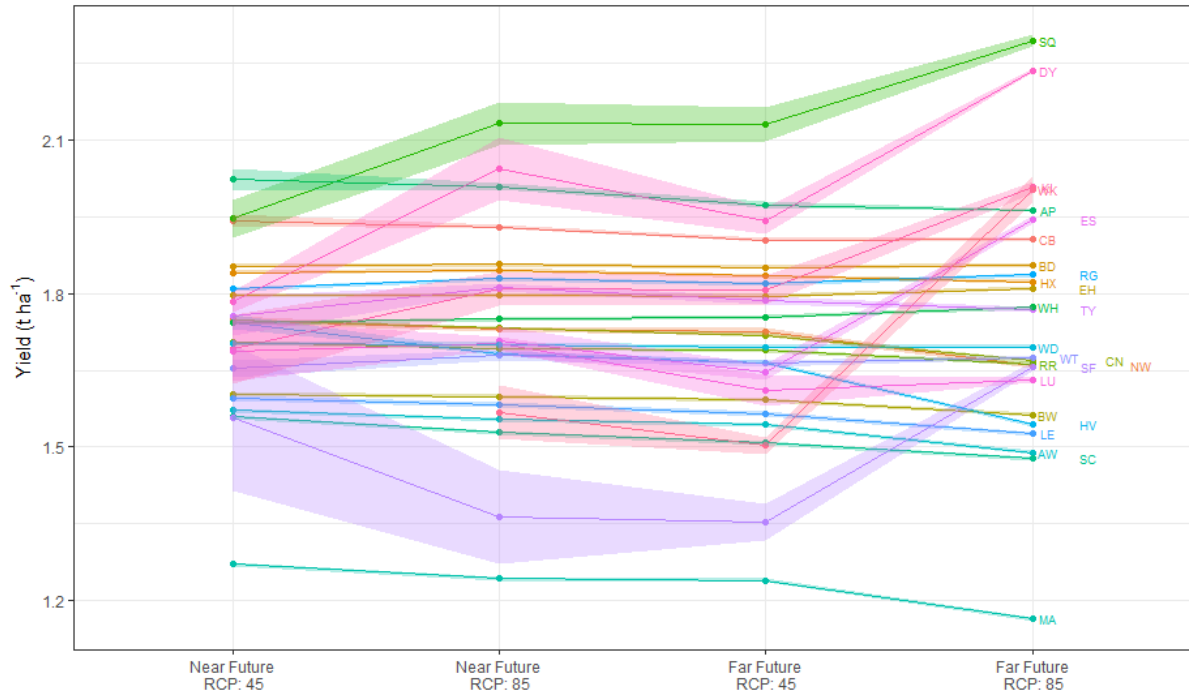


425

426 **Fig. 6.** – Boxplots of the expected yield under different climate scenarios at 26 locations in the  
 427 UK. Values shown under current climate are the average of up to 300 individual simulations.  
 428 Boxplots under future scenarios are constructed from up to 18 individual values (actual values  
 429 indicated above each box in figure), one per GCM, each of which is the result of averaging  
 430 over a maximum of 300 individual simulations.

431

432



433

434

435

436

437

438

439

440

441

**Fig. 7.** Average yield per site under each future climate scenario after having adjusted for GCM (points). Predictions are obtained by first forming the full table of predictions for all Site x GCM x RCP x Climate combinations that are present and then by averaging over GCM. The shading around each point indicates the standard errors based on marginal weights, which here reflects the number of unique GCMs for each scenario. We note that the interpolations between each point are an to aid visual interpretation but have no physical meaning.



442

443 **Table 5**

444 F-statistics assessing the effect of each term on soybean yield. Marginal F-statistics are  
 445 associated with including that term to the simplest possible model (respecting marginality),  
 446 e.g. For a model fitting  $A + B + A.B$ , the marginal statistic for A is associated with fitting only  
 447 A, the marginal statistic for B is associated with fitting only B, the marginal statistic for A.B is  
 448 associated with A.B after fitting the respective main effects A and B. Conditional F-statistics  
 449 are associated with including that term to the most complicated model (excluding terms to  
 450 which it is marginal). E.g. for a model fitting  $A + B + A.B$ , the conditional statistic for A is  
 451 associated with fitting A after accounting for B, the conditional statistic for B is associated  
 452 with fitting B after accounting for A, the conditional statistic for A.B is associated with fitting  
 453 A.B after fitting A and B.

Term	Marginal F Statistic	Conditional F Statistic	ndf	ddf (full model)
Climate	368.54	1072.42	1	417825
Site	7261.41	7529.58	25	
Climate.RCP	585.61	274.75	1	
Climate.Period	118.82	903.93	1	
Climate.GCM	1334.72	1529.34	17	
Climate.AtCO2	339.24	339.24	1	
Climate.Site	8.46	8.47	10	
Climate.RCP.Period	365.99	96.56	1	
Climate.RCP.GCM	62.58	86.88	17	
Climate.Period.GCM	76.32	102	17	
Climate.Site.RCP	259.86	83.73	25	
Climate.Site.Period	202.45	71.53	25	
Climate.Site.GCM	46.3	40.47	400	
Climate.Site.AtCO2	3.19	3.19	10	
Climate.RCP.Period.GCM	78.6	115.86	17	
Climate.Site.RCP.Period	32.32	24.66	23	
Climate.Site.RCP.GCM	4.8	5.38	333	
Climate.Site.Period.GCM	4.59	5.43	329	
Climate.Site.RCP.Period.GCM	2.92	2.92	305	

454

455

#### 456 **4. Discussion**

457 Our results suggest that by 2050 soybean should be a viable crop across most of  
458 England and south Wales under both RCP scenarios. In southern England the soybean early-  
459 maturing variety parameterised in our model is predicted to be certain to mature and so it is  
460 extremely likely that varieties that mature later will also be viable. This could have implications  
461 for increased yield as the growing season would be extended. To test this further we would  
462 need to calibrate the soybean model for these different types of variety, including a maturity  
463 group specific functions of the effect of daylength on development (Setiyono et al., 2007). Only  
464 after 2040 and with the RCP85 prediction does soybean appear viable for Scottish agriculture.

465 A number of soybean models exist in the literature (Jego et al., 2010; Sinclair et al.,  
466 2003). These range from quite complex models such as the CROPGRO-soybean model  
467 (Hoogenboom et al., 1992; Jones et al., 2003) which has successfully simulated a number of  
468 cultivars and in a range of environments including Australia and across the USA (sites ranging  
469 from Florida to Idaho), and SOYDEV (Setiyono et al., 2010; Setiyono et al., 2007; Setiyono et  
470 al., 2008) which was developed to simulate soybean development under high-yield conditions  
471 of North-Central U.S. Corn Belt, to simpler models such as Sinclair-Soybean (Setiyono et al.,  
472 2010; Sinclair et al., 2003). Soybean have also been parameterised in the WOFOST (Abadi  
473 et al., 2018) and LINTUL (Corrêa, 2008) crop models for studies in Indonesia and Brazil  
474 respectively. Most parameterisations are cultivar specific, although efforts have been made  
475 to make models more parsimonious by parameterising according to maturity grouping and  
476 cultivar stem termination type (Setiyono et al., 2010). For our analysis we chose to use a more  
477 generic model to simulate our early maturing varieties, and to avoid issues of overfitting, the  
478 model parameter values were largely based on crop physiology-based values from the  
479 literature with only four parameters fitted to the data. The data provided by our trials proved a  
480 good resource to parameterise our model. In particular, data on canopy expansion (which are  
481 relatively rare) gave us confidence that the crop development was captured by the model.  
482 The calibrated model was able to reproduce the canopy expansion and decline well (Fig. 2)

483 and accurately predicted the variation in expected yield across the seasons. This was born  
484 out in our model validation (Fig. 3). In particular, our simulation predicted the drought  
485 conditions in the soil in 2018 (as described above) and the poor yields that resulted (Table 3,  
486 trials 1847 and 1848). Our values of crop N uptake ( $156 \text{ kg N ha}^{-1}$ , S.I. Table S9 average of  
487 all trials) are in accordance with those reported in Bender et al. (2015), which were  $164 \text{ kg N}$   
488  $\text{ha}^{-1}$  for Soybean yielding  $2 \text{ t ha}^{-1}$ . In our model we chose not to include an energy cost to the  
489 plant for biological N fixation. Whilst we acknowledge that any form of BNF in crops has an  
490 energy cost associated with it (Liu et al., 2011; Minchin and Witty, 2005; Vance and Heichel,  
491 1991), Santachiara et al. (2018), found no evidence to suggest that BNF constitutes an extra  
492 energy cost to soybean crops in terms of growth or yield. This suggests that under  
493 agronomically relevant conditions this energy cost is somewhat compensated for, and that it  
494 does not substantially alter the relationship between crop biomass and crop N accumulation,  
495 particularly when yields and N uptake are relatively low (as observed in our field trials). We  
496 note however, that some models represent such C-N interactions in more detail, whereas  
497 others do not and treat N uptake more independently (see Fisher et al., 2010). Tamagno et al.  
498 (2018) list a number of mechanisms by which soybean might yield as well from BNF as it does  
499 from chemical fertiliser: increased photosynthesis, the availability of N throughout growth as  
500 opposed to dosage at a specific time and change in the nitrogen harvest index. In their  
501 experiments, however, Tamagno et al. (2018) found that at the crop level, soybean met the  
502 cost of BNF by a reduction in seed yield mediated by lower harvest index (HI), particularly in  
503 stressful environments, and a secondary contribution from reduced seed oil concentration.  
504 The soybean crops in our experiments did not receive fertiliser N and so we are unable to  
505 assess the contribution of fertilizer N compared to BNF. Given the low yields and the lack of  
506 data on the cost of BNF, we chose to disregard it in our model, which is also in line with how  
507 other soybean models have treated N fixation, uptake and partitioning. See, for example,  
508 Sinclair et al. (2003). Should yields improve through breeding or climate change, it might  
509 become necessary to revisit this part of the model and determine what mechanisms, if any,  
510 compensate for the carbon cost of N fixation.

511           The yields from our experiments ranged between 0.4 t ha<sup>-1</sup> and 2.9 t ha<sup>-1</sup> with an  
512 average of 1.7 t ha<sup>-1</sup> (Table 3 and Supplementary Information). These yields are slightly low  
513 within the context of global and European average yields which are reported to be closer to  
514 2.8 t ha<sup>-1</sup> and 2.08 t ha<sup>-1</sup> respectively (Terzić et al., 2018). It follows that our predicted yield  
515 across the UK are generally low (for current climate average yields for a given site-year range  
516 between 0.9 and 2.0 t ha<sup>-1</sup>). There was no obvious spatial pattern in determining where yields  
517 were likely to be greatest under current climate, and this is likely to be because yield depends  
518 on both soil and weather (unlike phenological timing which is driven by temperature and  
519 daylength). Our predictions show that increasing CO<sub>2</sub> levels will have a significant effect on  
520 yield increase (Fig 6). However, there is a slight decrease in yield at the majority of the sites  
521 as RCP changes from 4.5 to 8.5 or when Period changes from 2041–2060 to 2081–2100. This  
522 suggests that the effects of water and heat stress may compensate for the positive effects on  
523 yield of increased CO<sub>2</sub>. These factors could be addressed through variety choice and breeding  
524 as explored by Semenov (2009) for wheat crops in the UK.

525           Despite the observed and predicted low yields of soybean in UK conditions the crop  
526 may still be financially viable for farmers. Besides land rental, the operational cost of  
527 production of soybeans is currently modest. Few pests or diseases have been observed  
528 meaning that no pesticides other than herbicides are needed, although it is acknowledged that  
529 growing a greater area of soybean is likely to result in greater pest and disease incidence  
530 (Engering et al., 2013; Legg, 1999). Based on the estimated price of soybean and associated  
531 variable costs a 2 t ha<sup>-1</sup> crop could result in a gross margin of 468 £ ha<sup>-1</sup> which is comparable  
532 to the profit margin of field beans (Nix, 2020; Soya UK, 2018). When the rotational benefits of  
533 soybeans (as described in the introduction) are also considered the crop is an attractive  
534 proposition for farmers. More viable, however, would be a scenario in which soybean  
535 consistently yields around 2.5-3 t ha<sup>-1</sup> under UK conditions. Our experiments suggest that this  
536 is possible in principle, but will require further genetic and agronomic fine-tuning.

537           Hence, a key question is what are the major crop phenological or physiological  
538 constraints that need to be overcome to make soybean a competitive crop in the UK and other  
539 parts of Northern Europe nearby maritime Northern Europe that are at similar latitudes to the  
540 UK but have slightly warmer summers and so where soybean is equally or more likely to  
541 mature. Our canopy measurements showed that peak LAI values were similar to crops grown  
542 in Nebraska, USA that typically yield 4.5 – 5 t ha<sup>-1</sup> (Setiyono et al., 2008). That is to say,  
543 canopy development and closure did not seem to play a major role in limiting the yields we  
544 observed. Setiyono et al. (2008) found that their green leaf persisted longer than ours; this  
545 may be because they irrigated their crops. It is worth noting that our experiments report on  
546 early-developing varieties chosen for the current UK climate. To our knowledge, there are no  
547 breeding programmes for soybean in the UK at this current time, which raises the question for  
548 breeders of whether it is possible to breed varieties that retain green leaf for longer than at  
549 present. In the future, the last frost day in spring is likely to occur up to one month earlier than  
550 now (data not shown). Although this does not necessarily translate into one month's earlier  
551 sowing and longer growing season, it suggests that later developing, and potentially, higher  
552 yielding varieties will become viable in the UK and other Northern European countries in  
553 coming decades. There is also need to better tailor agronomic practices of growing soybean  
554 to UK soil and climatic conditions. Tillage, row spacing, seed rate, inoculation, starter fertilizer  
555 along with the seed, other nutrient applications, or irrigation are all practices we did not study  
556 in our work, but which are likely to be critical for exploiting the attainable yield potential.

## 557 **5. Conclusions**

558           Model-based prediction shows that early maturing varieties of soybean can be grown  
559 in the UK at latitudes lower than approximately 52.3°, although yields are slightly less than the  
560 average for other European countries. Under climate change, the potential for successfully  
561 growing soybean increases enormously, with predictions under far-future-RCP8.5 suggesting  
562 the crop could be viable as far North as southern Scotland with site DY (latitude 57.21 and  
563 longitude -2.2) predicted to mature 76% of the time.

564 Yields are expected to respond positively to increases in CO<sub>2</sub> with average increases  
565 associated with CO<sub>2</sub> only ranging from 9.1% (site EH) and 29.4% across sites (site RG), but  
566 this is tempered by increased water stress due to more evaporation meaning that only certain  
567 sites might see a positive effect of climate change on yield. With climate change, however,  
568 varieties that mature later will become viable in the south and this will also have positive  
569 implications on yield potential.

#### 570 **CRedit authorship contribution statement**

571 **Kevin Coleman:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data  
572 Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, Project administration.

573 **Andrew P. Whitmore:** Conceptualization, Methodology, Software, Investigation, Writing - Original Draft, Writing -  
574 Review & Editing, Supervision, Project administration, Funding acquisition.

575 **Kirsty L. Hassall:** Methodology, Software, Formal analysis, Investigation, Data Curation, Writing - Original Draft,  
576 Writing - Review & Editing, Visualization.

577 **Ian Shield:** Conceptualization, Methodology, Validation, Investigation, Data Curation, Writing - Original Draft,  
578 Writing - Review & Editing, Funding acquisition.

579 **Mikhail A. Semenov:** Methodology, Software, Investigation, Writing - Original Draft, Writing - Review &  
580 Editing.

581 **Achim Dobermann:** Conceptualization, Methodology, Writing - Review & Editing, Funding acquisition.

582 **Yoann Bourhis:** Software, Investigation, Writing - Review & Editing.

583 **Aryena Eskandary:** Software, Investigation, Writing - Review & Editing.

584 **Alice E. Milne:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data  
585 Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, Project administration,  
586 Funding acquisition.

587

#### 588 **Declaration of competing interest**

589 The authors declare that they have no known competing financial interests or personal  
590 relationships that could have appeared to influence the work reported in this paper.

591

#### 592 **Acknowledgements**

593 Rothamsted Research receives grant aided support from the Biotechnology and Biological  
594 Sciences Research Council (BBSRC) of the United Kingdom. This research was supported  
595 by the Biotechnology and Biological Sciences Research Council (BBSRC) Institute Strategic  
596 Programme (ISP) grants, “Soils to Nutrition” (S2N) grant number BBS/E/C/000I0330, the joint  
597 Natural Environment Research Council (NERC) and the Biotechnology and Biological  
598 Sciences Research Council (BBSRC) ISP grant “Achieving Sustainable Agricultural Systems”  
599 (ASSIST) grant number BBS/E/C/000I0130 ASSIST, Met data for Rothamsted and Brooms  
600 Barn was supplied by the e-RA database, part of the Rothamsted Long-term Experiments  
601 National Capability (LTE-NC), supported by the UK BBSRC (BBS/E/C/000J0300) and Lawes  
602 Agriculture Trust. Tim Barraclough, Peter Fruen, Christoph Demeester and Guillaume  
603 Brancourt provided technical assistance on the field experiments.

604

605

## 606 **References**

- 607 Abadi FR, Tastra IK, Koentjoro BS. Preliminary Study of WOFOST Crop Simulation in Its  
608 Prospect for Soybean (*Glycine max* L.) Optimum Harvest Time and Yield Gap  
609 Analysis in East Java. *AGRIVITA, Journal of Agricultural Science*; Vol 40, No 3  
610 (2018): OCTOBERDO - 10.17503/agrivita.v40i3.1832 2018.
- 611 Addiscott TM, Whitmore AP. Simulation of solute leaching in soils of differing permeabilities.  
612 *Soil Use and Management* 1991; 7: 94-102.
- 613 Allen RG, Pereira LS, Raes D, Smith M. Crop evapotranspiration - Guidelines for computing  
614 crop water requirements - FAO Irrigation and drainage paper 56, FAO - Food and  
615 Agriculture Organization of the United Nations, Rome, 1998.
- 616 Avery BW, Catt JA. The soil at Rothamsted. Harpenden, UK: Lawes Agricultural Trust, 1991.
- 617 Bender RR, Haegele JW, Below FE. Nutrient Uptake, Partitioning, and Remobilization in  
618 Modern Soybean Varieties. *Agronomy Journal* 2015; 107: 563-573.
- 619 Boons-Prins ER, de Koning GHJ, van Diepen CA, Penning de Vries FWT. Crop-specific  
620 simulation parameters for yield forecasting across the European Community.  
621 *Simulation Reports CARO-TT. 32. cabo-dlo*, Wageningen, The Netherlands, 1993.
- 622 Bouniols A, Cabelguenne M, Jones CA, Chalamet A, Charpentreau JL, Marty JR. Simulation  
623 of soybean nitrogen nutrition for a silty clay soil in Southern France. *Field Crops  
624 Research* 1991; 26: 19-34.
- 625 Cheng A, Raai MN, Zain NAM, Massawe F, Singh A, Wan-Mohtar WAAQI. In search of  
626 alternative proteins: unlocking the potential of underutilized tropical legumes. *Food  
627 Security* 2019; 11: 1205-1215.
- 628 Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the  
629 binomial. *Biometrika* 1934; 26: 404-413.
- 630 Coleman K, Jenkinson DS. RothC - A Model for the turnover of carbon in soil: Model  
631 description and users guide (updated June 2014). Harpenden, UK: Lawes  
632 Agricultural Trust, 2014.

633 Coleman K, Muhammed SE, Milne AE, Todman LC, Dailey AG, Glendining MJ, et al. The  
634 landscape model: A model for exploring trade-offs between agricultural production  
635 and the environment. *Science of The Total Environment* 2017; 609: 1483-1499.

636 Corrêa STR. Adaptação do modelo LINTUL (Light Interception and Utilization) para  
637 estimação da produtividade potencial da cultura de soja. Escola Superior de  
638 Agricultura "Luiz de Queiroz". Mestrado em Fitotecnia. Universidade de São Paulo,  
639 2008.

640 Curtis CJ, Emmett BA, Reynolds B, Shilland J. How important is N<sub>2</sub>O production in removing  
641 atmospherically deposited nitrogen from UK moorland catchments? *Soil Biology &  
642 Biochemistry* 2006; 38: 2081-2091.

643 efeca. UK Roundtable on Sustainable Soya: Baseline study 2018, 2018.

644 Engering A, Hogerwerf L, Slingenbergh J. Pathogen–host–environment interplay and  
645 disease emergence. *Emerging Microbes & Infections* 2013; 2: 1-7.

646 European Commission. EU-U.S. Joint Statement: the United States is Europe's main soya  
647 beans supplier with imports up by 121%. IP/19/2154, Brussels, 2019.

648 FarmingUK. Boosted research to look at ways of growing soybean in UK. FarmingUK, sc,  
649 2018.

650 Fisher JB, Sitch S, Malhi Y, Fisher RA, Huntingford C, Tan SY. Carbon cost of plant nitrogen  
651 acquisition: A mechanistic, globally applicable model of plant nitrogen uptake,  
652 retranslocation, and fixation. *Global Biogeochemical Cycles* 2010; 24.

653 Harkness C, Semenov MA, Areal F, Senapati N, Trnka M, Balek J, et al. Adverse weather  
654 conditions for UK wheat production under climate change. *Agricultural and Forest  
655 Meteorology* 2020; 282.

656 Hoogenboom G, Jones JW, Boote KJ. Modeling growth, development, and yield of grain  
657 legumes using SOYGRO, PNUTGRO, and BEANGRO - a review. *Transactions of  
658 the Asae* 1992; 35: 2043-2056.

659 IDH and IUCN NL. European Soy Monitor. Researched by B. Kuepper and M. Riemersma of  
660 Profundo. Coordinated by N. Sleurink of IDH, The Sustainable Trade Initiative and H.  
661 van den Hombergh of IUCN National Committee of the Netherlands., 2019.

662 Jago G, Pattey E, Bourgeois G, Morrison MJ, Drury CF, Tremblay N, et al. Calibration and  
663 performance evaluation of soybean and spring wheat cultivars using the STICS crop  
664 model in Eastern Canada. *Field Crops Research* 2010; 117: 183-196.

665 Jones JW, Hoogenboom G, Porter CH, Boote KJ, Batchelor WD, Hunt LA, et al. The DSSAT  
666 cropping system model. *European Journal of Agronomy* 2003; 18: 235-265.

667 Legg JP. Emergence, spread and strategies for controlling the pandemic of cassava mosaic  
668 virus disease in east and central Africa. *Crop Protection* 1999; 18: 627-637.

669 Liu YY, Wu LH, Baddeley JA, Watson CA. Models of biological nitrogen fixation of legumes.  
670 A review. *Agronomy for Sustainable Development* 2011; 31: 155-172.

671 Minchin FR, Witty JF. Respiratory/Carbon Costs of Symbiotic Nitrogen Fixation in Legumes.  
672 In: Lambers H, Ribas-Carbo M, editors. *Plant Respiration: From Cell to Ecosystem*.  
673 Springer Netherlands, Dordrecht, 2005, pp. 195-205.

674 Monteith JL. Conservative behaviour in the response of crops to water and light. In:  
675 Rabbinge R, Goudriaan J, Van Keulen H, Penning de Vries FWT, Van Laar HH,  
676 editors. *Theoretical Production Ecology: Reflections and Prospects*. Pudoc,  
677 Wageningen, The Netherlands, 1990, pp. 3-16.

678 Monteith JL, Moss CJ. Climate and the efficiency of crop Production in Britain [and  
679 Discussion]. *Philosophical Transactions of the Royal Society of London B: Biological  
680 Sciences* 1977; 281: 277-294.

681 Nix. *The John Nix Pocketbook for farm management*, 2020.

682 Parton WJ, Schime DS, Ojima DS, Cole CV. A general model for soil organic matter  
683 dynamics: sensitivity to litter chemistry, texture and management. In: Bryant RB,  
684 Arnold RW, editors. *Quantitative Modeling of Soil Forming Processes*. SSSA Special  
685 Publication 39, Madison, WI, 1994, pp. 147-167.

686 Reijnders L, Soret S. Quantification of the environmental impact of different dietary protein  
687 choices. *American Journal of Clinical Nutrition* 2003; 78: 664S-668S.



688 Román S, Sánchez-Siles LM, Siegrist M. The importance of food naturalness for consumers:  
689 Results of a systematic review. *Trends in Food Science & Technology* 2017; 67: 44-  
690 57.

691 RStudio Team. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL  
692 <http://www.rstudio.com/>. 2020.

693 Sabate J, Soret S. Sustainability of plant-based diets: back to the future. *American Journal of*  
694 *Clinical Nutrition* 2014; 100: 476S-482S.

695 Santachiara G, Salvagiotti F, Gerde JA, Rotundo JL. Does biological nitrogen fixation modify  
696 soybean nitrogen dilution curves? *Field Crops Research* 2018; 223: 171-178.

697 Semenov MA. Development of high-resolution UKCIP02-based climate change scenarios in  
698 the UK. *Agricultural and Forest Meteorology* 2007; 144: 127-138.

699 Semenov MA. Impacts of climate change on wheat in England and Wales. *Journal of the*  
700 *Royal Society Interface* 2009; 6: 343-350.

701 Semenov MA, Donatelli M, Stratonovitch P, Chatzidaki E, Baruth B. ELPIS: a dataset of  
702 local-scale daily climate scenarios for Europe. *Climate Research* 2010; 44: 3-15.

703 Semenov MA, Stratonovitch P. Use of multi-model ensembles from global climate models for  
704 assessment of climate change impacts. *Climate Research* 2010; 41: 1-14.

705 Semenov MA, Stratonovitch P. Adapting wheat ideotypes for climate change: accounting for  
706 uncertainties in CMIP5 climate projections. *Climate Research* 2015; 65: 123-139.

707 Setiyono TD, Cassman KG, Specht JE, Dobermann A, Weiss A, Yang H, et al. Simulation of  
708 soybean growth and yield in near-optimal growth conditions. *Field Crops Research*  
709 2010; 119: 161-174.

710 Setiyono TD, Weiss A, Specht J, Bastidas AM, Cassman KG, Dobermann A. Understanding  
711 and modeling the effect of temperature and daylength on soybean phenology under  
712 high-yield conditions. *Field Crops Research* 2007; 100: 257-271.

713 Setiyono TD, Weiss A, Specht JE, Cassman KG, Dobermann A. Leaf area index simulation  
714 in soybean grown under near-optimal conditions. *Field Crops Research* 2008; 108:  
715 82-92.

716 Shibu ME, Leffelaar PA, van Keulen H, Aggarwal PK. LINTUL3, a simulation model for  
717 nitrogen-limited situations: Application to rice. *European Journal of Agronomy* 2010;  
718 32: 255-271.

719 Sinclair TR, Farias JR, Neumaier N, Nepomuceno AL. Modeling nitrogen accumulation and  
720 use by soybean. *Field Crops Research* 2003; 81: 149-158.

721 Soil Survey Staff. Soil taxonomy: A basic system of soil classification for making and  
722 interpreting soil surveys. Vol 436. Washington DC, 1999.

723 Song W, Sun S, Ibrahim SE, Xu Z, Wu H, Hu X, et al. Standard Cultivar Selection and Digital  
724 Quantification for Precise Classification of Maturity Groups in Soybean. *Crop Science*  
725 2019; 59: 1997-2006.

726 Soya UK. Soya is now a commercial reality in the UK, 2018.

727 Springmann M, Godfray HCJ, Rayner M, Scarborough P. Analysis and valuation of the  
728 health and climate change cobenefits of dietary change. *Proceedings of the National*  
729 *Academy of Sciences of the United States of America* 2016; 113: 4146-4151.

730 Tamagno S, Sadras VO, Haegerle JW, Armstrong PR, Ciampitti IA. Interplay between  
731 nitrogen fertilizer and biological nitrogen fixation in soybean: implications on seed  
732 yield and biomass allocation. *Scientific Reports* 2018; 8.

733 Taylor KE, Stouffer RJ, Meehl GA. An overview of CMIP5 and the experiment design.  
734 *Bulletin of the American Meteorological Society* 2012; 93: 485-498.

735 Terzić D, Popović V, Tatić M, Vasileva V, Đekić V, Ugrenović V, et al. Soybean area, yield  
736 and production in world. XXII ECO-CONFERENCE® 2018 Ecological movement of  
737 Novi Sad, 2018.

738 Tessari P, Lante A, Mosca G. Essential amino acids: master regulators of nutrition and  
739 environmental footprint? *Scientific Reports* 2016; 6.

740 Trnka M, Hlavinka P, Semenov MA. Adaptation options for wheat in Europe will be limited by  
741 increased adverse weather events under climate change. *Journal of The Royal*  
742 *Society Interface* 2015; 12: 20150721.

743 Trnka M, Rötter RP, Ruiz-Ramos M, Kersebaum KC, Olesen JE, Žalud Z, et al. Adverse  
744 weather conditions for European wheat production will become more frequent with  
745 climate change. *Nature Climate Change* 2014; 4: 637-643.

746 Tuorila H, Hartmann C. Consumer responses to novel and unfamiliar foods. *Current Opinion*  
747 *in Food Science* 2020; 33: 1-8.

748 Van Ittersum MK, Leffelaar PA, Van Keulen H, Kropff MJ, Bastiaans L, Goudriaan J.  
749 Developments in modelling crop growth, cropping systems and production systems in  
750 the Wageningen school. *NJAS - Wageningen Journal of Life Sciences* 2003; 50: 239-  
751 247.

752 van Laar H, Goudriaan J, van Keulen H. SUCROS97: Simulation of crop growth for potential  
753 and water-limited production situations, as applied to spring wheat. *Quantitative*  
754 *Approaches in Systems Analysis No.*, September 1997, ab-dlo, 1997, pp. 75.

755 van Vuuren DP, Edmonds J, Kainuma M, Riahi K, Thomson A, Hibbard K, et al. The  
756 representative concentration pathways: an overview. *Climatic Change* 2011; 109: 5.

757 Vance CP, Heichel GH. Carbon in N2 fixation - limitation or exquisite adaptation. *Annual*  
758 *Review of Plant Physiology and Plant Molecular Biology* 1991; 42: 373-392.

759 Vanuytrecht E, Raes D, Willems P, Semenov MA. Comparing climate change impacts on  
760 cereals based on CMIP3 and EU-ENSEMBLES climate scenarios. *Agricultural and*  
761 *Forest Meteorology* 2014; 195-196: 12-23.

762 VSN International. *Genstat for Windows 20th Edition*. VSN International, Hemel Hempstead,  
763 UK. [Genstat.co.uk](http://Genstat.co.uk), 2019.

764 Weir AH, Bragg PL, Porter JR, Rayner JH. A winter-wheat crop simulation-model without  
765 water or nutrient limitations. *Journal of Agricultural Science* 1984; 102: 371-382.

766 Willett W, Rockstrom J, Loken B, Springmann M, Lang T, Vermeulen S, et al. Food in the  
767 Anthropocene: the EAT-Lancet Commission on healthy diets from sustainable food  
768 systems. *Lancet* 2019; 393: 447-492.

769 Williams JR, Jones CA, Kiniry JR, Spalton DA. The EPIC crop growth model. *Transactions of*  
770 *the Asae* 1989; 32: 497-511.

771 Wolf J. User guide for LINTUL5: Simple generic model for simulation of crop growth under  
772 potential, water limited and nitrogen, phosphorus and potassium limited conditions,  
773 Wageningen UR 2012.

774

775