# scientific reports

OPEN

# European soybean to benefit people and the environment

Jose L. Rotundo[1,4✉], Rachel Marshall[2], Ryan McCormick[3], Sandra K. Truong[4], David Styles[5], Jose A. Gerde[6], Emmanuel Gonzalez-Escobar[2], Elizabete Carmo-Silva[2], Victoria Janes-Bassett[7], Jennifer Logue[8], Paolo Annicchiarico[9], Chris de Visser[10], Alice Dind[11], Ian C. Dodd[2], Louise Dye[12], Stephen P. Long[13,2], Marta S. Lopes[14], Joke Pannecoucque[15], Moritz Reckling[16,17], Jonathan Rushton[18], Nathaniel Schmid[11], Ian Shield[19], Marco Signor[20], Carlos D. Messina[21] & Mariana C. Rufino[2,22]

Europe imports large amounts of soybean that are predominantly used for livestock feed, mainly sourced from Brazil, USA and Argentina. In addition, the demand for GM-free soybean for human consumption is project to increase. Soybean has higher protein quality and digestibility than other legumes, along with high concentrations of isoflavones, phytosterols and minerals that enhance the nutritional value as a human food ingredient. Here, we examine the potential to increase soybean production across Europe for livestock feed and direct human consumption, and review possible effects on the environment and human health. Simulations and field data indicate rainfed soybean yields of $3.1 \pm 1.2$ t ha$^{-1}$ from southern UK through to southern Europe (compared to a 3.5 t ha$^{-1}$ average from North America). Drought-prone southern regions and cooler northern regions require breeding to incorporate stress-tolerance traits. Literature synthesized in this work evidenced soybean properties important to human nutrition, health, and traits related to food processing compared to alternative protein sources. While acknowledging the uncertainties inherent in any modelling exercise, our findings suggest that further integrating soybean into European agriculture could reduce GHG emissions by 37–291 Mt CO$_{2e}$ year$^{-1}$ and fertiliser N use by 0.6–1.2 Mt year$^{-1}$, concurrently improving human health and nutrition.

Soybean is the largest source of vegetable protein for animal worldwide and it is an important component of human diets in some areas of the world. The EU-27 + UK imports 33 M tonnes of soybean products annually at a cost of €14bn[1]. Soybeans have higher protein, poly- and mono-unsaturated fat, and lower carbohydrate content than other legumes[2]. Clinical studies indicate that soybean consumption benefits human health by lowering low-density-lipoprotein (LDL)-cholesterol concentrations[3] and blood pressure compared to other legume grains[4]. High soybean diets are associated with lower the incidence of heart attack and stroke, with variation across study populations[5], although protein replacement resulting in lower meat consumption might also play a role. Whilst other grain legumes may provide these benefits and are important for diversification of the system, soybean has potential to be the most productive, requiring fewer hectares per unit of product[6]. Global average dry bean yields (1980–2020) rose from 0.5 to 0.8 t ha$^{-1}$ (60%), dry peas 1.3 to 2.0 t ha$^{-1}$ (53%) and soybean 1.6 to 2.8 t ha$^{-1}$ (74%)[7]. Historically, soybean yield increases over the years have been associated with both genetic progress

[1]Corteva Agriscience, Seville, Spain. [2]Lancaster Environment Centre, Lancaster University, Lancaster, UK. [3]Gro Intelligence, New York City, NY, USA. [4]Corteva Agriscience, Johnston, USA. [5]School of Biological and Chemical Sciences, University of Galway, Galway, Ireland. [6]Instituto de Ciencias Agrarias de Rosario, UNR, CONICET, Zavalla, Argentina. [7]School of Environmental Sciences, University of Liverpool, Liverpool, UK. [8]Lancaster Medical School, Lancaster University, Lancaster, UK. [9]Council for Agricultural Research and Economics (CREA), Rome, Italy. [10]Wageningen University and Research, Wageningen, The Netherlands. [11]Research Institute of Organic Agriculture (FiBL), Frick, Switzerland. [12]School of Psychology and Food Science and Nutrition, University of Leeds, Leeds, UK. [13]Departments of Crop Sciences and of Plant Biology, University of Illinois, Champaign, USA. [14]Sustainable Field Crops, Institute of Agrifood Research and Technology (IRTA), Lleida, Spain. [15]Flanders Research Institute for Agriculture, Fisheries and Food (ILVO), Merelbeke, Belgium. [16]Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany. [17]Department of Crop Production Ecology, Swedish University of Agricultural Sciences (SLU), Uppsala, Sweden. [18]Centre of Excellence for Sustainable Food Systems, University of Liverpool, Liverpool, UK. [19]Rothamsted Research, Harpenden, UK. [20]Regional Agency for Rural Development (ERSA), Gorizia, Italy. [21]University of Florida, Gainesville, USA. [22]School of Life Sciences, Technical University of Munich, München, Germany. ✉email: jose.rotundo@corteva.com

and improved crop husbandry[8]. The increase in soybean yield associated with genetic progress is not related to incorporating genetically modified (GM) defensive traits, since all the genetic gain studies are conducted in the absence of pests and weeds[9]. Soybean was one of the first crop species to be fully sequenced, enabling adaptive breeding to different environments and to accelerate crop improvement[10–12]. Although soybean is higher yielding with a faster pace of yield improvement than other grain legumes, its cultivation has been associated with a large carbon footprint due to deforestation and negative biodiversity impacts[13]. The current soybean value chain configuration of the European food system and agricultural policies inhibit harnessing the benefits of greater human soybean consumption[14].

European soybean yields are low in comparison to this crop's potential[14,15]. Suboptimal air temperatures for growth and development likely limit soybean production in northern Europe by delaying early vegetative growth, causing flower and pod abscission, poor grain filling, and high moisture for harvest[16]. Low soil temperatures limit root hydraulic conductance and decrease leaf turgor and expansion whilst high irradiance can cause photoinhibition[17]. Soil water deficits constrain soybean yields in central and southern Europe[18]. Leaf expansion and photosynthesis vary among soybean genotypes, suggesting genotypes better suited to the diversity of European environments could be selected[19]. Many areas are well suited to soybean cultivation, giving immediate potential for expansion. A comprehensive European soybean trial network has the potential to supply essential data for effective adaptation to diverse environments and the identification of high-yielding varieties suitable for different regions.

European farmers favour cereal production over legumes partly because of their higher yields, profitability[20], yield stability[15], well-developed markets, and ineffective incentives for legume production[20,21]. Nitrogen (N) fixation in European soybean cultivation would reduce fertiliser use[22], as soybean is typically cultivated without adding N fertiliser, benefitting farmers in regions that restrict N use to mitigate water pollution and avoiding impacts of high fertiliser costs[23]. Production within Europe would also address regulatory constraints on genetically modified (GM) foods and consumer preferences[24]; the EU aims to transition to more locally-produced protein-rich plant-based proteins[1]. This paper aims to critically assess the evidence and prospects for increased soybean production by identifying suitable regions within Europe for soy production whilst analysing opportunities for adaptation and the likely environmental benefits of alternative scenarios. We also review health and nutritional factors associated with consuming soy foods, identifying specific traits desirable to the food industry.

## Results

### Increasing soybean production in Europe

In year 2020, EU-27 + UK cultivated about 0.9 M hectares of soybeans to produce 2.7 M tonnes, approximately 8% of the total imports[7]; this area is projected to increase as demand for GM-free soybean for human consumption increases[1]. Nine countries account for ~ 85% of total EU-27 + UK imports (Fig. 1a). Except for Italy, these countries do not produce significant amounts of soybean, with eight of such countries (Netherlands, Germany, France, Poland, UK, Belgium, Denmark, and Ireland) located at the northern edge of current soybean cultivation. Soybean acreage, whilst still small relative to consumption, has significantly expanded in the last decade (Fig. 1b), notably in Italy, France, and Romania, currently at 12,000–21,000 ha year$^{-1}$. Russia and Ukraine, non-EU countries with a tradition of soybean cultivation, substantially increased acreage by 214,000 and 99,800 ha year$^{-1}$ to 2.7 and 1.36 Mha between 2010 and 2020, although the Russia-Ukraine war is expected to sharply decrease production and export, with 2022 official statistics pending. Comparing current yields with yield potential for 12 countries (with soy cultivation exceeding 20,000 ha) shows untapped opportunities to increase yields through improved agronomy (Fig. 1c), and improved cultivars. Nine of these 12 countries show a water-limited yield gap larger than 20%, and five larger than 40%. There are clear opportunities to close yield gaps in countries where no sustained yield gains have been observed in the last two decades (e.g. France, Italy and Germany).

Crop modelling shows a large area (clusters D, E, F, and G, Fig. 2) of 70 M hectares from continental western, through central to eastern Europe where economic yields of soy (> 3 t ha$^{-1}$) are feasible under rainfed conditions with a crop success rate > 85% and where better agronomic practices would close a yield gap of about 1 t ha$^{-1}$ (Table 1, Fig. 1c). These practices involve weed control, reducing pest and disease incidence, managing planting density and row spacing, and selecting cultivars adapted to local environments, tolerating cold stress, in-season dry spells, and humid conditions during dry-down. These simulations show that improved agronomy could allow maturity groups between I and 000 to reach their yield potential. South of this optimal region are the Mediterranean and Balkan regions (cluster I, 56 Mha), with a larger yield gap (> 2 t ha$^{-1}$) related to water deficits. Breeding for drought tolerance and adjusting maturity group could also benefit large areas of western Asia such as most of Türkiye (cluster K) and Kazakhstan (cluster J). Cluster H (Spain and Portugal) includes marginal agroecologies for soybean, where production currently requires irrigation. Cool temperatures and a shorter growing season in north-eastern Europe and southern UK (clusters B and C), limit potential yields to < 3 t ha$^{-1}$; overcoming cold-tolerance requires breeding investments to identify phenotypes adapted to low soil temperature at sowing time and ability to develop photosynthetically competent leaves in chilling spring temperatures. A temperature-sensitive polymer seed coat is successfully used in the US Midwest to prevent seed and seedling damage during cold events following early season sowing[25]. Early spring air temperatures of 8–10 °C cause photodamage[26] and slow crop development by limiting leaf expansion thereby compromising radiation capture, leading to crops failing to reach maturity or achieving low yields (< 1.5 t ha$^{-1}$) in northern UK, Ireland and Scandinavia (cluster A). Current soybean cultivars cannot establish in the coldest environments of mountain and sub-arctic regions (clusters L, M, N and O, Table 1).
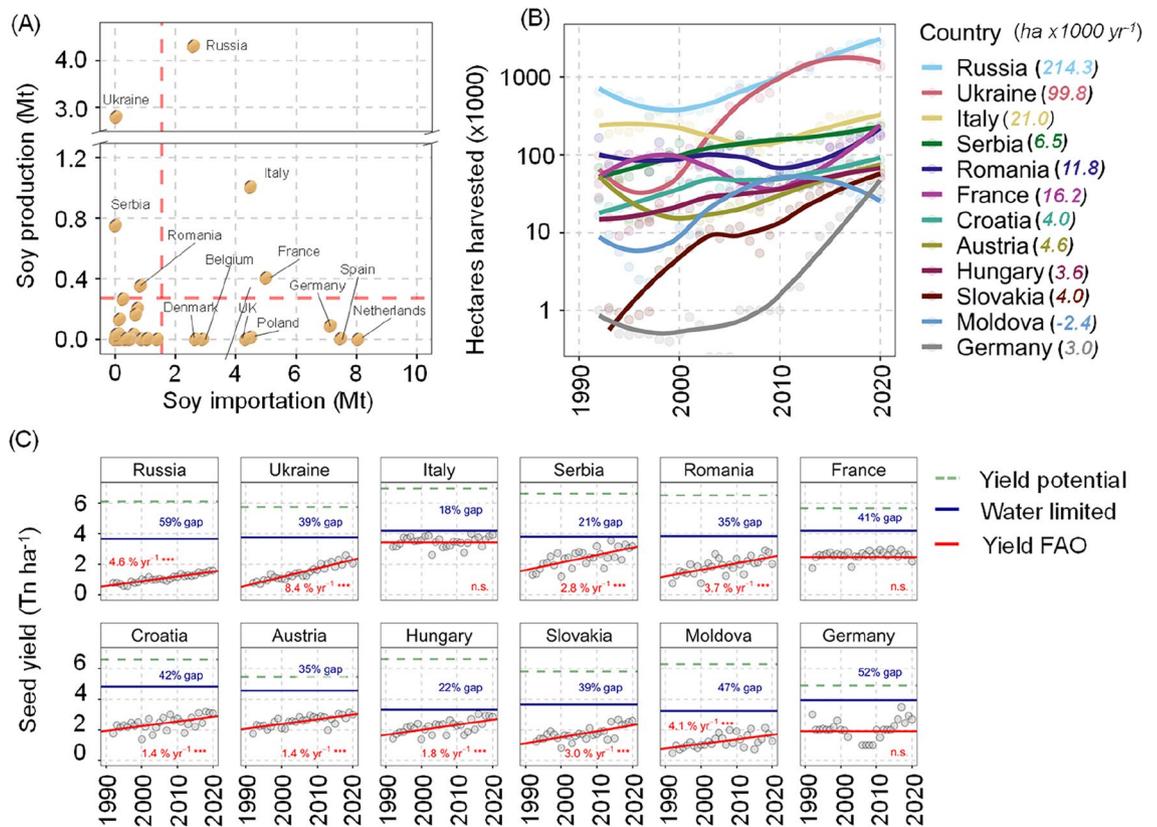
**Figure 1.** European soy production, imports, cultivated area and yield gaps. (**A**) Soybean seed production and import of soybean seed equivalents in millions of tons (Mt) for 2020 in 39 countries (EU-27 + UK plus 12 neighbouring European countries). Vertical and horizontal red dashed lines indicate mean imports and production. Countries above these averages are labelled. (**B**) Change in soy cultivated area (1992–2020) of 12 European countries that cultivate > 20,000 hectares since 2017. Numbers in parenthesis on the legend indicate the slope in thousand hectares per year for the period 2010–2020 (all linear models $p < 0.05$) (**C**) Change in soybean yield from 1992 to 2020 (grey dots). Red lines show the linear trend with the increase in yield as percentage per year (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). Blue lines indicate simulated water-limited yield (mean of 20-year weather) and percent yield gap calculated as: (water-limited yield − observed yield in 2020)/ water-limited yield. Green dashed lines indicate irrigated potential yield. All data except simulations are from FAOSTAT.

## Environmental sustainability

Alternative scenarios representing agronomic potential for European soybean expansion show significant prospective environmental benefits. First, taking hypothetical scenarios that avoid displacing existing crop cultivation, planting half of set-aside area[27] in appropriate agroclimatic regions, soybean would produce 9–13 Mt soy annually at current (Scenario 1, Sc-1) or potential (Sc-2) yields (Table 2). This would replace up to 40% of current EU imports, avoiding 13–19 Mt $CO_2$ equivalent ($CO_2$e) from soybean production and transport from Brazil, plus an avoided carbon opportunity cost (COC) of 24–35 Mt $CO_2$e by sparing 1.8–2.7 Mha of soy cultivation in Brazil. Alternatively, diversifying crop rotations by replacing 10% of current EU wheat cultivation (in areas where soy could yield ≥ 3 t ha$^{-1}$) with soybean cultivation, could produce 11–23 Mt soybean at current (Sc-3) or potential (Sc-4) yields, avoiding 16–31 Mt $CO_2$e annually from imported soybean. However, based on average European wheat yield based on Agribylase LCA data for compensatory wheat production, this would displace 26–34 Mt wheat production, requiring a net increase in global cropland area of ca 1 Mha in Sc-3 as weighted-average annual soybean yields in Europe (3.2 t ha$^{-1}$) are lower than Brazil's (ca 5 t ha$^{-1}$). A COC of 13 Mt $CO_2$e would negate over 80% of the production and transport GHG saving (Table 2). Linking soy production with diet shifts, scenarios 5 and 6 indicate large area (6.5–15 Mha), fertiliser-N (0.6–1.2 Mt N year$^{-1}$), and production-related GHG (47–93 Mt $CO_2$e year$^{-1}$) savings by substituting chicken, pork, and milk with half of the soybean produced on 10% of land currently used to cultivate wheat, and where high soy yields (≥ 3 t ha$^{-1}$) can be achieved. On a protein basis, soy produced in these scenarios could substitute 30–58% of chicken consumption and 8–16% of pork consumption, alongside substitution of 8–16% of milk consumption on a volume basis across the EU[28]. Including COC, these scenarios 5 and 6 could mitigate up to 291 Mt $CO_2$e year$^{-1}$. The area planted with soybean (3.6–4.8 Mha) would be only slightly larger than projected plantings with soybean and pulses (3.4 Mha) over the coming years[27].
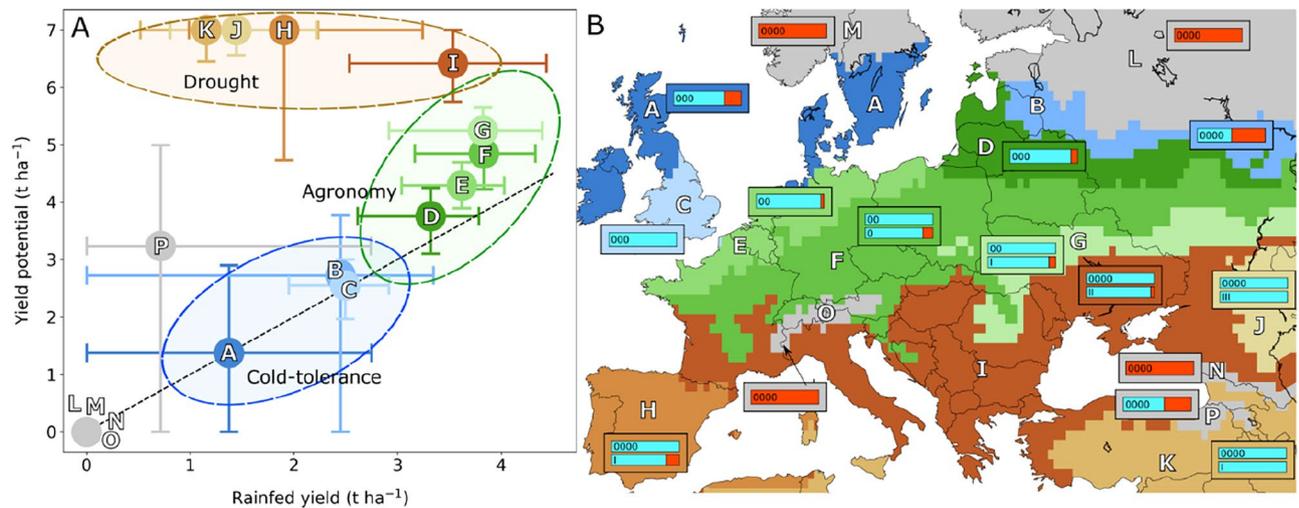
**Figure 2.** Soybean crop performance across Europe and its potential growing regions (16 clusters with similar outcomes for maturity groups 0000 to III). Clusters L to P were shaded dark grey because soy cannot complete its growth cycle in these regions. (**A**) Median and interquartile ranges of potential (y-axis) and rainfed yields (x-axis) for the maturity group with the largest yield within each region. Interquartile ranges for yields originate from temporal and spatial variation in weather, site-to-site variation in soil properties, and uncertainty associated with satellite-derived soil water initial conditions. Regions were assigned to crop improvement strategies: breeding for cold-tolerance in cool climates (blue shade), agronomic optimisation (green shade), and improving drought tolerance (brown shade); dashed line is the 1:1 regression line. (**B**) Geographical extent of each cluster; a bar plot inset is labelled with maturity group with the largest median rainfed and/or potential yield, and the bar is coloured cyan according to the proportion of years the maturity group successfully reached R7. For insets with two bars, the upper bar represents the maturity group with the largest median rainfed yield, and the lower bar represents the maturity group with the largest median yield potential. Insets with one bar indicate the same maturity group had the largest median potential and rainfed yields. The proportion of years failing to achieve R7 is coloured red, and the proportion of years with crop harvest success are coloured in light blue. The map was created using Python 3.10.0 (https://www.python.org/downloads/).

## Soybean for nutrition

The large soybean import demand of the EU would be mitigated if more soybean was used for direct human consumption rather than to feed livestock. Here we review reasons and opportunities for such a shift. Studies indicate that soy-based foods provide simultaneously protein, essential fatty acids, carbohydrates, dietary fibre, minerals and vitamins and beneficial phytochemicals such as isoflavones and phytosterols (Supplemental Table 1). Most soybean imported into Europe is currently used as livestock feed, with soybean breeding aiming to increase yield instead of enhancing soy functional properties[29]. Processing improves quality of feed- and food-grade soybean by reducing the concentrations of anti-nutritional compounds for humans and livestock, such as trypsin inhibitors, allergens, saponins, lectins, phytate, and the raffinose family oligosaccharides (RFO)[30], justifying common breeding goals with specific targets for fresh edible and mature grain products (Fig. 3).

Soybeans are rich in essential amino acids, especially lysine. When considering protein quality, soybeans score almost as high as milk proteins and well above proteins from other legume or cereal sources[31]. The limiting amino acids to fully meet human and monogastric requirements are the sulphur-containing ones: methionine and cysteine[30]. There is genetic diversity in cysteine and methionine in storage proteins, glycinin (11S) and β-conglycinin (7S), with 11S having up to 4 × higher concentration of these amino acids than the 7S protein[32]. Increasing the 11S/7S ratio improves soy quality, with enhanced texture, chewiness, springiness, and hardness of tofu[33]. Modifying soybean amino acid profile has seen limited success, as high-throughput germplasm screening has not been cost-effective and bioengineering is unlikely to be accepted in the EU[34]. Genotype and growing regions affect seed protein concentration, and high protein concentration is usually associated with lower yields[35], requiring simultaneous screening of traits. However, a wide range of protein, oil and mineral composition exists within soybean germplasm collections, providing clear breeding opportunities to deliver cultivars to meet nutritional requirements[36].

Compared to other legumes and cereals, soybeans have more diverse nutritional properties and higher concentrations of minerals (potassium, calcium, magnesium, iron and zinc) and B vitamins (Supplementary Table 1). Riboflavin, vitamin A and vitamin E, whilst not in high concentrations, enhance the nutritional profile of soy foods[37]. Certain genotypes and environments can enrich key minerals for human nutrition such as selenium, copper, iron, manganese and zinc in soy[38]. For some minerals, most notably calcium and iron, soy has higher bioavailability than other legumes[39]. Soy's high iron concentrations can help raise iron levels in deficient populations; the ferritin form of iron found in soy has greater bioavailability than the phytate-complexed form found in cereal grains[40]. Increased soybean iron concentration in varieties with higher nodulation[41], could help improve iron status in humans[42].

| Cluster | Region name | Climate | Current land use % of arable (top 3 crops by area) | Arable land (ha) | Land area as arable (%) | Yield potential (t ha⁻¹) | Rainfed yield (t ha⁻¹) | Crop success rainfed (%) | Maturity group potential yield | Maturity group rainfed yield |
|---|---|---|---|---|---|---|---|---|---|---|
| A | Northern Europe | Temperate maritime | Wheat (37%), barley (31%), rapeseed (10%) | 6,124,324 | 18.1 | $1.6 \pm 1.5$ | $1.5 \pm 1.3$ | 75 | 000 | 000 |
| B | North-eastern Europe | Continental | Wheat (45%), barley (13%), sunflower (8%) | 2,247,233 | 9.1 | $2.0 \pm 1.9$ | $1.8 \pm 1.6$ | 50 | 0000 | 0000 |
| C | Southern UK | Maritime | Wheat (41%), barley (26%), rapeseed (12%) | 4,646,324 | 34.7 | $2.5 \pm 0.8$ | $2.4 \pm 0.7$ | 100 | 000 | 000 |
| D | North-eastern Europe | Humid continental | Wheat (38%), barley (11%), rapeseed (11%) | 6,976,458 | 18.4 | $3.3 \pm 1.5$ | $2.9 \pm 1.4$ | 90 | 000 | 000 |
| E | Western Europe | Temperate maritime | Wheat (35%), barley (15%), rapeseed (8%) | 13,225,677 | 39.4 | $4.1 \pm 1.2$ | $3.4 \pm 1.1$ | 95 | 00 | 00 |
| F | West to eastern Europe | Continental to intermediate | Wheat (31%), barley (13%), maize (10%) | 37,826,068 | 28.5 | $4.1 \pm 1.4$ | $3.6 \pm 1.1$ | 85 | 0 | 00 |
| G | East Europe | Humid continental | Wheat (27%), maize (20%), sunflower (20%) | 13,063,593 | 31.4 | $4.6 \pm 1.9$ | $3.6 \pm 1.1$ | 90 | I | 00 |
| H | Southern Europe | Mediterranean hot summer | Olives (22%), barley (20%), wheat (15%) | 12,487,126 | 18.6 | $5.1 \pm 2.9$ | $2.2 \pm 1.5$ | 80 | I | 0000 |
| I | Southern Europe | Mediterranean humid continental | Wheat (29%), maize (18%), sunflower (13%) | 55,649,190 | 25.8 | $5.8 \pm 1.9$ | $3.4 \pm 1.3$ | 95 | II | 0000 |
| J | Central Asia | Cold semi-arid | Wheat (48%), barley (14%), sunflower (11%) | 1,754,289 | 7.8 | $6.7 \pm 0.6$ | $1.6 \pm 0.9$ | 100 | III | 0000 |
| K | West Asia | Dry summer subtropical | Wheat (37%), barley (17%), olives (7%) | 18,165,133 | 19.6 | $6.2 \pm 1.9$ | $1.5 \pm 1.3$ | 100 | I | 0000 |
| *L* | *North-eastern Europe* | *Sub-arctic* | *Wheat (41%), barley (16%), sunflower (11%)* | *5,412,229* | *8.0* | *$0.2 \pm 0.8$* | *$0.2 \pm 0.7$* | *0* | *na* | *na* |
| *M* | *Northern Europe* | *Maritime sub-arctic* | *Wheat (35%), barley (31%), oats (14%)* | *729,597* | *4.2* | *$0.2 \pm 0.7$* | *$0.2 \pm 0.7$* | *0* | *na* | *na* |
| *N* | *Southwest Russia* | *Hot summer continental* | *Wheat (33%), barley (12%), maize (10%)* | *480,846* | *6.6* | *$0.8 \pm 1.8$* | *$0.7 \pm 1.5$* | *0* | *na* | *na* |
| *O* | *Alps* | *Alpine* | *Wheat (31%), maize (12%), barley (10%)* | *899,790* | *9.3* | *$0.3 \pm 1.0$* | *$0.3 \pm 1.0$* | *0* | *na* | *na* |
| *P* | *Asia Southeast Black Sea* | *Humid* | *Wheat (36%), barley (16%), olives (4%)* | *1,387,943* | *20.4* | *$2.6 \pm 2.4$* | *$1.4 \pm 1.5$* | *60* | *0000* | *0000* |

**Table 1.** Description of clusters and associated regions based on similar simulated soybean performance (potential and water-limited yields, and median crop success rate) for maturity groups from 0000 to III. Clusters L-P shaded grey represent the clusters where soy cannot complete its cycle under current climate. Significant values are in italics.

Anti-nutritional compounds associated with soy proteins such as trypsin inhibitors reduce protein digestibility in the human gut[43] and in monogastric livestock[44] or inhibit nutrient and vitamin uptake resulting in deficiencies[30]. Fermentation (for miso and tempeh), soaking (tofu) and cooking can reduce anti-nutritional effects for humans[30] and thermal processing reduces trypsin inhibitors activity in livestock feed although protein denaturation reduces digestibility[30]. Genetic variability in trypsin inhibitor, phytate and tannin content[45] could be exploited by using molecular methods to reduce concentrations of these compounds[43]. Thermal processing inactivates carbohydrate-binding proteins such as lectins[44], but heat-resistant RFOs such as stachyose and raffinose may remain[43]. RFOs are substrates for microbial fermentation in the human gut, causing bowel discomfort[46]. RFOs reduce feed digestibility for monogastric animals[47], although they ferment into fatty acids in pig intestines showing prebiotic properties[48]. Soy breeding aims to reduce RFOs concentrations whilst increasing total sugars[49] since high sucrose, fructose and glucose increase the sweetness of soymilk, tofu and natto[46,50,51].

| Scenario | European soybean | Imported soybean | Displaced wheat | Barley feed | Maize feed | Rapeseed feed | Sunflower feed | Cropland area | Fertiliser-N | Production GHG | Carbon opportunity cost |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mt year$^{-1}$ | | | | | | | Mha | Mt year$^{-1}$ | Mt CO$_2$e year$^{-1}$ | |
| Sc-1 | 9 | − 9 | 0 | | | | | − 1.8 | | − 13 | − 24 |
| Sc-2 | 13 | − 13 | 0 | | | | | − 2.7 | | − 19 | − 35 |
| Sc-3 | 11 | − 11 | 26 | | | | | 1.0 | | − 16 | 13 |
| Sc-4 | 23 | − 23 | 34 | | | | | − 0.1 | | − 31 | − 2 |
| Sc-5 | 11 | − 16 | 13 | − 0.7 | − 42 | − 0.8 | − 0.3 | − 6.5 | − 0.6 | − 47 | − 86 |
| Sc-6 | 23 | − 31 | 10 | − 1.3 | − 83 | − 1.5 | − 0.6 | − 15.0 | − 1.2 | − 93 | − 198 |

**Table 2.** New and avoided crop production, and associated area and fertiliser nitrogen (N) application changes, associated with six scenarios of soybean cultivation across agroclimatic regions and fractions of areas indicated in Table 1. Also shown are net production and carbon opportunity cost greenhouse gas emissions savings associated with each of the scenarios, including avoided animal production (and associated animal feed crop production). Sc-1 and Sc-2 represent soybean cultivation on 50% of EU set-aside at current or potential yields from Table 1, respectively. Sc-3 and Sc-4 represent replacement of 10% of wheat cultivation (by area) with soy cultivation in regions where soybean yields of at least 3 t ha$^{-1}$ could be achieved, at current or potential yields, respectively. Sc-5 and Sc-6 are derivatives of Sc-3 and Sc-4, assuming that half of the European soybean production displaces animal protein, equally split across chicken, pork and milk.
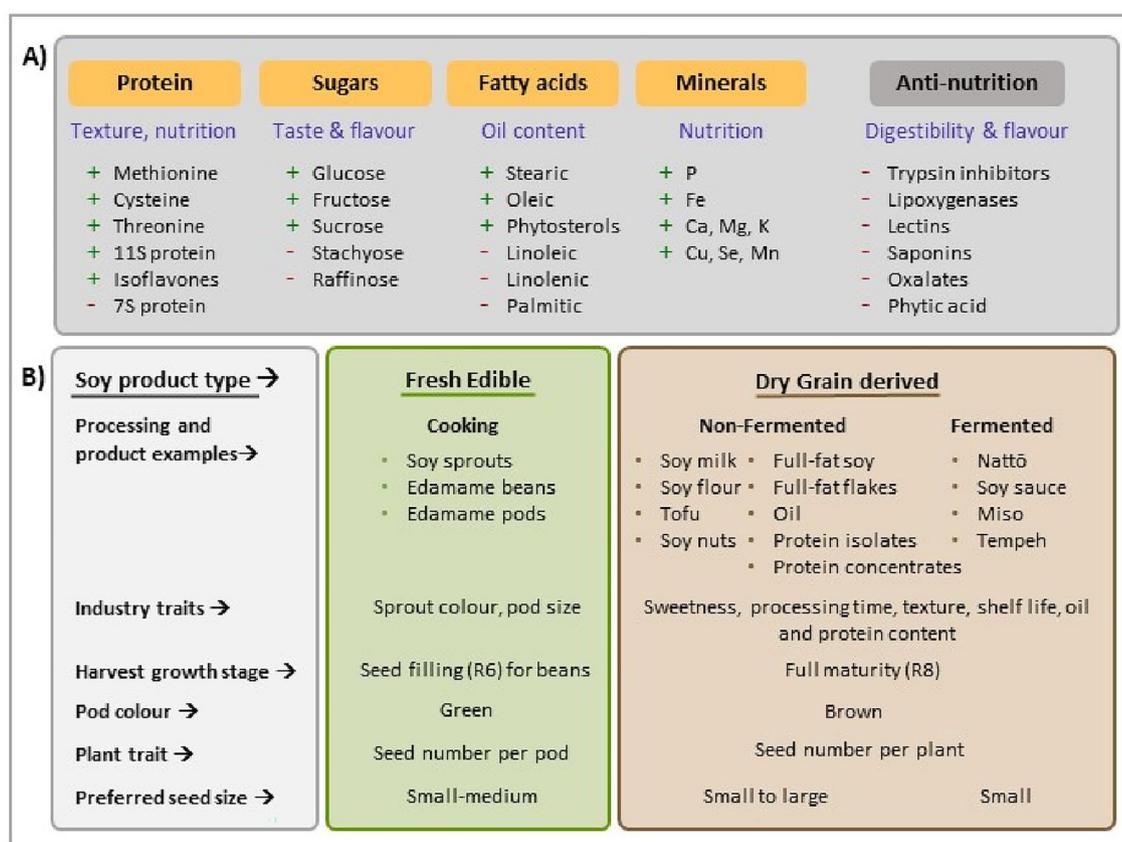


**Figure 3.** Soybean and soy-product traits, qualities, and breeding aims: (**A**) Crop traits for soy recommended to increase (green plus symbol) or to decrease (red minus symbol). The grey box indicates negative selection for anti-nutritional compounds and the orange boxes indicate positive selection of traits (protein, sugars, fatty acids and minerals). (**B**) Soy products from fresh edible and mature grain and target traits for crop improvement contributing to better food and feed quality.

Soybeans are classified as allergenic foods, however, the prevalence of this allergenicity among the population with food allergies is ~ 3%, while milk, peanuts, tree nuts, eggs, and shellfish account for ~ 75%[52]. At least 16 protein fractions in soybeans are considered allergens[53]. This makes it difficult to tackle this issue solely through plant breeding, therefore additional processing strategies are being explored to reduce soybean allergenicity through thermal, ultrasound, and cold plasma treatments, enzymatic hydrolysis, and chemical modifications[54].

## Soybean and health

Consuming soy foods can lower cholesterol concentrations by displacing high-saturated fat animal proteins in human diets[3]. A meta-analysis of 46 studies indicated soy protein consumption decreased cholesterol by 3–4% more than with other non-soybean protein sources[55,56]. The mechanisms by which soy reduces LDL-cholesterol are not well understood and may relate to the 7S globulin upregulation of LDL receptor[57,58], the effect of isoflavones[59], or phytosterols which can reduce cholesterol absorption[60]. Isoflavones are associated with 7S and mimic the effects of estrogen in the human body[61]. Their moderate consumption is associated with reduced risk of heart disease and hormone-related cancers, and reduction in adverse effects of hormone imbalances[62]. Soy has higher isoflavone concentrations than pulses, with sowing date, density[63] and cultivar[64] promising further enhancement. Seed phytosterol concentrations vary between 202 and 843 μg g$^{-1}$ and positively correlate with oil content[64], and phytosterols can be extracted for dietary supplements.

Soy oil has a relatively high ratio of linoleic to α-linolenic (ca. 7:1)[65] and efforts to change it are currently undergoing[66]. Since linoleic is inflammatory and α-linolenic is anti-inflammatory, they should be consumed at a ratio of 1:1–2:1 (linoleic:α-linolenic)[67]. Research showed that reducing the α-linolenic and saturated fats does not affect the concentrations of beneficial phytochemicals and antioxidants[68]. Recent discoveries on the function of key genes in fatty acids synthesis can be used to improve soy oil profile[69]. Concentrations of oleic acid higher than 70% and concentrations of α-linolenic lower than 3% improve oil stability and provide health benefits as food and feed[70,71]. Soybean breeders have developed cultivars with modified fatty acid compositions targeted to increase the oxidative stability and the improvement of other technological properties such as plasticity or melting temperature through the reduction of the concentration of polyunsaturated fatty acids (linoleic and α-linolenic)[66].

## Soybean comparison to other protein sources

Soybean has one of the highest concentrations of protein among grain legumes and cereals, making it a suitable feedstock to produce substitutes for animal-sourced foods such as tofu, soy milk, edamame, tempeh, and, increasingly, textured vegetable protein (meat analogue). The best quality soybean varieties produce about 40–50% protein, 20–25% total edible oil and 15% soluble carbohydrates[72]. Soybeans provide a higher ratio of protein to saturated fat than animal-based products and a ca. 4×higher protein to carbohydrate ratio than cereals (Supplementary Table 1). Although soy has lower sulphur amino acid (methionine and cysteine) content than cow milk, the protein quality of soy milk is higher than that of other plant-based milk substitutes such as almond, rice, coconut[73]. Soy is also the main source of protein for pigs[74], and of lysine to both pigs and poultry[20]; improving the amino acid profile of soy could benefit the livestock industry. Whilst plant proteins generally have lower digestibility than animal proteins[75], soy has the highest digestibility among legumes with 80% for soybean and >95% for soy protein isolate[76], and therefore it is a preferred ingredient for meat substitutes in human diets.

## Soybean traits for food use

Nutritional and industrial properties of food-grade soy vary with plant developmental stages (green *versus* mature beans), with processing, and different soy products. When producing soybean sprouts, the light used for germination can promote chlorophyll formation affecting nutritional quality and marketability[77]. Sprouting decreases the activity of hemagglutinin, lipoxygenases, and trypsin inhibitors that impair protein digestibility. Seed concentrations of bitter-tasting aglycones, which regulate the flavour of sprouts decrease in partially shaded maize-soybean intercropping[78]. At the green picking stage, edible pods (consumed as Edamame) are rich in vitamins B1 and B2, iron, calcium, phosphorus, and protein content, but low in lipoxygenases, which increase in concentration with pod maturity[79]. Water deficit during seed filling increases protein concentration but decreases seed size[80]. Time to pod maturity influences glucose content, seed colour, seed texture, seed size and quality[81], with essential amino acid concentrations increasing in mature beans[1]. Pod harvest time is critical for quality and nutritional properties of fresh edible soy.

Soaking soybeans before processing is common practice to soften the beans and reduce cooking time for human consumption. However, soaking can leach beneficial water-soluble antioxidants into water[82]. Cooking can decrease concentrations of other health-promoting compounds such as flavonoids and phenolics by 68–78% and 43–62% respectively[82]. Larger seeds reduce the negative effects of soaking, with 5.5–7.3 mm seeds preferred for natto production for their higher water absorption capacity and rapid fermentation, which retain flavour[50]. Therefore, improvement in soybean properties and size can help improve specific production efficiencies.

## Discussion

Europe has historically had a deficit in plant-based protein sources, relying on imports from the Americas to feed livestock[1]. European aspirations to increase protein production to shorten agri-food supply chains, reduce the demand for imported protein rich feedstocks, and to supply the burgeoning market of plant-based products have become more acute with the Russia-Ukraine war leading to urgent calls to develop a comprehensive European protein strategy[83]. In 2018, The European Commission suggested avenues to increase domestic soybean production, including (i) R&D investment to increase yields via extension, training, and dedicated breeding programmes, (ii) additional subsidies for protein-rich crops, and ecological focus areas (EFA) for nitrogen fixing crops, (iii) phasing-out import of oil palm to increase the demand for soy oil and (iv) raising awareness about the benefits of soybean based products to change consumer behaviour, and to increase consumption of plant-based products[84]. Economic modelling of the first three strategies concluded that the most effective policies would support an annual increase in yields of 1% at EU level for 2020–2030[84]. Several European countries have observed these yield gains (Fig. 1c), and Eastern European climate and soils (similar to those of the US Midwest) should allow high yields (>5 t ha$^{-1}$)[14].

Climate change is expected to increase the frequency of severe drought and heat waves in southern Europe, whilst northern areas may benefit from higher late season temperatures[85]. Drought is expected to constrain soybean production more than heat stress[86] yet expansion of production areas as a result of temperature and $CO_2$ effects may overcompensate yield reductions due to drought[87]. Four strategies are proposed here to grow soybean under future climate: (i) selecting for cold-tolerance of early growth and higher photosynthetic efficiency in maturity groups ranging from 0 to 0000 to adapt to the diversity of environments of a large area (clusters A, B, C, D and E) becoming increasingly warmer and some areas with sub-optimal radiation (clusters A and C); (ii) early-sowing of long-cycle cultivars (groups I to III) to escape end-of-season water deficit[88] (clusters I, J, and K); (iii) relay and double-cropping cereals and soy to better utilise land and the thermal-time window in central Europe (clusters E, F and G); (iv) selecting drought tolerant cultivars for south-European environments especially cluster I, and cluster H where double cropping can be practiced. Suitability of variety selection by location is therefore crucial for expansion of soybean[89].

To ensure that soy cultivars produce > 3 t ha$^{-1}$ with a failure rate < 10% under future warmer and drier climates, a systematic assessment of genotypes is required. Diversity panels could be screened to identify drought and cold tolerance in soybean cultivars[12]. Limited variation in drought tolerance is reported in modern cultivars probably because of breeding in favourable environments[90]. However, genes associated with drought tolerance have been identified[91] and genomic selection can effectively and efficiently select for drought tolerance by exploiting gene-marker linkages over several parts of the genome[92]. Projections indicate that Europe could achieve 50% self-sufficiency in soy production by further integrating soy into crop rotations[93]. Relay intercropping of cereals and legumes provides another option in specific conditions, with increased land use equivalent ratio probably derived from increased nutrient use efficiency[94]. This could help increase biodiversity and reduce nutrient surpluses[95] although the practice needs optimisation and specialised machinery[96].

Reducing animal feed demand (where soy production and consumption is linked with a diet shift away from meat) would reduce displacement of other crops in Europe, and spare significant areas of cropland overall, so that an additional 86–198 Mt $CO_2$e year$^{-1}$ could be sequestered via carbon farming activities (e.g. afforestation) on spared land[14,97]. Thus, European soybean cultivation could mitigate substantial quantities of GHG emissions when planted on set aside or ecological focus areas, or when cultivated within a food system that replaces crop-fed animal protein with plant protein. Combining animal protein substitution with carbon farming activities on spared land, could save 291 Mt $CO_2$e annually.

Increasing soybean foods in human diets can improve human health by reducing LDL-cholesterol concentrations, blood pressure, the risk of cardiovascular events and supplementing iron-deficient populations. Increasing soy production in Europe is feasible, bringing additional benefits by greater alternatives for crop rotation, stability and food system self-sufficiency, and more sustainable farming by reducing negative impacts associated with intensive fertiliser use. Further, creating awareness of soybean health benefits needs to be accompanied by market development because the food industry has specific requirements to process soy such as protein concentration, oil quality and the reduction of anti-nutritional compounds.

The market for plant-based proteins is growing rapidly[1], with most hot and cold meat replacements comprising soybean[98], and soy milk having a higher protein content than other plant-based milk alternatives[99]. The increased demand for soy and the feasibility of increasing its production in Europe justify crop and agronomic improvement programmes that concurrently achieve higher productivity and better quality to deliver plant-based protein. Value chains based on European soy must allow cultivars and husbandry methods that meet industry and consumer demand. This appears to be critical as food-grade soy requires more flexible approaches than those currently on offer from exporting countries dependent on large-scale infrastructure and shipping. Large amounts of premium quality soybeans are required to achieve a dietary transition towards more climate-friendly foods.

## Methods

### Assessing European soy production and cultivated area

Data from EUROSTAT at NUTS 2 level was used to estimate 2020 area of arable land use and soy production per region[100]. As EUROSTAT data covers only EU nations, this data was supplemented by FAOSTAT country level data[7] for the remaining model-simulated areas. Spatial analysis of the data was conducted in R using the 'sf' package[101,102]. Area of each NUTS 2 region/country per model simulated region was calculated using spatial intersection. The area of arable land/soy production per model simulated region was estimated using the fractional area of each NUTS 2 region/country within each model simulated region. Predominant arable crops for each model-simulated region were calculated in the same way, using FAOSTAT 2020 data only (due to differences in reporting of arable crop types between the two data sources).

### European soybean production and imports

Soybean "Area harvested", "Yield", and "Production quantity" was downloaded from FAOSTAT[7] under the "Production", "Crops and livestock" domains. Country filter was "Europe > (List)", and Items were "Crops Primary", "Soybeans" and years from 1992 to 2020 (29 years). The list of 24 countries with 29 years of records for these variables were: Albania, Austria, Bosnia, Bulgaria, Croatia, Czechia, France, Germany, Greece, Hungary, Italy, Latvia, Moldova, Netherlands, North Macedonia, Poland, Romania, Russia, Serbia, Slovakia, Slovenia, Spain, Switzerland and Ukraine. Analysis of soybean production, area harvested, and yield progress focused on the subset of 12 countries that had on average more than 20,000 ha of soybean cultivation since 2017 (Russia, Ukraine, Italy, Serbia, Romania, France, Croatia, Austria, Hungary, Slovakia, Moldova, and Germany). Data on soybean imports (including intra Europe trade accounting for approximately 25% of total imports) for year 2020 was downloaded from FAOSTATS[7] under the "Trade", "Crops and livestock products" domains. Country filter was "Europe > (List)", and Items were "Soybeans", "Oil, Soybeans", and "Cake, Soybeans". The components "Oil,

Soybeans" and "Cake, Soybeans" were converted to soybean seed equivalents using 0.178 and 0.792 oil-to-seed and cake-to-seed ratios, respectively[103]. The list of 39 countries for this dataset included: Albania, Austria, Belarus, Belgium, Bosnia, Bulgaria, Croatia, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Montenegro, Netherlands, North Macedonia, Norway, Poland, Portugal, Moldova, Romania, Russia, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Ukraine and UK. Data processing and visualisation was performed with the dplyr[104] and ggplot2 in R[105]. Increase in soybean cultivation was fitted using a Local Polynomial Regression (loess) from the R statistical software. Yield potential and water-limited simulations for each of the 12 selected soybean producing countries was performed with the model described in the next section. Yield trends from FAO data were analysed using linear regression (*lm()* function) in R. We lean toward using a linear regression instead of a non-linear approach to describe yield trends over time because genetic and agronomical improvements can be attributed to the gradual accumulation of small gains and the consistent application of these advancements. Additionally, our focus was to use the most parsimonious statistical model, which, for the reason mentioned above, is the linear regression model[106].

## Crop modelling and simulations

Soybean performance was simulated by employing a phenological model and physiological model across Europe and surrounding regions with boundaries of a bounding box defined by a minimum and maximum latitude of 35.8 and 61.2 decimal degrees, respectively, and a minimum and maximum longitude of − 10.5 and 48.9 decimal degrees, respectively. The simulations were run on a 0.5º resolution grid (the centroid of which is hereby referred to as a site) for a total of 4,036 sites. Weather inputs of daily minimum temperature, maximum temperature, precipitation, and solar radiation were obtained for each site from The Weather Company's (an IBM business) Cleaned Historical Observations for 20 years from 1999 to 2018, inclusive. The timing of soybean developmental stages was simulated using an ensemble dynamic phenology model, previously trained on transcontinental data[107]. The publicly available phenology model was modified to include a planting algorithm; planting date for each site and year was determined adaptively based on a rolling mean of growing degree units (GDUs) calculated with base 10 °C and maximum 30 °C. If the seven-day rolling daily mean exceeded 4.4 GDUs, planting would occur, with March 1st as the earliest possible planting date. Emergence was assumed to occur after accumulating 70 GDUs[108], and therefore planting under conditions that allow emergence within fewer than 16 days. The timing of soybean developmental stages was simulated using an ensemble phenology model with a daily time step[107]. This model ensemble simulates phenological stages based on two physiological models (CROPGRO and SOYDEV) and one machine learning model. Both model components are driven by weather variables such as mean temperature, and daylength. The model was calibrated with ~ 13,500 data points from cultivars maturity group 000 to X across[106]. The phenological stages predicted by the model, emergence, flowering (R1), initiation of seed filling (R5), and physiological maturity (R7), were used to drive a physiology model, also referred to as a crop growth model. The mean absolute error of prediction for the phenological model was ~ 5 days. The public soybean crop growth model, previously parameterized from soybean observations, is based on established principles[109] using a modified plant leaf area expansion function[110] and empirical coefficients for the parameters of the expolinear growth equation of the leaf area index (LAI) and leaf partitioning[108]. For each site and year, the model was run for each maturity group from 0000 to III. This analysis did not parameterize neither the phenological or physiological models to the European observations, and so model performance reported is strictly out of sample fits. The quantification of performance is to illustrate the relative potential or plausibility of soybean crops in European geographies.

Four model configurations were run: a "yield potential" simulation calculated outcomes when only solar radiation and temperature were limiting factors (i.e., no water or nutrient limitation), and "rainfed" simulations (i.e. rainfed yield potential) with three different soil moisture initial conditions used to calculate the outcomes under water limitation, generating a total of 2,260,160 unique simulations (4036 sites × 20 years × 7 maturity groups × 4 configurations). For the rainfed simulations, local soil properties (including water holding capacity and wilting point) and initial water availability were obtained using the soil moisture active passive SMAP mission's v4 L4 land model constants (SPL4SMLM), analysis updates (SPL4SMAU), and geophysical data (SPL4SMGP) products. The SMAP root zone depth of 1 m was used to define maximum rooting depth for soybeans[111], and a daily average was used for non-SPL4SMLM variables reported at a temporal resolution finer than 1 day[112,113]. Initial conditions corresponding to planting dates outside of the SMAP mission (prior to 2015-03-31) were assumed to be the mean daily value of each site for that calendar day for the six years from 2015 to 2020 (inclusive). Initial plant available water was derived from the estimated SMAP rootzone soil moisture for the planting date (i.e. SPL4SMGP "sm_rootzone"). The ensemble standard deviation of the SMAP root zone soil moisture (i.e. SPL4SMAU "sm_rootzone_ensstd") was used to estimate the uncertainty in initial soil water conditions, and three initial conditions were used for the rainfed simulations: − 1, 0 and + 1 standard deviation conditions. Given the initial rootzone soil moisture, initial plant available water was defined using the rootzone wilting point (i.e. SPL4SMLM "clsm_wp"); the SMAP holding capacity constant (i.e. SPL4SMLM "clsm_poros") was used as a basis for determining runoff. Ultimately, each site was characterised by 20 years × 7 maturity groups × 4 water availability configurations (560 simulations per site). In all cases, simulated years were reported as the mean of 20-year weather data).

## Simulation data processing

Outputs were processed to define plausible attainable yields and crop success rates. A given simulation at a site and season (i.e. a given cropping season) was considered a failure if: (1) the mean temperature fell below 0 °C on any day between plant emergence and physiological maturity (R7); (2) the duration between planting and R7 was greater than 170 days following literature[114] that indicated crops taking longer than 170 days in European

environments would be uncommon; or (3) the adaptive planting algorithm never resulted in planting. The yield for a failed site and season was assigned as 0 Mg ha$^{-1}$. For crops that completed their cycle successfully, 7 Mg ha$^{-1}$ was considered as a maximum plausible attainable yield. Whilst yield potentials greater than this are plausible[115], available European soybean yield trial data (Supplementary Table 2, Supplementary Fig. 1) did not support that yields greater than 7 Mg ha$^{-1}$ would be achievable, and this ceiling was set in the simulations as a conservative maximum yield estimate. Comparing simulated rainfed yield potential to observed rainfed yield trials indicated that simulated potentials were plausible. Rainfed yield potentials of trials from Belgium, Germany, Italy, the Netherlands, Spain, and the United Kingdom were performed without any calibration to the data. The simulations were compared to the observed mean trial yield of each grown maturity, indicating plausible simulated water-limited yield potential (n = 84; Pearson correlation = 0.48; Mean Bias Error = -0.60 Mt ha$^{-1}$; Mean Absolute Percentage Error = 36%; Supplementary Fig. 1).

For each site, three site-level summary statistics were generated: the mean rainfed yield for each of the seven maturity groups (7 variables representing, for each of the 7 maturity groups, the mean of 20 years × 3 initial water conditions); the proportion of years reaching physiological maturity (7 variables representing, for each of the 7 maturity groups, the proportion of non-failures as defined above across 20 years); and the mean planting day of the year (1 variable representing the mean planting day across 20 years, where planting date was identical across maturities). These 15 statistics (7 maturity groups, with their planting date (1 date for all), success rate for each group, and their yield) for each site were used to define regions of similar potential soybean performance across the study area using spatial clustering. Spatial weights were defined based on a 75 km distance band and, following min–max normalization of each variable, and a bottom-up agglomerative hierarchical clustering of individual pixels. The Ward's method on Euclidean distance was performed with PySal and Scikit-learn to generate 16 clusters[116,117].

## Environmental savings

To indicate the magnitude of potential environmental savings achievable through enhanced European production of soybean, six simplified scenarios were generated to represent maximum agronomic potential and associated land and GHG emission savings under different contexts of soy production and consumption. These highly stylised scenarios do not represent predictions or forecasts based on current trends, but are designed to indicate the magnitude of possible GHG mitigation associated with possible supply and demand side changes. Two scenarios were created to represent soybean cultivation on 50% of set-aside land, estimated for relevant agroclimatic regions (soybean yields of at least 3 t ha$^{-1}$) based on an average set-aside ratio of 6.1% of arable land in 2022[118]. These scenarios assumed no crop displacement effects, and reflected current (Scenario 1, Sc-1) and potential (Scenario 2, Sc-2) yields from Table 1. Aggregate soybean production was therefore calculated as the sum product of 3.05% of arable areas (50% of 6.1% set aside) for regions supporting at least 3 t ha$^{-1}$ yield (Table 1) multiplied by respective yields in those regions. Another pair of scenarios considered displacement of wheat based on data presented in Table 1. These scenarios assume that 10% of wheat acreage was converted to soy cultivation, in regions where average soybean yields of at least 3 t ha$^{-1}$ could be achieved—at current (Scenario 3, Sc-3) or potential yields (Scenario 4, Sc-4), respectively (same approach to sum product calculation by area per region multiplied by yield per region). Wheat displacement is a simplified yet relevant and conservative assumption for the following reasons: (i) wheat cultivation represents 11–48% of arable areas in regions studied here, proving an obvious target for land sparing and crop system diversification; (ii) large volumes of wheat grain go to animal feed, and could be avoided through the diet change component of scenarios studied here (explicitly modelled in results). The soybean produced was assumed to replace soybean imported into Europe from Brazil, with land and GHG emission intensities taken from the Agribalyse life cycle assessment (LCA)[119] database accessed via OpenLCA[119] – a GHG intensity of 1.39 kg CO$_2$e per kg soybean is conservative (relatively low) compared with emissions intensities reported in previous studies and Ecoinvent v3.9[120]. The tonnage of wheat displaced was calculated based on the area of soybean cultivation multiplied by an average yield of 7.1 t ha$^{-1}$ based on conventional soft wheat grain production for animal feed in France from Agribalyse. Finally, Sc-3 & 4 were adapted to consider half of the soybean production displacing animal protein, equally split across chicken, pork and milk o the basis of 1:1 protein substitution (Table 1) for chicken and pork, and substitution of 7 kg of cows' milk with 1 kg soybeans[120] at current or potential yields (Scenario 5, Sc-5 and Scenario 6, Sc-6, respectively). This substitution provides an estimate of the approximate magnitude of environmental savings that could be leveraged by soy consumption as part of a more radical sustainable food system transformation, as advocated by, *inter alia*, Springmann et al.[121] Agribalyse inventory data for European production of broilers, pig meat and milk were used to estimate avoided quantities of main feed ingredients (wheat, barley, soybean, maize, rapeseed and sunflower), as well as total avoided production GHG emissions. In turn, inventory data from Agribalyse for European production of the main feed ingredients, excluding soybean, were used to estimate avoided land areas and fertiliser-N applications associated with avoidance of feed production. Avoided soybean production associated with animal protein substitution (50% of new European soybean cultivation) and direct substitution of imported soybean (50% of new European soybean cultivation) was represented based on inventory data for imported Brazilian soybean, as already described. These effects were summed to indicate aggregate GHG, land and fertiliser-N savings associated with animal protein substitution. Finally, land availability will be a major constraining factor for carbon dioxide removal (CDR) activities, such as afforestation, necessary to achieve Paris Agreement commitments[121]. Thus, sparing land from food production could leverage considerable GHG mitigation via CDR – represented here by applying a default carbon sequestration potential value of 3.6 t ha$^{-1}$ year$^{-1}$ proposed by Searchinger et al.[97] to net land areas spared from animal feed production. These scenarios do not intend to predict the future, but to illustrate the magnitude of environmental savings technically achievable if different agri-food strategies incorporate ambitious expansion of European soybean cultivation. These strategies

range from business-as-usual plus simple substitution of imported feed, to more a more radical food system transformation that previous studies have indicated is necessary to respect planetary boundaries[122].

## Soybean effects on health and industrial properties

Relevant literature was included to identify and synthesize evidence on soybean properties important to human nutrition, health and traits related to food processing in order to compare soy with other protein sources (both plant and animal based). Data on the nutritional value of soy compared to other food products was drawn primarily from the UK Government Composition of Foods Integrated Dataset 2019[123]. This database was selected as it provides data for a wide range of foodstuffs and enables direct comparison across a number of nutritional properties. Additional data for nutritional quality of milks, tofu and maize flour were not available from the database. This data was drawn from other sources as shown in Supplementary Table 1.

## Data availability

## References

1. Commission, E. *EU Agricultural Outlook for Markets—Income and Environment*. 2021–2031 (2021).
2. Messina, M. J. Legumes and soybeans: Overview of their nutritional profiles and health effects. *Am. J. Clin. Nutr.* **70**, 439S-450S. https://doi.org/10.1093/ajcn/70.3.439s (1999).
3. Jenkins, D. J. *et al.* Soy protein reduces serum cholesterol by both intrinsic and food displacement mechanisms. *J. Nutr.* **140**, 2302S-2311S. https://doi.org/10.3945/jn.110.124958 (2010).
4. Taku, K. *et al.* Effects of soy isoflavone extract supplements on blood pressure in adult humans: Systematic review and meta-analysis of randomized placebo-controlled trials. *J. Hypertens.* **28**, 1971–1982. https://doi.org/10.1097/HJH.0b013e32833c6edb (2010).
5. Yan, Z., Zhang, X., Li, C., Jiao, S. & Dong, W. Association between consumption of soy and risk of cardiovascular disease: A meta-analysis of observational studies. *Eur. J. Prevent. Cardiol.* **24**, 735–747. https://doi.org/10.1177/2047487316686441 (2017).
6. Poore, J. & Nemecek, T. Reducing food's environmental impacts through producers and consumers. *Science* **360**, 987–992. https://doi.org/10.1126/science.aaq0216 (2018).
7. Faostats. (Food and Agricultural Organisation of the United Nations, Rome, Italy, 2022).
8. Specht, J. E. *et al. Soybean.* **59901**, 311–356. https://doi.org/10.2135/cssaspecpub33.c12 (2014).
9. de Felipe, M., Gerde, J. A. & Rotundo, J. L. Soybean genetic gain in maturity groups III to V in Argentina from 1980 to 2015. *Crop Sci.* **56**, 3066–3077. https://doi.org/10.2135/cropsci2016.04.0214 (2016).
10. Dong, L. *et al.* Genetic basis and adaptation trajectory of soybean from its temperate origin to tropics. *Nat. Commun.* **12**, 5445–5445. https://doi.org/10.1038/s41467-021-25800-3 (2021).
11. Schmutz, J. *et al.* Genome sequence of the palaeopolyploid soybean. *Nature* **463**, 178–183. https://doi.org/10.1038/nature08670 (2010).
12. Xie, M. *et al.* A reference-grade wild soybean genome. *Nat. Commun.* **10**, 1216–1216. https://doi.org/10.1038/s41467-019-09142-9 (2019).
13. Song, X. P. *et al.* Massive soybean expansion in South America since 2000 and implications for conservation. *Nat. Sustain.* **4**, 784–792. https://doi.org/10.1038/s41893-021-00729-z (2021).
14. van Loon, M. P. *et al.* Grain legume production in Europe for food, feed and meat-substitution. *Glob. Food Secur.* https://doi.org/10.1016/j.gfs.2023.100723 (2023).
15. Cernay, C., Ben-Ari, T., Pelzer, E., Meynard, J. M. & Makowski, D. Estimating variability in grain legume yields across Europe and the Americas. *Sci. Rep.* **5**, 1–11. https://doi.org/10.1038/srep11171 (2015).
16. Yamaguchi, N., Ohnishi, S. & Miyoshi, T. Screening for chilling-tolerant soybeans at the flowering stage using a seed yield- and maturity-based evaluation method. *Crop Sci.* **58**, 312–320. https://doi.org/10.2135/cropsci2017.06.0392 (2018).
17. Long, S. P., Humphries, S. & Falkowski, P. G. Photoinhibition of photosynthesis in nature. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **45**, 633–662. https://doi.org/10.1146/annurev.pp.45.060194.003221 (1994).
18. Santini, M., Noce, S., Antonelli, M. & Caporaso, L. Complex drought patterns robustly explain global yield loss for major crops. *Sci. Rep.* **12**, 5792–5792. https://doi.org/10.1038/s41598-022-09611-0 (2022).
19. Montes, C. M. *et al.* High-throughput characterization, correlation, and mapping of leaf photosynthetic and functional traits in the soybean (*Glycine max*) nested association mapping population. *Genetics* https://doi.org/10.1093/genetics/iyac065 (2022).
20. de Visser, C. L. M., Schreuder, R. & Stoddard, F. The EU's dependency on soya bean import for the animal feed industry and potential for EU produced alternatives. *Oilseeds Fats Crops Lipids* **21**, D407–D407. https://doi.org/10.1051/ocl/2014021 (2014).
21. European, P. A European strategy for the promotion of protein crops. Report No. 2017/2116(INI), (Brussels, 2018).
22. Peoples, M. B., Giller, K. E., Jensen, E. S. & Herridge, D. F. Quantifying country-to-global scale nitrogen fixation for grain legumes: I. Reliance on nitrogen fixation of soybean, groundnut and pulses. *Plant Soil* **469**, 1–14. https://doi.org/10.1007/s11104-021-05167-6 (2021).
23. European, C. (European Comission, Directorate-General for Trade, Brussels, Belgium, 2022).
24. Eriksson, D. *et al.* A comparison of the EU regulatory approach to directed mutagenesis with that of other jurisdictions, consequences for international trade and potential steps forward. *New Phytol.* **222**, 1673–1684. https://doi.org/10.1111/nph.15627 (2019).
25. Gesch, R. W., Archer, D. W. & Spokas, K. Can using polymer-coated seed reduce the risk of poor soybean emergence in no-tillage soil?. *Field Crops Res.* **125**, 109–116. https://doi.org/10.1016/j.fcr.2011.09.005 (2012).
26. Tambussi, E. A., Bartoli, C. G., Guiamet, J. J., Beltrano, J. & Araus, J. L. Oxidative stress and photodamage at low temperatures in soybean (*Glycine max* L. Merr.) leaves. *Plant Sci.* **167**, 19–26. https://doi.org/10.1016/j.plantsci.2004.02.018 (2004).
27. European, C., Development, D. G. A. & Rural. Short-term outlook for EU agricultural markets, Spring 2022. (European Commission, DG Agriculture and Rural Development, Brussels, 2022)
28. Shahbandeh, M. Forecast volume of pig meat consumed in the European Union (EU-27) from 2015 to 2031.
29. Sudarić, A., Matosa Kocar, M., Duvnjak, T., Zdunić, Z. & Markulj Kulundžić, A. In *Soybean for Human Consumption and Animal Feed* (ed Aleksandra Sudarić) 23–61 (InteChopen, 2019).

30. Liener, I. E. Implications of antinutritional components in soybean foods. *Crit. Rev. Food Sci. Nutr.* **34**, 31–67 (1994).
31. Hertzler, S. R., Lieblein-Boff, J. C., Weiler, M. & Allgeier, C. Plant proteins: Assessing their nutritional quality and effects on health and physical function. *Nutrients* **12**, 1–27. https://doi.org/10.3390/nu12123704 (2020).
32. Boehm, J. D. Jr. *et al.* Genetic mapping and validation of the loci controlling 7S alpha' and 11S A-type storage protein subunits in soybean (*Glycine max* L. Merr.). *Theor. Appl. Genet.* **131**, 659–671. https://doi.org/10.1007/s00122-017-3027-9 (2018).
33. Wang, F. *et al.* Study on the tofu quality evaluation method and the establishment of a model for suitable soybean varieties for Chinese traditional tofu processing. *LWT Food Sci. Technol.* **117**, 108441–108441. https://doi.org/10.1016/j.lwt.2019.108441 (2020).
34. Krishnan, H. B. & Jez, J. M. Review: The promise and limits for enhancing sulfur-containing amino acid content of soybean seed. *Plant Sci.* **272**, 14–21. https://doi.org/10.1016/j.plantsci.2018.03.030 (2018).
35. Rotundo, J. L., Miller-Garvin, J. E. & Naeve, S. L. Regional and temporal variation in soybean seed protein and oil across the United States. *Crop Sci.* **56**, 797–808. https://doi.org/10.2135/cropsci2015.06.0394 (2016).
36. Zhang, D. *et al.* Elevation of soybean seed oil content through selection for seed coat shininess. *Nat. Plants* **4**, 30–35. https://doi.org/10.1038/s41477-017-0084-7 (2018).
37. Riaz, M. N. *Soy Applications in Food* 1st edn. (CRC Press, Boca Raton, 2005).
38. Hacisalihoglu, G. & Settles, A. M. Quantification of seed ionome variation in 90 diverse soybean (*Glycine max*) lines. *J. Plant Nutr.* **40**, 2808–2817. https://doi.org/10.1080/01904167.2017.1382528 (2017).
39. Messina, M. & Messina, V. The role of soy in vegetarian diets. *Nutrients* **2**, 855–888. https://doi.org/10.3390/nu2080855 (2010).
40. Murray-Kolb, L. E., Welch, R., Theil, E. C. & Beard, J. L. Women with low iron stores absorb iron from soybeans. *Am. J. Clin. Nutr.* **77**, 180–184. https://doi.org/10.1093/ajcn/77.1.180 (2003).
41. Burton, J. W., Harlow, C. & Theil, E. C. Evidence for reutilization of nodule iron in soybean seed development. *J. Plant Nutr.* **21**, 913–927. https://doi.org/10.1080/01904169809365453 (1998).
42. Rizzo, G. & Baroni, L. Soy, soy foods and their role in vegetarian diets. *Nutrients*. https://doi.org/10.3390/nu10010043 (2018).
43. Gupta, S. K. & Manjaya, J. G. *Advances in Improvement of Soybean Seed Composition Traits Using Genetic, Genomic and Biotechnological Approaches* Vol. 218 (Springer, New York, 2022).
44. Sakkas, P., Royer, E., Smith, S., Oikeh, I. & Kyriazakis, I. Combining alternative processing methods for European soybeans to be used in broiler diets. *Anim. Feed. Sci. Technol.* **253**, 45–55. https://doi.org/10.1016/j.anifeedsci.2019.03.010 (2019).
45. Gu, C., Pan, H., Sun, Z. & Qin, G. Effect of soybean variety on anti-nutritional factors content, and growth performance and nutrients metabolism in rat. *Int. J. Mol. Sci.* **11**, 1048–1056. https://doi.org/10.3390/ijms11031048 (2010).
46. Kumar, V. *et al.* Sucrose and raffinose family oligosaccharides (RFOs) in soybean seeds as influenced by genotype and growing location. *J. Agric. Food Chem.* **58**, 5081–5085. https://doi.org/10.1021/jf903141s (2010).
47. Qiu, D. *et al.* Identification and characterization of a stachyose synthase gene controlling reduced stachyose content in soybean. *Theor. Appl. Genet.* **128**, 2167–2176. https://doi.org/10.1007/s00122-015-2575-0 (2015).
48. Zheng, R. *et al.* Effect of soybean oligosaccharides on immunity and TLR2-NF-κB signal pathway response for weanling pigs. *J. Food Agric. Environ.* **10**, 273–279 (2012).
49. Mozzoni, L. & Chen, P. Correlations of yield and quality traits between immature and mature seed stages of edamame soybean. *J. Crop Improv.* **33**, 67–82. https://doi.org/10.1080/15427528.2018.1542366 (2019).
50. Escamilla, D. M., Rosso, M. L., Holshouser, D. L., Chen, P. Y. & Zhang, B. Improvement of soybean cultivars for natto production through the selection of seed morphological and physiological characteristics and seed compositions: A review. *Plant Breed.* **138**, 131–139. https://doi.org/10.1111/pbr.12678 (2019).
51. Kocar, M. M. *et al.* Assessment of phenotypic variability of saccharides in soybean genotypes suitable for growing in Europe. *J. Central Eur. Agric.* **21**(1), 92–103. https://doi.org/10.5513/JCEA01/21.1.2452 (2020).
52. Wiederstein, M., Baumgartner, S. & Lauter, K. Soybean (*Glycine max*) allergens—a review on an outstanding plant food with allergenic potential. *ACS Food Sci. Technol.* **3**, 363–378. https://doi.org/10.1021/acsfoodscitech.2c00380 (2023).
53. Krishnan, H. B., Kim, W.-S., Jang, S. & Kerley, M. S. All three subunits of soybean β-conglycinin are potential food allergens. *J. Agric. Food Chem.* **57**, 938–943. https://doi.org/10.1021/jf802451g (2009).
54. L'Hocine, L. & Boye, J. I. Allergenicity of soybean: New developments in identification of allergenic proteins, cross-reactivities and hypoallergenization technologies. *Crit. Rev. Food Sci. Nutr.* **47**, 127–143 (2007).
55. Blanco-Mejia, S. *et al.* A meta-analysis of 46 studies identified by the FDA demonstrates that soy protein decreases circulating LDL and total cholesterol concentrations in adults. *J. Nutr.* **149**, 968–981. https://doi.org/10.1093/jn/nxz020 (2019).
56. Harland, J. I. Food combinations for cholesterol lowering. *Nutr. Res. Rev.* **25**, 249–266. https://doi.org/10.1017/S0954422412000170 (2012).
57. Lovati, M. R. *et al.* Soy protein peptides regulate cholesterol homeostasis in Hep G2 cells. *J. Nutr.* **130**, 2543–2549. https://doi.org/10.1093/jn/130.10.2543 (2000).
58. Manzoni, C. *et al.* Subcellular localization of soybean 7S globulin in HepG2 cells and LDL receptor up-regulation by its alpha' constituent subunit. *J. Nutr.* **133**, 2149–2155. https://doi.org/10.1093/jn/133.7.2149 (2003).
59. Xiao, C. W., Mei, J. & Wood, C. M. Effect of soy proteins and isoflavones on lipid metabolism and involved gene expression. *Front. Biosci. Landmark* **13**, 2660–2673. https://doi.org/10.2741/2873 (2008).
60. Miras-Moreno, B., Sabater-Jara, A. B., Pedreno, M. A. & Almagro, L. Bioactivity of phytosterols and their production in plant in vitro cultures. *J. Agric. Food Chem.* **64**, 7049–7058. https://doi.org/10.1021/acs.jafc.6b02345 (2016).
61. Rowland, I. *et al.* Bioavailability of phyto-oestrogens. *Br. J. Nutr.* **89**, S45-58. https://doi.org/10.1079/BJN2002796 (2003).
62. Zaheer, K. & Humayoun Akhtar, M. An updated review of dietary isoflavones: Nutrition, processing, bioavailability and impacts on human health. *Crit. Rev. Food Sci. Nutr.* **57**, 1280–1293. https://doi.org/10.1080/10408398.2014.989958 (2017).
63. Al-Tawaha, A. M. & Seguin, P. Seeding date, row spacing, and weed effects on soybean isoflavone concentrations and other seed characteristics. *Can. J. Plant Sci.* **86**, 1079–1087. https://doi.org/10.4141/p06-043 (2006).
64. Hoeck, J. A., Fehr, W. R., Murphy, P. A. & Welke, G. A. Influence of genotype and environment on isoflavone contents of soybean. *Crop Sci.* **40**, 48–51. https://doi.org/10.2135/cropsci2000.40148x (2000).
65. Yamaya, A., Endo, Y., Fujimoto, K. & Kitamura, K. Effects of genetic variability and planting location on the phytosterol content and composition in soybean seeds. *Food Chem.* **102**, 1071–1075. https://doi.org/10.1016/j.foodchem.2006.07.001 (2007).
66. Gerde, J. A. *et al.* In *Bailey's Industrial Oil and Fat Products* 1–68 (2020).
67. Saini, R. K. & Keum, Y. S. Omega-3 and omega-6 polyunsaturated fatty acids: Dietary sources, metabolism, and significance - A review. *Life Sci.* **203**, 255–267. https://doi.org/10.1016/j.lfs.2018.04.049 (2018).
68. Slavin, M., Cheng, Z. H., Luther, M., Kenworthy, W. & Yu, L. L. Antioxidant properties and phenolic, isoflavone, tocopherol and carotenoid composition of Maryland-grown soybean lines with altered fatty acid profiles. *Food Chem.* **114**, 20–27. https://doi.org/10.1016/j.foodchem.2008.09.007 (2009).
69. Turquetti-Moraes, D. K., Moharana, K. C., Almeida-Silva, F., Pedrosa-Silva, F. & Venancio, T. M. Integrating omics approaches to discover and prioritize candidate genes involved in oil biosynthesis in soybean. *Gene* **808**, 145976–145976. https://doi.org/10.1016/j.gene.2021.145976 (2022).
70. Fehr, W. R. Breeding for modified fatty acid composition in soybean. *Crop Sci.* **47**, S-87. https://doi.org/10.2135/cropsci2007.04.0004IPBS (2007).

71. Combs, R. & Bilyeu, K. Novel alleles of FAD2-1A induce high levels of oleic acid in soybean oil. *Mol. Breed.* **39**, 79–79. https://doi.org/10.1007/s11032-019-0972-9 (2019).
72. Ghani, M., Kulkarni, K. P., Song, J. T., Shannon, J. G. & Lee, J.-D. Soybean sprouts: A review of nutrient composition, health benefits and genetic variation. *Plant Breed. Biotechnol.* **4**, 398–412. https://doi.org/10.9787/PBB.2016.4.4.398 (2016).
73. Chalupa-Krebzdak, S., Long, C. J. & Bohrer, B. M. Nutrient density and nutritional value of milk and plant-based milk alternatives. *Int. Dairy J.* **87**, 84–92. https://doi.org/10.1016/j.idairyj.2018.07.018 (2018).
74. Florou-Paneri, P. *et al.* Alternative protein sources to soybean meal in pig diets. *J. Food Agric. Environ.* **12**, 655–660. https://doi.org/10.1234/4.2014.5214 (2014).
75. Agnoli, C. *et al.* Position paper on vegetarian diets from the working group of the Italian Society of Human Nutrition. *Nutr. Metab. Cardiovasc. Dis.* **27**, 1037–1052. https://doi.org/10.1016/j.numecd.2017.10.020 (2017).
76. Carbonaro, M., Maselli, P. & Nucara, A. Relationship between digestibility and secondary structure of raw and thermally treated legume proteins: A Fourier transform infrared (FT-IR) spectroscopic study. *Amino Acids* **43**, 911–921. https://doi.org/10.1007/s00726-011-1151-4 (2012).
77. Lee, S. J. *et al.* Comparison of isoflavone concentrations in soybean (*Glycine max* L. Merrill) sprouts grown under two different light conditions. *J. Agric. Food Chem.* **55**, 9415–9421. https://doi.org/10.1021/jf071861v (2007).
78. Liu, J. *et al.* Partial improvements in the flavor quality of soybean seeds using intercropping systems with appropriate shading. *Food Chem.* **207**, 107–114. https://doi.org/10.1016/j.foodchem.2016.03.059 (2016).
79. Kumar, V., Rani, A., Billore, S. D. & Chauhan, G. S. Physico-chemical properties of immature pods of Japanese soybean cultivars. *Int. J. Food Prop.* **9**, 51–59. https://doi.org/10.1080/10942910500471727 (2006).
80. Rotundo, J. L. & Westgate, M. E. Meta-analysis of environmental effects on soybean seed composition. *Field Crops Res.* **110**, 147–156. https://doi.org/10.1016/j.fcr.2008.07.012 (2009).
81. Rao, M. S. S., Bhagsari, A. S. & Mohamed, A. I. Fresh green seed yield and seed nutritional traits of vegetable soybean genotypes. *Crop Sci.* **42**, 1950–1958. https://doi.org/10.2135/cropsci2002.1950 (2002).
82. Xu, B. J. & Chang, S. K. C. Total phenolics, phenolic acids, isoflavones, and anthocyanins and antioxidant properties of yellow and black soybeans as affected by thermal processing. *J. Agric. Food Chem.* **56**, 7165–7175. https://doi.org/10.1021/jf8012234 (2008).
83. European, P. Resolution on Food Security and Ukraine (2022/2593(RSP). (Brussels, 2022).
84. European, C., Development, D. G. A. & Rural. *Report from the Commission to the Council and the European Parliament on the Development of Plant Proteins in the European Union* (Directorate-General for Agriculture and Rural Development, Brussels, 2018).
85. Ipcc. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (eds V. Masson-Delmotte *et al.*) 3–32 (Cambridge University Press, 2021).
86. Veas, R. E. A. *et al.* Soybean seed growth dynamics exposed to heat and water stress during the filling period under field conditions. *J. Agron. Crop Sci.* **208**, 472–485. https://doi.org/10.1111/jac.12523 (2022).
87. Nendel, C. *et al.* Future area expansion outweighs increasing drought risk for soybean in Europe. 1340–1358. https://doi.org/10.1111/gcb.16562 (2023).
88. Schoving, C. *et al.* Combining simple phenotyping and photothermal algorithm for the prediction of soybean phenology: Application to a range of common cultivars grown in Europe. *Front. Plant Sci.* **10**, 1–14. https://doi.org/10.3389/fpls.2019.01755 (2020).
89. Coleman, K. *et al.* The potential for soybean to diversify the production of plant-based protein in the UK. *Sci. Total Environ.* **767**, 14490–14490 (2022).
90. Saleem, A. *et al.* Response of a diverse European soybean collection to "short duration" and "long duration" drought stress. *Front. Plant Sci.* **13**, 818766–818766. https://doi.org/10.3389/fpls.2022.818766 (2022).
91. Arya, H., Singh, M. B. & Bhalla, P. L. Towards developing drought-smart soybeans. *Front. Plant Sci.* https://doi.org/10.3389/fpls.2021.750664 (2021).
92. Annicchiarico, P. *et al.* Development and proof-of-concept application of genome-enabled selection for pea grain yield under severe terminal drought. *Int. J. Mol. Sci.* https://doi.org/10.3390/ijms21072414 (2020).
93. Guilpart, N., Iizumi, T. & Makowski, D. Data-driven projections suggest large opportunities to improve Europe's soybean self-sufficiency under climate change. *Nat. Food* **3**, 255–265. https://doi.org/10.1038/s43016-022-00481-3 (2022).
94. Yu, Y., Stomph, T. J., Makowski, D. & van der Werf, W. Temporal niche differentiation increases the land equivalent ratio of annual intercrops: A meta-analysis. *Field Crops Res.* **184**, 133–144. https://doi.org/10.1016/j.fcr.2015.09.010 (2015).
95. Simon-Miquel, G., Reckling, M., Lampurlanés, J. & Plaza-Bonilla, D. A win-win situation—Increasing protein production and reducing synthetic N fertilizer use by integrating soybean into irrigated Mediterranean cropping systems. *Eur. J. Agron.* https://doi.org/10.1016/j.eja.2023.126817 (2023).
96. Tanveer, M., Anjum, S. A., Hussain, S., Cerdà, A. & Ashraf, U. Relay cropping as a sustainable approach: Problems and opportunities for sustainable crop production. *Environ. Sci. Pollut. Res.* **24**, 6973–6988. https://doi.org/10.1007/s11356-017-8371-4 (2017).
97. Searchinger, T. D., Wirsenius, S., Beringer, T. & Dumas, P. Assessing the efficiency of changes in land use for mitigating climate change. *Nature* **564**, 249–253. https://doi.org/10.1038/s41586-018-0757-z (2018).
98. Pointke, M. & Pawelzik, E. Plant-based alternative products: Are they healthy alternatives? Micro- and macronutrients and nutritional scoring. *Nutrients* **14**, 601–601 (2022).
99. Scholz-Ahrens, K. E., Ahrens, F. & Barth, C. A. Nutritional and health attributes of milk and milk imitations. *Eur. J. Nutr.* **59**, 19–34. https://doi.org/10.1007/s00394-019-01936-3 (2020).
100. Eurostats. (European Commission, DG Agriculture and Rural Development, 2020).
101. Pebesma, E. Simple features for R: Standardized support for spatial vector data. *R J.* **10**, 439–446. https://doi.org/10.32614/rj-2018-009 (2018).
102. Pebesma, E. & Bivand, R. *Spatial Data Science: With Applications in R* 1st edn. (Chapman and Hall, Boca Raton, 2023).
103. United States Soybean Exportation Council. Conversion Table - Soy Markets. Retrieved October 10 from https://ussec.org/resources/conversion-table/ (2015).
104. Wickham, H., François, R., Henry, L. & Müller, K. A grammar of data manipulation [R package dplyr version 1.0.2] (2020).
105. Wickham, H. *Ggplot2: Elegant graphics for data analysis*, 2nd edn (Springer International Publishing, 2016).
106. Hammer, G., Messina, C., Wu, A. & Cooper, M. Biological reality and parsimony in crop models—why we need both in crop improvement!. *In Silico Plants* **1**(1), diz010 (2019).
107. McCormick, R. F. *et al.* Intercontinental prediction of soybean phenology via hybrid ensemble of knowledge-based and data-driven models. *In Silico Plants* **3**, 1–12. https://doi.org/10.1093/insilicoplants/diab004 (2021).
108. Soltani, A. & Sinclair, T. R. *Modeling Physiology of Crop Development, Growth and Yield* (CAB International, Wallingford, 2012).
109. Sinclair, T. R. Water and nitrogen limitations in soybean grain production. 1. Model development. *Field Crops Res.* **15**, 125–141 (1986).
110. Goudriaan, J. In *Resource Capture by Crops* (ed J. L. Monteith) 99–110 (Nottingham University Press, 1994).
111. Fan, J., McConkey, B., Wang, H. & Janzen, H. Root distribution by depth for temperate agricultural crops. *Field Crops Res.* **189**, 68–74. https://doi.org/10.1016/j.fcr.2016.02.013 (2016).

112. Entekhabi, D. *et al.* The soil moisture active passive (SMAP) mission. *Proc. IEEE* **98**, 704–716. https://doi.org/10.1109/JPROC.2010.2043918 (2010).
113. Reichle, R. *et al.* SMAP L4 Global 3-hourly 9 km EASE-Grid Surface and Root Zone Soil Moisture Geophysical Data, Version 4 (2018).
114. Kurasch, A. K. *et al.* Identification of mega-environments in Europe and effect of allelic variation at maturity *E. loci* on adaptation of European soybean. *Plant Cell Environ.* **40**, 765–778. https://doi.org/10.1111/pce.12896 (2017).
115. Setiyono, T. D. *et al.* Simulation of soybean growth and yield in near-optimal growth conditions. *Field Crops Res.* **119**, 161–174. https://doi.org/10.1016/j.fcr.2010.07.007 (2010).
116. Pedregosa, F. *et al.* Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830. https://doi.org/10.48550/arXiv.1201.0490 (2011).
117. Rey, S. J. & Anselin, L. In (eds Fischer & Arthur Getis) 175–193 (Springer, Heidelberg, 2010).
118. Commission, E. *Short-Term Outlook for EU Agricultural Markets, Spring 2022* (European Commission, DG Agriculture and Rural Development, Brussel, 2022).
119. OpenLca. (France, 2002).
120. Wernet, G. *et al.* The ecoinvent database version 3 (part I): Overview and methodology. *Int. J. Life Cycle Assess.* **21**, 9 (2016).
121. Intergovernmental Panel on Climate Change. *Climate Change and Land: IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems* (Cambridge University Press, Cambridge, 2022).
122. Springmann, M. *et al.* Options for keeping the food system within environmental limits. *Nature* **562**, 519–525. https://doi.org/10.1038/s41586-018-0594-0 (2018).
123. Public Health England. *Composition of Foods Integrated Dataset (CoFID)—McCance and Widdowson's 'Composition of Foods Integrated Dataset' on the Nutrient Content of the UK Food Supply.* 25/03/2015 edn, (Public Health England, 2015).

## Acknowledgements

## Author contributions

J.L.R. and M.C.R. drafted and wrote the manuscript. J.L.R., R.Mc., S.K.T., and C.D.M. were responsible for the crop modelling work and interpretation. R.M., L.D., and J.L. were responsible for the human health section. E.G.E., J.A.G., E.C.S., and M.C.R. were responsible for the nutritional traits section. M.S.L., M.S., P.A., C.dV., M.R., A.D., J.P., I.S., and N.S. provided soybean yield field data. All authors provided critical feedback, interpretation, and revisions to the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-57522-z.

**Correspondence** and requests for materials should be addressed to J.L.R.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.