



**HAL**  
open science

## **Envirotyping to control genotype x environment interactions for efficient soybean breeding**

Chloé Elmerich, Michel-Pierre Faucon, Milagros Garcia, Patrice Jeanson, Guéno­lé  
Boulch, Bastien Lange

### ► **To cite this version:**

Chloé Elmerich, Michel-Pierre Faucon, Milagros Garcia, Patrice Jeanson, Guéno­lé Boulch, et al.. Envirotyping to control genotype x environment interactions for efficient soybean breeding. *Field Crops Research*, 2023, 303, pp.109113. <10.1016/j.fcr.2023.109113>. <hal-04218126>

**HAL Id: hal-04218126**

**<https://normandie-univ.hal.science/hal-04218126v1>**

Submitted on 26 Sep 2023

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

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



HAL Authorization

# 1 **Envirotyping to control genotype x environment interactions for efficient** 2 **soybean breeding**

3  
4 **Chloé Elmerich**<sup>1,2</sup>, Michel-Pierre Faucon<sup>1</sup>, Milagros Garcia<sup>2</sup>, Patrice Jeanson<sup>2</sup>, Guénoé Boulch<sup>1</sup>,  
5 Bastien Lange<sup>1</sup>

6 <sup>1</sup> AGHYLE, UniLaSalle, 19 rue Pierre Wagué, 60000, Beauvais, France

7 <sup>2</sup> Lidea Seeds, 6 chemin de Panedautes, 31700 Mondonville, France

8 \*author for correspondence: [chloe.elmerich@unilasalle.fr](mailto:chloe.elmerich@unilasalle.fr); [bastien.lange@unilasalle.fr](mailto:bastien.lange@unilasalle.fr)

## 9 10 **Highlights**

- 11 • Northward soybean deployment requires environmental characterisation using eco-climatic  
12 factors
- 13 • Envirotyping explained 88% of the genotype by environment interactions
- 14 • Five environment types contrasted by the intensity and timing of stresses
- 15 • Combine specific and broad adaptation is best suited in the target population of environments

## 16 17 **Abstract**

18 In the context of the European protein deficit and the need for climate change mitigation by agriculture,  
19 soybean (*Glycine max* L. Merr) is of major interest to farmers and breeders. Expanding the crop to new  
20 cultivation areas requires understanding and control of the Genotype by Environment Interactions (GEI) that  
21 impede the genetic gain. New envirotyping methods, including the Target Population of Environments (TPE)  
22 characterisation, are key to the development of efficient breeding programs for specific or broad adaptations.  
23 The objectives were to (i) determine the environment types describing the European early soybean TPE, (ii)  
24 characterise the distribution and repeatability of the environment types across the TPE to identify the best  
25 suited adaptation strategy and (iii) demonstrate the importance of assessing the alignment between multi-  
26 environmental trials and the TPE for breeding decisions. In this study, 602 environments from France to  
27 Russia were clustered into five environment types using twelve eco-climatic factors, *i.e.* environmental

28 variables that were calculated over specific phenological periods, such as the number of days below 15°C  
29 between flower induction and first flower stages or the solar radiation quantity between first pod and first  
30 seed stages. These factors were previously identified as main drivers of GEI for yield in early maturity  
31 soybean (maturity groups '000' and '00'). The environmental clustering explained 88% of the GEI effect on  
32 soybean yield. The five environment types that composed the TPE, mainly contrasted in the intensity and  
33 timing of stresses related to temperature (cold stress during the vegetative growth and heat stress during the  
34 reproductive growth) and water availability (precipitation amount, evapotranspiration and drought  
35 throughout the crop cycle). Interestingly, we observed geographical and temporal variations in the  
36 environment types distributions across the TPE as well as in their repeatability. These variations attested to  
37 the TPE heterogeneity and thus suggested that selection strategies based on either specific or broad  
38 adaptations should be combined. For example, specific adaptations to the third and fourth environment types  
39 were best suited in Eastern Europe while the broad adaptation to all environment types could be  
40 recommended in Western Europe. When broad adaptation was required, we demonstrated the need to assess  
41 the alignments between the environment types frequencies in the TPE and those observed when multi-  
42 environmental trials were conducted. This work will contribute to improving the existing soybean germplasm  
43 by considering the risks linked with weather variations and unpredictability so as to design elite soybean  
44 according to environment type.

45

46 **Key words**

47 Envirotyping, Multi-environment trial, Target population of environments, Soybean, Weighted selection

## 48 1. Introduction

49 Breeders aim to create cultivars that outperform the existing germplasm under a large set of growing  
50 conditions across the cultivation area. Spatial considerations and year-to-year variations in weather scenarios  
51 can considerably affect the relative performances of cultivars and make breeding decisions more complex  
52 (Cooper & DeLacy, 1994). This phenomenon, called Genotype by Environment Interactions (GEI), limits  
53 the genetic gain by decreasing the effective heritability of traits, *i.e.* the ability to identify superior genotypes  
54 statistically (Brennan et al., 1981; Basford & Cooper, 1998). When the GEI are low, cultivars can be  
55 identified that perform well across a wide range of conditions; this is referred to as broad adaptation (Cooper  
56 & Byth, 1996). Alternatively, when there are greater GEI, certain cultivars can perform better than others  
57 within a restricted set of conditions. This is referred to as specific adaptation (Cooper & Byth, 1996). Crop  
58 growth models have become increasingly popular to unravel the complexity of the ecophysiological  
59 mechanisms and processes underlying GEI (Millet et al., 2019; Rincent et al., 2019; Casadebaig et al., 2021;  
60 Bustos-Korts et al., 2022). These models offer the possibility of connecting plant physiological processes  
61 with environmental variables (Chenu et al., 2017). For instance, they can be used to calculate climatic  
62 variables over specific simulated phenological periods. Known as environmental covariates or Eco-climatic  
63 Factors (EFs), they can be treated as explanatory variables of complex traits, including yield (Schoving et  
64 al., 2022, Elmerich et al., 2023).

65 To capture the GEI effect, breeders traditionally evaluate genotypes across Multi-Environment Trials  
66 (METs) that include a set of locations and years, referred to as environments. Such trials aim to characterise  
67 the environmental conditions in which the crops are likely to be grown. These conditions correspond to the  
68 Target Population of Environments (TPE) (Podlich & Cooper, 1998). Due to seasonal variability and the  
69 practical constraints that restrict the number of genotypes and environments tested in METs the environments  
70 sampled in METs often offer a biased representations of the TPE (Podlich et al., 1999). An extension of the  
71 breeder's equation used to quantify genetic gain was recently developed to explicitly account for the  
72 influence of the MET-TPE alignment on trait predictions (Cooper et al., 2023). Thus, weighted analyses  
73 based on representativity assessments of trials can enable breeders to correct for this bias and select  
74 germplasm that is better adapted to a TPE. This requires an upstream characterisation and understanding of  
75 the TPE through identification of the key environmental classes affecting the performance of the genotypes.

76 These environmental classes, referred to as environment types , regroup environments (location x year) in  
77 which genotypes have similar performances (Chenu et al., 2013). The classification can be performed  
78 according to geographical regions (Döttinger et al., 2023), stress factors patterns (Chapman et al., 2000) or  
79 effects on performance traits such as yield levels or maturity dates (Chauhan et al., 2013; Kurasch et al.,  
80 2017). While useful to describe environmental variables affecting crop productivity, these classification  
81 criteria do not attest to the performance stability of the genotypes. Consequently, the best approach for  
82 breeding purposes is grouping or clustering environments based on environment variables that are key drivers  
83 of GEI (Annicchiarico, 2021). The challenge is to identify these drivers and use them to cluster environments  
84 into environment types that define the TPE. This approach minimises GEI within environment types but  
85 maximises them between environment types (Chenu et al., 2011; Bustos-Korts et al., 2022).

86 Leguminous crop deployments have been accelerated to address European protein deficits (Marraccini et  
87 al., 2020; Ayerdi Gotor & Marraccini, 2022). Among legumes, soybean (*Glycine max* L. (Merrill)) seems to  
88 be a good candidate to become a major oleo-protein crop for both food and feed (Sudarić, 2020). European  
89 soybean production is currently low, and importations greatly exceed exportations (FAOSTAT, 2021).  
90 Although recent studies demonstrate the potential of soybean crop in Europe, few breeding efforts have been  
91 made to adapt soybean in Northern Europe (Boulch et al. 2021; Karges et al., 2022). Under highly divergent  
92 weather conditions in France, Elmerich et al. (2023) identified the main EFs driving GEI in semi-  
93 indeterminate soybeans from two early maturity groups ('000' and '00'). Various GEI-drivers were identified  
94 depending on maturity groups and the nature of the GEI (genotypes by location or locations by year). The  
95 GEI-drivers occurred mainly during the overlapping period of vegetative and reproductive growths. Cold  
96 stress during the vegetative growth, solar radiation intensity during the pod emission and maximum  
97 temperature before maturity were the main climatic variables that impacted GEI.

98 The potential early soybean cultivation area is wide in Northern Europe as it extends from France to  
99 Russia (approximately between the 45° and the 50° parallels). However, areas dedicated to soybean  
100 production remain low (3.6% of the total crops harvested in 2021) (FAOSTAT, 2021). Compared with other  
101 regions of the globe, Europe presents a substantial diversity of climatic scenarios (Woollings, 2010). Thus,  
102 GEI for yield are expected to be large, which will reduce the heritability of yield-related traits (Holland et  
103 al., 2003).

104 To date, the spatial and temporal characterisation of the European soybean TPE is unknown, especially  
105 when GEI-drivers are used. Attempts to show the potential of dividing the TPE for specific adaptation have  
106 not considered GEI *per se* nor the environmental characterisation (Kurasch et al., 2017; Döttinger et al.,  
107 2023). Those advances would enable breeders to position their cultivars on the market and contribute to  
108 designing better breeding strategies. The objectives will be to (i) determine the environment types that  
109 describe the European early soybean TPE using GEI-drivers, (ii) characterise the distribution and  
110 repeatability of environment types across the TPE to propose the best-suited adaptation strategy, and (iii)  
111 demonstrate the importance of MET-TPE alignment in breeding decisions and cultivars recommendations.  
112

## 113 **2. Material and Methods**

### 114 *2.1. Datasets*

115 Two datasets were used in this study; they differed in data sources and use purposes. Tested and virtual  
116 environments (a combination of location x year) were used in order to establish an adequate number of  
117 environments. Dataset A consisted of 112 environments in an unbalanced combination of 60 locations and  
118 five years (2017-2021). Terres Inovia, the French technical institute for oil and protein crops, used Multi-  
119 environment trials (METs) to test post-inscription cultivars in those environments  
120 (<https://www.myvar.fr/resultats/campagne-177.html>). The trials were conducted with rigorous chemical and  
121 mechanical weed management. Even though biotic stresses are limited for soybean in Northern Europe,  
122 annual trials are evaluated by a committee and then validated or rejected (and thus were not used in this  
123 study). A total of 57 cultivars from two early maturity groups ('000' and '00') were tested (supplementary  
124 table S1). The METs for the two maturity groups were conducted independently in France, between the forty-  
125 fifth and the fiftieth north parallels.

126 Dataset B consisted of a panel of untested environments across Europe. The studied early maturity groups  
127 '000' and '00' soybean cultivars are grown from France to Russia in locations situated between the forty-  
128 fifth° and the fifty-second parallels. This area includes regions in France, Germany, Belgium, Switzerland,  
129 the Czech Republic, Poland, Ukraine and Russia. Forty-nine locations were selected across this axis over a  
130 span of ten years (2012 to 2021) to produce 490 virtual environments.

131 Meteorological data for the two datasets were extracted from the Agri4Cast Resources Portal  
132 (<https://agri4cast.jrc.ec.europa.eu/dataportal/>). The closest grid point (in kilometres) was attributed to each  
133 environmental location. Daily minimum and maximum temperatures ( $^{\circ}\text{C}$ ), daily solar radiation ( $\text{MJ m}^{-2}$   
134  $\text{day}^{-1}$ ) and daily precipitation (mm) were extracted from 1 January to 31 December. Soil parameters extracted  
135 from raster files produced by the European Soil Data Centre (<https://esdac.jrc.ec.europa.eu/>) were used to  
136 characterise the soil of each location (1 km/1 km grid) using QGIS3 (v. 3.14.1). Clay content (%), silt content  
137 (%), sand content (%), gravel content (%), organic carbon content (%), total nitrogen content (%) and bulk  
138 density ( $\text{g cm}^{-3}$ ) from topsoil and subsoil were extracted as well as the depth available to roots (cm).

139

## 140 2.2. *Eco-climatic factors calculation*

141 A previous study using Dataset A (Elmerich et al., 2023), identified 20 out of 126 Eco-climatic Factors  
142 (EFs) for their primary impact on Genotype by Environment Interactions (GEI). The 126 EFs used in this  
143 study consisted of climatic variables calculated between two developmental stages (*i.e.* phenophases). Five  
144 major categories of environmental variables were used: period duration, temperature, water, solar radiation  
145 and stresses. Seven soybean phenophases were used: Sowing to EMergence (SEM), EMergence to Flower  
146 Initiation (EMFI), Flower Induction to First Flower (FIFF), First Flower to First Pod (FFFP), First Pod to  
147 First Seed (FPFS), First Seed to End of Pod (FSEP), End of Pod to Physiological Maturity (EPPM). These  
148 phenophases were calculated based on the simulation of stages using the DSSAT-CROPGRO-soybean model  
149 (Boote et al., 1998; Jones et al., 2003; Hoogenboom et al., 2019). The 126 EFs were sorted using partial least  
150 square regression and the variable importance in projection scores. This enabled identification of the main  
151 GEI-drivers.

152

## 153 2.3. *Environmental classification based on k-means clustering*

154 Clustering consists of grouping individuals into an optimal number of clusters based on one or a set of  
155 variables. In our study, the aim was to group the 602 environments (Datasets A and B) into clusters that  
156 represented the environment types; the intra-cluster GEI were minimised and the inter-cluster GEI were  
157 maximised. The environmental classification should be defined according to crucial environmental factors  
158 that affect GEI in the Target Population of Environments (TPE) (Annicchiarico, 2021).

159 The set of 20 EFs identified by Elmerich et al. (2023) appeared to be too large for the clustering. Usually,  
160 the number of variables used for clustering ranges from 6 to 12 (Corlouer et al., 2019; Schoving et al., 2022).  
161 Thus, 888 896 clustering models were tested (*i.e.* the combination of 6 to 12 EFs among the 20). K-means  
162 clustering was performed with R software (v4.2.1) using the Euclidean distance measure and the complete  
163 linkage method. For each clustering model, the optimal number of clusters was determined using the *nbclust*  
164 package in the R software (v4.2.1).

165

#### 166 2.4. Linear models

167 Each model was tested using R software (v4.2.1).

168 For each k-means model, a linear mixed-effects model – Equation 1 (E1) – was applied on the data of  
169 Dataset A (112 environments). Only cultivars tested at least three times in each cluster were kept. The model  
170 tested the fixed effects of Genotype (G), Cluster (C) and Genotype by Cluster Interactions (G x C), and the  
171 random effect of the Environments nested in the Clusters C(E) across the network:

$$172 Y_{ijk} = \mu + G_i + C_j + (G \times C)_{ij} + C_j(E_k) + \varepsilon_{ijk}, \quad (E1)$$

173 where  $Y_{ijk}$  is the seed yield of genotype  $i$  in cluster  $j$  in the environment  $k$ ,  $\mu$  is the population mean,  $G_i$  stands  
174 for the effect of genotype  $i$ ,  $C_j$  stands for the effect of cluster  $j$ ,  $(G \times C)_{ij}$  stands for the effect of the interaction  
175 between genotype  $i$  and cluster  $j$  and  $C_j(E_k)$  stands for the effect of the environment  $k$  nested in the cluster  $j$ .  
176 The residual is  $\varepsilon_{ijk}$ .

177 A linear model – Equation 2 (E2) – was applied to the data of Dataset A (112 environments) to test the  
178 effects of the EFs used in the clustering on the GxC:

$$179 GC_{ijk} = \sum_{n=1}^p \alpha_n EF_n + \varepsilon_{ijk}, \quad (p \in [6;12]), \quad (E2)$$

180 where,  $GC_{ijk}$  is the genotype by cluster interaction effect of genotype  $i$  in cluster  $j$  in the environment  $k$ , and  
181  $\alpha_n$  stands for the linear regression coefficient of the eco-climatic factor  $n$ ,  $EF_n$ . The residual is  $\varepsilon_{ijk}$ .

182 The evaluation of each model was based first on the significance of the GxC effect. A significant GxC  
183 effect indicated that there was some variation among genotypes in their rank orders between clusters. To  
184 ensure that the significance of the GxC effect was not driven by a single or a few genotypes and clusters, the

185 percentage of significant interactions between specific genotypes and clusters was calculated. Finally, the  
186 capacity to explain the GxC effect using the EFs drawn on in the clustering model was evaluated.

187  
188 2.5. *Genotype plus genotype by block of environments biplot*

189 Complementary principal component analyses were conducted on environment-centred genotype by  
190 environment, grouped by environment type, matrix (Cooper & DeLacy, 1994; Yan & Kang, 2003; Laffont  
191 et al., 2013). The Genotype plus Genotype by Block of environments biplot (GGB biplot) displays the  
192 genotype and genotype by block effects of a multi-environment trial dataset. The blocks of environments in  
193 our study were the environment types.

194 The GGB biplots can be interpreted such that genotypes located near the origin might either have all their  
195 values close to the environment means (low performance variation), given that the data were environment  
196 centred, or their variability is located in another dimension. Similarly, environments close to the origin may  
197 have little variability across genotypes or may not fit well in two dimensions. Genotypes that are close  
198 together have similar performance across environment types (Chapman et al., 1997). The R package *gge* was  
199 used for the analyses (Laffont et al., 2013).

200 The GGB biplots allowed us to assess genotypic performance across environment types. Their  
201 construction required a balanced multi-environmental trial dataset. Thus, two multi-environmental trials  
202 including varied years and locations were used in our study. The first multi-environmental trial (MET1)  
203 included seven cultivars from maturity group ‘00’ and 28 environments grouped in three environment types.  
204 The second multi-environment trial (MET2) included five cultivars from maturity group ‘000’ and 44  
205 environments grouped in three environment types.

206  
207 2.6. *Weighted and unweighted multi-environment trial data*

208 A MET dataset from 2018 was used that included ten ‘00’ cultivars that were tested in twelve locations.  
209 For the unweighted selection strategy, the mean performance by genotypes was calculated as the arithmetic  
210 mean of the performance in the twelve locations. The means of the ten genotypes were ranked from 1 to 10.  
211 For the weighted selection strategy, the mean performance was calculated – Equation 3 (E3) – (Podlich et  
212 al., 1999).

213 
$$w_k = \sum_{i=1}^E e_i \sum_{j=1}^{E_i} \frac{y_{kj}}{E_i} \quad (\text{E3})$$

214 where  $w_k$  is the weighted mean performance of the genotype  $k$ ,  $e_i$  is the frequency of occurrence of the  
215 environment type  $i$  in the TPE,  $y_{jk}$  is the phenotypic performance of the genotype  $k$  in environment  $j$ , and  $E_i$   
216 is the number of environments of type  $i$  in the MET.

217

### 218 **3. Results**

#### 219 *3.1. Eco-climatic factors set used for environmental clustering*

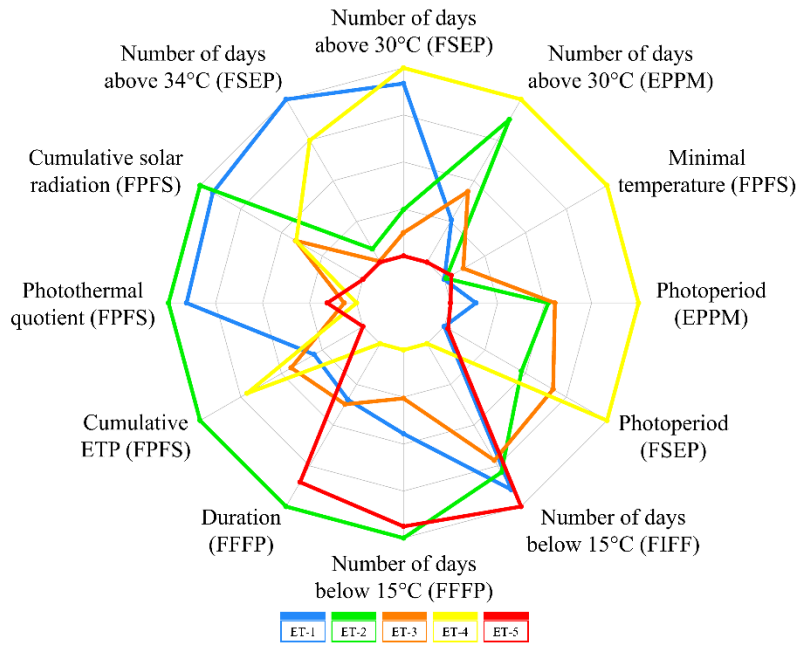
220 A total of 888 896 clustering models were tested. Each model used a unique combination of Eco-climatic  
221 Factors (EFs) that were identified as major Genotype by Environment Interactions (GxE) drivers in a previous  
222 study conducted by Elmerich et al. (2023). The best clustering model was identified based on the selection  
223 criteria detailed in *section 2.5*. In this model, 1.3% of the genotype yield variance was explained by the  
224 Genotype by Cluster Interactions (GxC) effect. The complete analyse of variance table can be found in  
225 supplementary table 2. Finally, the combination of EFs used in this clustering model allowed 88% of the  
226 GxC effect to be explained.

227 This model used a set of 12 EFs to define five clusters. The 12 EFs were: the number of days below 15°C  
228 from Flower Induction to First Flower (FIFF) and from the First Flower to First Pod (FFFP), the duration of  
229 the FFFP period, the photothermal quotient during FFFP, the minimal temperature from First Pod to First  
230 Seed (FPFS), the solar radiation quantity during FPFS, the evapotranspiration potential during FPFS, the  
231 photoperiod from First Seed to End of Pod (FSEP), the number of days above 30 and 34°C during FSEP, the  
232 photoperiod from End of Pod to Physiological Maturity (EPPM) and the number of days above 30°C during  
233 EPPM.

234

#### 235 *3.2. Characterisation of the five environment types*

236 The selected clustering model allowed the identification of five clusters that were referred to as  
237 environment types. The twelve EFs used for the clustering strongly discriminated the environment types (p-  
238 value < 0.001 for each EF) (Fig. 1).



239

240 **Fig. 1.** Characterisation of the five environment types by the 12 eco-climatic factors used for the clustering.  
 241 The phenophases are indicated in parenthesis with the following stages: Flower Induction (FI), First Flower  
 242 (FF), First Seed (FS), First Pod (FP), End Pod (EP) and Physiological Maturity (PM). The smaller inner  
 243 circle represents the minimum value and the outer cycle represents the maximal value for each eco-climatic  
 244 factor. The colours of the lines indicate the environment type: blue for the first environment type (ET-1),  
 245 green for the second environment type (ET-2), orange for the third environment type (ET-3), yellow for the  
 246 fourth environment type (ET-4) and red for the fifth environment type (ET-5).

247

248 As each environment was initially characterised by a set of 126 EFs, the latter were used to precisely  
 249 describe the environment types if a significant difference was observed ( $p$ -value  $< 0.05$ ) (Table 1). The first  
 250 environment type (ET-1) differed greatly in the high quantity of solar radiation (*ca.* 1088 MJ m<sup>-2</sup>),  
 251 precipitation (*ca.* 153 mm) and potential evapotranspiration (*ca.* 231 mm) as well as important cold stress  
 252 (*ca.* 26 days below 10°C) during the vegetative growth. The ET-1 reproductive growth conditions showed  
 253 low precipitation (*ca.* 82 mm) and ETP (*ca.* 187 mm) as well as high thermal amplitude (*ca.* 13°C) and heat  
 254 stress (*ca.* 12.7 days above 30°C). The second environment type (ET-2) was characterised by non-stressful  
 255 conditions during the vegetative growth followed, during the reproductive growth, by high water stress (*ca.*  
 256 0.47), solar radiations (*ca.* 1131 MJ m<sup>-2</sup>) and thermal amplitude (*ca.* 12.8 °C), with both chilling (*ca.* 17 days  
 257 below 15°C) and heat stress (*ca.* 7.3 days above 30°C) at the beginning and the end of the period respectively.  
 258 The third environment type (ET-3) was distinguishable by the absence of stressful or extreme conditions  
 259 throughout the crop cycle. The fourth environment type (ET-4) was denoted by the low precipitation (*ca.* 70  
 260 mm), warm temperature (*ca.* 18 °C) and short duration (*ca.* 39 days) of the vegetative growth. The ET-4

261 reproductive growth conditions were differentiated by high solar radiation intensity (*ca.* 23 MJ m<sup>-2</sup> d<sup>-1</sup>), high  
262 water (*ca.* 0.37) and severe heat stress (*ca.* 6 days above 34°C). The fifth environment type (ET-5) showed  
263 high precipitation (*ca.* 154 mm), solar radiation quantity (*ca.* 1084 MJ m<sup>-2</sup>), cold stress (*ca.* 27 days below  
264 10°C) and period duration (*ca.* 56 days) during the vegetative growth. The ET-5 reproductive growth was  
265 differentiated by the high precipitation (*ca.* 164 mm) and the low water stress index (*ca.* 0.11), the solar  
266 radiation intensity (*ca.* 18.1 MJ m<sup>-2</sup> d<sup>-1</sup>), the thermal amplitude (*ca.* 10.3 °C) and the heat stress (*ca.* 4 days  
267 above 30°C). Among the 602 environments used for the clustering, 83 (13.8%) belonged to ET-1, 106  
268 (17.6%) belonged to ET-2, 161 (26.7%) belonged to ET-3, 93 (15.4%) belonged to ET-4 and 159 (26.4%)  
269 belonged to ET-5.

**Table 1. Description of the five environment types identified by clustering.** For each Eco-climatic Factor (EF), the environment type effect was tested by Kruskal-Wallis non-parametric tests. When this effect was significant, a Dunn's multiple comparison test was performed in order to determine environment types (using Bonferroni adjusted *p*-values). In the table, yellow and blue shadings correspond to the highest and lowest EF means, respectively, that significantly differed from the others. The EFs written in bold correspond to those used for the k-means clustering. Abbreviations used in the table: evapotranspiration (ET), potential evapotranspiration (ETP), photothermal quotient (PTQ). The duration ranged from 5 to 52 days. The photoperiod ranged from 12.83 to 16.31 hours. The solar radiation intensity ranged from 5.3 to 29.7 MJ m<sup>2</sup> d<sup>-1</sup>. The solar radiation quantity ranged from 86 to 758 MJ m<sup>2</sup>. The PTQ ranged from 0.59 to 3.72. The precipitation amount ranged from 0 to 246 mm. The ET ranged from 2 to 141 mm. The ETP ranged from 8 to 154 mm. The water stress index ranged from 0 to 0.99. The minimum, average and maximum temperatures ranged from 0 to 22.9°C, 1.8 to 29.5°C and 3 to 37.1°C, respectively. The number of days with a minimal temperature below 10 or 15°C ranged from 0 to 50 days in both cases. The number of days with a minimal temperature above 30 or 34°C ranged from 0 to 18 days and 0 to 10 days, respectively. The thermal amplitude ranged from 2.5 to 18.7°C.

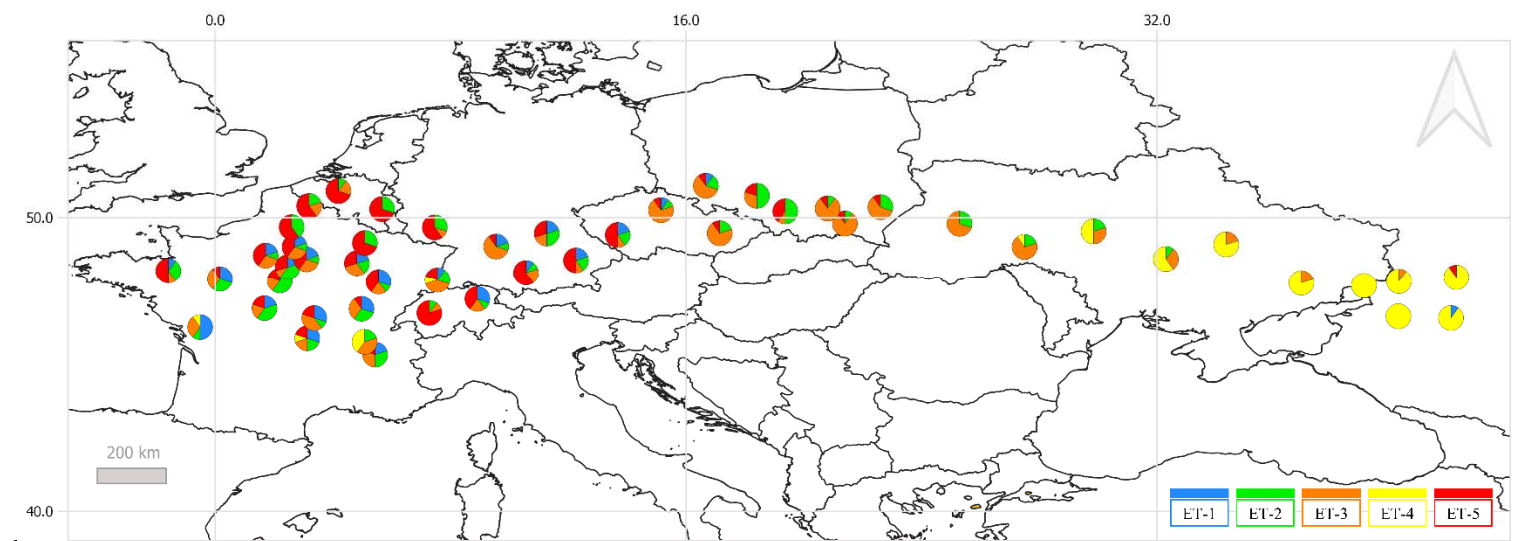
| Environment types | Variables            | Sowing - Emergence    | Emergence - Flower Induction | Flower Induction - First Flower | First Flower - First Pod | First Pod - First Seed | First Seed - End Pod | End Pod - Physiological Maturity | Overview  |
|-------------------|----------------------|-----------------------|------------------------------|---------------------------------|--------------------------|------------------------|----------------------|----------------------------------|---|
| ET-1              | Duration/Photoperiod |                       | Long period                  |                                 |                          | Short days             | Short days           |                                  | <i>Wet, cold and shaded conditions during vegetative growth. Dry, warm and sunny conditions during reproductive growth</i>        |
|                   | Solar radiation      | Quantity and PTQ      | Quantity                     |                                 | PTQ                      | Quantity               |                      |                                  |   |
|                   | Water availability   | Precipitation and ETP | Precipitation, ET and ETP    | ET                              | Precipitation            |                        | ETP                  | ETP                              |   |
|                   | Temperature min      | Cold stress           | Cold stress                  |                                 |                          | T° min                 |                      |                                  |   |
|                   | Temperature moy      |                       |                              |                                 | Amplitude                | Amplitude              | Amplitude            |                                  |   |
|                   | max                  | T° max                |                              | Heat stress                     | Heat stress              | Heat stress            |                      |                                  |   |
| ET-2              | Duration/Photoperiod |                       |                              |                                 | Long period              | Long period            | Short period         |                                  | <i>No stressful conditions during vegetative growth. Very dry, sunny and variable temperatures during the reproductive growth</i> |
|                   | Solar radiation      |                       |                              |                                 | PTQ                      | Quantity               | PTQ                  |                                  |   |
|                   | Water availability   |                       |                              |                                 | Precipitation            | ETP and Water stress   | Precipitation and ET | Precipitation and ET             |   |
|                   |                      |                       |                              |                                 | ETP and Water stress     |                        | Water stress         | Water stress                     |   |
|                   | Temperature min      |                       |                              |                                 | Chilling stress          | T° min                 |                      |                                  |   |
| Temperature moy   |                      |                       |                              |                                 | Amplitude                |                        | Amplitude            |                                  |   |
|                   | max                  |                       |                              |                                 |                          | Heat stress (34°C)     | Heat stress (30°C)   |                                  |   |
| ET-3              | Duration/Photoperiod |                       |                              |                                 |                          | Long days              |                      |                                  | <i>Low stressful conditions</i>   |
|                   | Solar radiation      |                       |                              | PTQ                             | PTQ                      |                        |                      |                                  |   |

|      |                      |                           |                           |                           |                        |                |                        |   |                      |
|------|----------------------|---------------------------|---------------------------|---------------------------|------------------------|----------------|------------------------|---|----------------------|
|      | Water availability   |                           |                           |                           |                        | ET             |                        | <i>throughout the crop cycle.</i>   |                      |
|      | min                  |                           |                           |                           | Cold stress            | T° min         |                        |   |                      |
|      | moy                  |                           | Amplitude                 |                           |                        |                |                        |   |                      |
|      | max                  |                           | Heat stress               |                           |                        | Heat stress    |                        |   |                      |
|      | Duration/Photoperiod | Short period              | Short period              | Short period              | Short period           | Long days      | Long days              | <i>Short, dry and warm conditions during the vegetative growth. Dry, hot and sunny during the reproductive growth</i>         |                      |
|      | Solar radiation      | Quantity and PTQ          | Quantity and PTQ          | Quantity and PTQ          | PTQ                    | Intensity, PTQ | Intensity              |   | Intensity            |
|      | Water availability   | Precipitation, ET and ETP | Precipitation, ET and ETP | Precipitation, ET and ETP | Precipitation          |                | Precipitation          |   | ET                   |
| ET-4 | min                  | T° min                    | T° min                    | Chilling stress           | Chilling stress        | T° min         | T° min                 |   | T° min               |
|      | moy                  | Cold stress               | Cold stress               |                           |                        |                |                        |   | T° average           |
|      | max                  | T° max                    | T° max and Heat stress    | T° max                    | T° max                 |                | T° max and Heat stress |   | Heat stress          |
|      |                      |                           |                           |                           |                        |                |                        |   |                      |
|      | Duration/Photoperiod | Long period               |                           | Long period               |                        | Short days     | Short days             | <i>Long, wet and cold conditions during vegetative growth. Wet, cool and shaded conditions during the reproductive growth</i> |                      |
|      | Solar radiation      | Intensity                 | Quantity                  |                           | PTQ                    | Quantity       | Intensity and PTQ      |   | Intensity            |
|      | Water availability   | Precipitation, ET and ETP | ET and ETP                | Precipitation             | Precipitation          | Precipitation  | Precipitation and ET   |   | Precipitation and ET |
| ET-5 | min                  | Cold stress               |                           | Chilling stress           | Chilling stress        | T° min         |                        |   | T° min               |
|      | moy                  |                           |                           |                           |                        |                | Amplitude              |   | Amplitude            |
|      | max                  | T° max                    |                           | T° max and Heat stress    | T° max and Heat stress |                | Heat stress            |   | Heat stress          |

224 3.3. Environment types occurrences in the European early soybean target population of environments

225 Fig. 2 displays the frequencies of environment types for the Dataset B (49 locations over 10 years). The  
226 first environment type (ET-1) (10.6% in Dataset B) occurred principally in France and Germany. The second  
227 environment type (ET-2) (19.1% in Dataset B) was present at least one year in all locations from France to  
228 the middle of Ukraine. The third environment type (ET-3) (28% in Dataset B) occurred at least one year in  
229 42 locations but showed high frequencies in locations situated in central Europe. The fourth environment  
230 type (ET-4) (16.7% in Dataset B) characterised mostly Eastern Europe locations and a few Southern France  
231 environments. The fifth environment type (ET-5) (25.5% in Dataset B) was present from France to Poland  
232 with high frequencies in Northern France, Belgium, Switzerland and Germany.

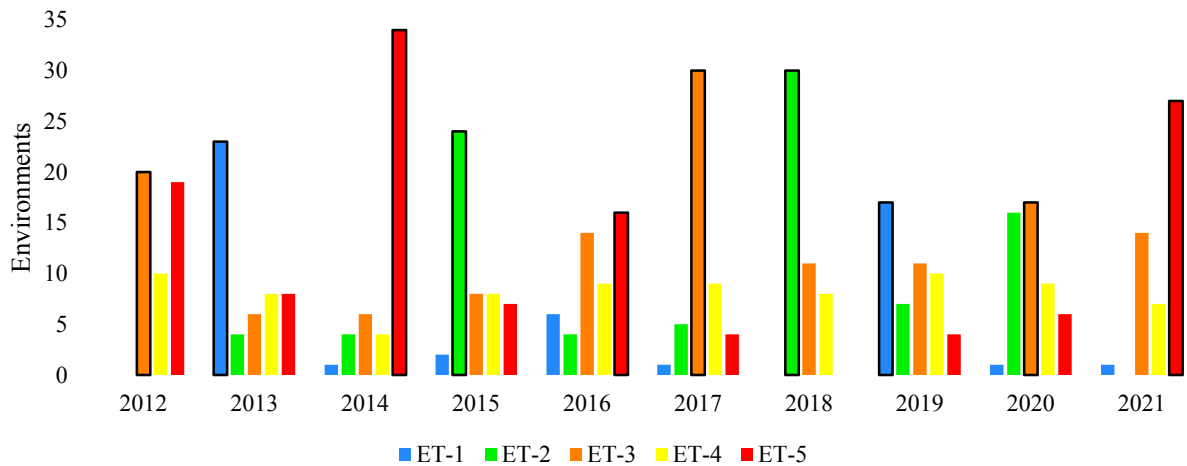
233 Moreover, around 12% of the TPE locations presented low environment types repeatability (*i.e.* most  
234 frequent environment type frequency < 33%), 53% presented intermediate environment types repeatability  
235 (*i.e.* 33% < most frequent environment type frequency ≤ 66%) and 35% presented high environment types  
236 repeatability (*i.e.* most frequent environment type frequency > 66%) (Couédel et al., 2021).



238 **Fig. 2.** Environment types frequency of occurrence from France to Russia (European early soybean  
239 production area) over ten years. For the 49 locations of Dataset B, the environment types' frequency of  
240 occurrence was calculated and represented on the map using pie charts. The colours of each pie section  
241 indicate the environment type: blue for the first environment type (ET-1), green for the second environment  
242 type (ET-2), orange for the third environment type (ET-3), yellow for the fourth environment type (ET-4)  
243 and red for the fifth environment type (ET-5).

244

245 The temporal distribution (2012-2021) of the five environment types was highly dependent on the  
 246 environment types considered (**Fig. 3**). In contrast to the other years, the years 2013, 2014, 2017, 2018  
 247 and 2020 were largely dominated by only one environment type. The ET-3 and ET-4 were present each year  
 248 (at least in 6 or 4 locations, respectively), while ET-1, ET-2 and ET-5 were not represented in 2, 2 and 1 year,  
 249 respectively.



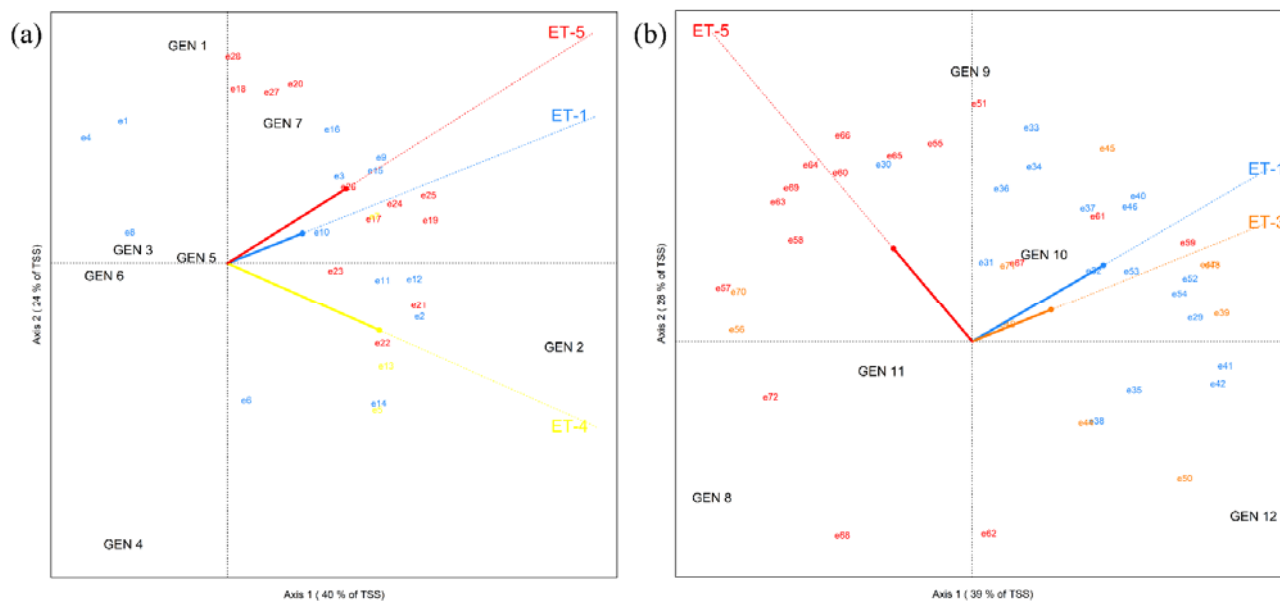
250

251 **Fig. 3.** Temporal distribution of the five environment types from 2012 to 2021 using the 49 locations in  
 252 Dataset B. The colours indicate the environment type: blue for the first environment type (ET-1), green for  
 253 the second environment type (ET-2), orange for the third environment type (ET-3), yellow for the fourth  
 254 environment type (ET-4) and red for the fifth environment type (ET-5).  
 255

### 256 3.4. Genotypic performance linked with envirotyping

257 Stratifying environments by environment types using Genotype plus Genotype by Block of environments  
 258 biplots (GGB biplot) enabled the interpretation of cultivar responses in the TPE. With the GGB biplot  
 259 representation cultivars can be compared by projecting a perpendicular from the genotype to the environment  
 260 type vector for any particular environment type. Cultivars that are further along in the positive direction of  
 261 the environment type vector have higher yields than other cultivars in said environment type, and vice versa.  
 262 In the multi-environmental trial 1 (MET1), which included seven cultivars and 28 environments grouped in  
 263 three environment types, the cultivar GEN2 was strongly positively correlated with ET-4 and positively  
 264 correlated with ET-1 and ET-5, and GEN4 was also positively correlated with ET-4 but negatively correlated  
 265 with ET-1 and ET-5 (Fig. 4a). In the multi-environmental trial 2 (MET2), which included five cultivars and  
 266 44 environments grouped in three environment types, the cultivars GEN8 and GEN11 were both negatively

267 correlated with all environment types whereas GEN9, GEN10 and GEN11 were positively correlated with  
 268 ET-1 but negatively correlated with ET-5 (Fig. 4b).



269

270 **Fig. 4.** Environment-standardised Genotype plus Genotype by Block of environments biplots illustrating  
 271 contrasted cultivars' performances depending on environment types in two multi-environmental trials from  
 272 the Dataset A. (a) The first multi-environmental trial included seven cultivars from maturity group '00'  
 273 (GEN 1 to GEN 7) and 28 environments (e1 to e28) grouped in three environment types (ET-1, ET-4 and ET-5).  
 274 (b) The second multi-environment trial included five cultivars from maturity group '000'  
 275 (GEN 8 to GEN 12) and 44 environments (e29 to e72) grouped in three environment types (ET-1, ET-3 and ET-5). The  
 276 percentages of the total genotype plus genotype by environment variation explained by the main two principal  
 277 components are shown in parentheses. The colours indicate the environment type: blue for the first (ET-1),  
 278 orange for the third (ET-3), yellow for the fourth (ET-4) and red for the fifth (ET-5).

279

### 280 3.5. MET-TPE alignment for MET data evaluation

281 In the 2018 Multi-Environment Trials (MET) dataset presented in Table 2, one environment belonged to  
 282 ET-1 (8.3%), seven belonged to ET-2 (58.3%), three belonged to ET-3 (25%), one belonged to ET-4 (8.3%)  
 283 and zero belonged to ET-5 (0%). This MET did not satisfactorily represent the frequencies of environment  
 284 types within the TPE (14, 18, 27, 15 and 26%, respectively for ET-1 to ET-5). Changes in cultivar ranks  
 285 between unweighted and weighted data were observed and ranged from no change (Genotype 3, 6 and 8) to  
 286 a six-rank position change (Genotype 9). If a breeder was allowed to keep only 10% of the tested cultivars  
 287 for future testing and/or deployment, his choices would differ depending on the strategy. For instance,  
 288 Genotype 4 would surely be advanced if data were weighted by the environment types frequencies observed  
 289 in the TPE but would most likely have been eliminated if data were unweighted.

**Table 2.** Example of the use of unweighted vs. weighted data analyses on the final breeding decisions. The Target Population of Environments (TPE) consists of five types of environments with a frequency of occurrence of 0.14, 0.18, 0.27, 0.15 and 0.26, respectively. Ten genotypes were evaluated across 12 environments within a Multi-Environment Trial (MET) included in the Dataset A. In this MET, the first environment type (ET-1) has been sampled in the MET environment 1, the second environment type (ET-2) has been sampled in MET environments 2 to 8, the third environment type (ET-3) has been sampled in MET environments 9 to 11, the fourth environment type (ET-4) has been sampled in MET environment 12 and the fifth environment type (ET-5) has not been sampled in the MET. The grain yield performance of the genotypes is expressed in quintals by hectare. The estimated mean yield performance and rank are given for each genotype.

| Env. in MET | Environment Types in MET              |      |      |                                       |      |      |      |      |                                       |      |      |                                       | Data analyses |            |          |           |
|-------------|---------------------------------------|------|------|---------------------------------------|------|------|------|------|---------------------------------------|------|------|---------------------------------------|---------------|------------|----------|-----------|
|             | ET-1<br>( $e_1 = 0.14$ ) <sup>1</sup> |      |      | ET-2<br>( $e_2 = 0.18$ ) <sup>1</sup> |      |      |      |      | ET-3<br>( $e_3 = 0.27$ ) <sup>1</sup> |      |      | ET-4<br>( $e_4 = 0.15$ ) <sup>1</sup> |               | Unweighted | Weighted |           |
|             | 1                                     | 2    | 3    | 4                                     | 5    | 6    | 7    | 8    | 9                                     | 10   | 11   | 12                                    | Mean          | Rank       | Mean     | Rank      |
| Genotype 1  | 51.1                                  | 28.4 | 46.4 | 25.4                                  | 30.2 | 34.1 | 43.7 | 33.6 | 41.7                                  | 34.9 | 38.6 | 37.7                                  | 37.2          | <b>9</b>   | 29.2     | <b>6</b>  |
| Genotype 2  | 50.3                                  | 40.0 | 44.1 | 24.0                                  | 27.5 | 36.2 | 48.8 | 30.6 | 44.5                                  | 38.8 | 42.2 | 34.6                                  | 38.5          | <b>4</b>   | 29.8     | <b>3</b>  |
| Genotype 3  | 41.1                                  | 28.7 | 43.2 | 20.9                                  | 30.4 | 36.6 | 43.1 | 31.5 | 45.3                                  | 37.4 | 32.1 | 37.1                                  | 35.6          | <b>10</b>  | 27.5     | <b>10</b> |
| Genotype 4  | 52.5                                  | 30.5 | 47.0 | 24.6                                  | 34.5 | 30.4 | 45.3 | 33.1 | 43.1                                  | 42.0 | 31.7 | 40.8                                  | 38.0          | <b>6</b>   | 30.1     | <b>2</b>  |
| Genotype 5  | 49.1                                  | 30.5 | 48.0 | 23.0                                  | 32.2 | 39.7 | 46.8 | 32.4 | 42.2                                  | 35.7 | 40.4 | 36.1                                  | 38.0          | <b>5</b>   | 29.2     | <b>5</b>  |
| Genotype 6  | 54.5                                  | 27.4 | 45.2 | 25.4                                  | 36.2 | 42.2 | 47.9 | 33.2 | 42.9                                  | 38.0 | 43.1 | 39.0                                  | 39.6          | <b>1</b>   | 31.0     | <b>1</b>  |
| Genotype 7  | 52.4                                  | 33.7 | 52.8 | 24.5                                  | 34.5 | 35.8 | 46.8 | 32.9 | 35.7                                  | 36.3 | 42.2 | 37.1                                  | 38.7          | <b>3</b>   | 29.7     | <b>4</b>  |
| Genotype 8  | 48.0                                  | 33.3 | 51.5 | 24.3                                  | 29.7 | 31.2 | 45.1 | 33.6 | 46.5                                  | 37.3 | 36.4 | 36.1                                  | 37.8          | <b>7</b>   | 29.1     | <b>7</b>  |
| Genotype 9  | 42.0                                  | 35.2 | 54.8 | 25.6                                  | 31.2 | 41.4 | 48.4 | 33.9 | 43.9                                  | 39.6 | 38.9 | 35.0                                  | 39.2          | <b>2</b>   | 28.9     | <b>8</b>  |
| Genotype 10 | 41.0                                  | 31.4 | 48.9 | 25.2                                  | 31.5 | 38.3 | 44.9 | 33.2 | 38.3                                  | 39.0 | 40.5 | 38.4                                  | 37.6          | <b>8</b>   | 28.4     | <b>9</b>  |

<sup>1</sup>Frequency of each environment type in the TPE

302

## 303 **4. Discussion**

### 304 *4.1. Five environment types characterised the European early soybean target population of environments*

305 The environmental clustering using twelve Eco-climatic Factors (EFs) that contrasted in terms of climatic  
 306 variables and phenophases explained 88% of the cultivar yield variation attributed to Genotype by Cluster  
 307 interaction (GxC). Thanks to the clustering, breeders will be able to minimise the GxC variance in their  
 308 prediction models and focus more on the genetic variation *per se* (Messina et al., 2018). The highly significant  
 309 and large GxC variance component compared to the genotype and cluster effects demonstrated the interest  
 310 in defining environment types within the European early maturity soybean production area. It would be  
 311 interesting to confront this model performance with a richer database.

312 The results highlighted five environment types that represent the European early soybean TPE. The first  
 313 environment type (ET-1) was characterised by wet, cold and shaded conditions during the vegetative growth

314 and warmth and high radiation but no drought conditions during the reproductive growth. The second  
315 environment type (ET-2) showed no stressful conditions during the vegetative growth but showed drought  
316 stress, high radiation and variable temperatures during the reproductive growth. The third environment type  
317 (ET-3) was distinguishable by the absence of stressful or extreme conditions throughout the crop cycle. The  
318 fourth environment type (ET-4) was marked by the short, dry and warm conditions during the vegetative  
319 growth as well as by the dry, hot and high radiation conditions during the reproductive growth. The fifth  
320 environment type (ET-5) was characterised by long, wet and cold conditions during the vegetative growth as  
321 well as wet, cool and shaded conditions during the reproductive growth. The environment types could be  
322 referred to as follows: ‘Temperate Humid’, ‘Temperate dry’, ‘Temperate with no stress’, ‘Warm dry’ and  
323 ‘Cool humid’, respectively (IPCC, 1997). The different environment types were primarily contrasted by the  
324 intensity and timing of stresses relative to temperatures (cold stress during the vegetative growth and heat  
325 stress during the reproductive growth) and water availability (precipitation amount, evapotranspiration and  
326 drought). Despite disconnected/juxtaposed geographical areas, our classification is consistent with the  
327 classification established by Schoving et al. (2022) that identified environment classes on the basis of levels  
328 of temperatures and drought stresses. Our non-a priori approach allowed us to identify the combinations of  
329 stresses and conditions to which soybean cultivars are expected to similarly respond. This envirotyping  
330 approach allowed to define the first European typology of environments for soybean and can be used to  
331 design new trial networks or improve existing ones.

332

#### 333 4.2. *Varied occurrences of environment types in the target population of environments*

334 The results demonstrated that the environment types distribution in the TPE over ten years (from 2012 to  
335 2021) was heterogenous and lack generalities. Contrary to the six mega-environments defined by Kurasch et  
336 al. (2017), in the early soybean TPE, environment types’ occurrences neither followed annual (one year  
337 corresponding to one environment type) nor geographical (longitudinal or latitudinal) distributions. This  
338 difference could be attributed to our wider study area and to the environmental descriptors that were based  
339 on maturity dates in their study as opposed to EFs. The ET-1 showed low repeatability, ET-2 and ET-5  
340 showed intermediate repeatability, the ET-4 showed high repeatability and the ET-3 equally showed  
341 intermediate and high repeatability. The identification of highly repeatable environment types (mainly

342 located in Eastern Europe) is key to designing breeding strategy for specific adaptation. In environments  
343 where specific adaptation appears best suited, the interactions between genotypes and other environmental  
344 characteristics, such as soil properties, cultural practices or biotic stress incidences could be investigated. On  
345 the other hand, the highlight of low environment types repeatability in Western Europe will help breeders  
346 aim for broad adaptation, as future growth conditions are highly unpredictable (de la Vega & Chapman,  
347 2006). The broad adaptation of cultivars could be reached by designing more plastic cultivars, *i.e.* cultivars  
348 having traits with a large reaction norm, leading to a stable performance under a large set of growing  
349 conditions (Sambatti & Caylor, 2007). Future work should evaluate soybean above and below ground traits  
350 plasticity in response to the main EFs driving GEI to identify relevant traits to focus on for broad adaptation.  
351 Among these, cold stress during the early vegetative growth has never been investigated and seems of great  
352 importance (Elmerich et al., 2023).

353 Variations in the spatial and temporal distribution of the environment types across the TPE revealed the  
354 complexity to design efficient breeding programs that will control possible genotype by environment  
355 interactions and select for climate adaptation as much as. In the context of climate change, these five  
356 environment types could be conserved but their spatial and temporal distributions could be affected. Thus,  
357 monitoring the environment types distribution will be essential. The perspectives will be to simulate virtual  
358 environments using different forecasting climate scenarios, *e.g.* SSP5-8.5 and SSP2-4.5 (IPCC, 2022) and  
359 confront our results to a projected environment types distribution (Bustos-Korts et al., 2016).

360

#### 361 4.3. *Environmental characterisation, an informative tool for breeding decisions*

362 The environmental characterisation aimed to increase the understanding of GEI effects. This is a crucial  
363 step when attempting to expand a new crop northward, and to assist breeders dealing with GEI (Chenu,  
364 2015). In our study, the GEI effect, was larger for yield, compared with the genotype effect. The same ranking  
365 was previously reported in the literature (Dreccer et al., 2007; Chenu et al., 2011).

366 Using METs conducted in France, differences in genotypes performances were observed across  
367 environment types. Thanks to the environmental characterisation of the European early soybean TPE,  
368 predictions can be made on genotypes performances in non-tested locations that belong to the same  
369 environment type. Better precision in these predictions could be achieved especially in Eastern Europe where

370 specific adaptation appears to be best suited. For instance, GEN 2 showed its best performance in ET-4;  
371 therefore, it could be advised for use in Eastern Europe or used as a parent to develop bi-parental populations.  
372 Envirotyping could increase performances predictions for new targeted areas in locations where trials have  
373 not been conducted. The environment types frequencies within the TPE constitute an interesting result to  
374 position current elite soybean cultivars on the market, according to their performance in each environment  
375 type. It could also be used to design new cultivars specifically adapted to an environment type. This approach  
376 allows the crop yield potential to be maximised for a targeted region and to match farmers' needs for  
377 improved and locally designed germplasm (Annicchiarico, 2021).

378 A huge challenge when analysing METs for breeding purposes is evaluating how METs data match the  
379 TPE in order to improve breeding decisions and enhance the genetic gain over generations. Comparisons  
380 between weighted and unweighted MET phenotypic data support the importance of MET-TPE alignment  
381 evaluation (Podlich et al., 1999; Cooper et al., 2023). Our results confirmed the efficiency of this approach  
382 when climatic scenarios at the MET scale derived from the TPE (Table 2). When dealing with large and  
383 complex GEI effects, as observed in the Western European soybean production area, breeders should weight  
384 their phenotypic data and consider the MET-TPE alignment, particularly if they are selecting for broad  
385 adaptation.

386

## 387 **5. Conclusion**

388 This study is the first to European described early soybean Target Population of Environments (TPE) to  
389 control Genotype by Environment Interactions (GEI) for breeding purposes. Environmental characterisation  
390 based on GEI-drivers is a powerful tool to tackle GEI and accelerate soybean genetic gain. Our method  
391 explained a large part (88%) of the yield variation that is due to the GEI effects. The five environment types  
392 observed across Europe and their repeatability varied in terms of geographical and temporal distributions.  
393 The results underlined the necessity to combine specific (maximising cultivars performances in a given  
394 environment type) and broad adaptations (maximising cultivars yield stabilities over different environment  
395 types) depending on locations. Envirotyping represents a necessary progression within the era of data-driven  
396 plant improvement (Xu, 2016; Crossa et al., 2021). These approaches allowed us to better predict genotypic  
397 performances within specific environment types and improve the positioning of cultivars in the marketplace.

398 Moreover, as demonstrated in this study, a better MET data interpretation and utilisation can be reached  
399 through MET-TPE alignment evaluation. To enhance the deployment of soybean northward, perspectives  
400 should be to design ideotypes that outperform the existing germplasm for specific or multiple environment  
401 types across the TPE.

402

### 403 **Acknowledgements**

404 The authors would like to thanks Roman FERRANT for his comments on the adequacy between our  
405 observations and his knowledge and field experience. The authors are thankful to Johanna YOUNOUS for  
406 her advises on statistical analysis. The authors are grateful to H el ene FREROT and Philippe DEBAEKE for  
407 their insightful comments on the ecophysiological modelling. The ‘‘SOJA Terres Inovia-GEVES-  
408 Partenaires’’ are acknowledged for the open access trial dataset.

409

### 410 **Authors contribution**

411 CE, GB and BL designed the study. CE, MG and BL conducted the data analysis. CE, GB, PJ, MG, MPF  
412 and BL wrote the manuscript. All authors read, commented on and approved the final version.

413

### 414 **References**

415 Annicchiarico, P. (2021). Breeding gain from exploitation of regional adaptation: An alfalfa case study.  
416 *Crop Science*, 61(4), 2254–2271. <https://doi.org/10.1002/csc2.20423>

417 Ayerdi Gotor, A., & Marraccini, E. (2022). Innovative Pulses for Western European Temperate Regions:  
418 A Review. *Agronomy*, 12(1), Article 1. <https://doi.org/10.3390/agronomy12010170>

419 Basford, K. E., & Cooper, M. (1998). Genotype×environment interactions and some considerations of  
420 their implications for wheat breeding in Australia This review is one of a series commissioned by the  
421 Advisory Committee of the Journal. *Australian Journal of Agricultural Research*, 49(2), 153–174.  
422 <https://doi.org/10.1071/a97035>

423 Boote, K. J., Jones, J. W., Hoogenboom, G., & Pickering, N. B. (1998). The CROPGRO model for grain  
424 legumes. In G. Y. Tsuji, G. Hoogenboom, & P. K. Thornton (Eds.), *Understanding Options for Agricultural*  
425 *Production* (Vol. 7, pp. 99–128). Springer Netherlands. [https://doi.org/10.1007/978-94-017-3624-4\\_6](https://doi.org/10.1007/978-94-017-3624-4_6)

426 Brennan, P. S., Byth, D. E., Drake, D. W., Lacy, Ihd., & Butler, D. G. (1981). Determination of the  
427 location and number of test environments for a wheat cultivar evaluation program. *Australian Journal of*  
428 *Agricultural Research*, 32(2), 189–201. <https://doi.org/10.1071/ar9810189>

429 Bustos-Korts, D., Boer, M. P., Layton, J., Gehringer, A., Tang, T., Wehrens, R., Messina, C., de la Vega,  
430 A. J., & van Eeuwijk, F. A. (2022). Identification of environment types and adaptation zones with self-  
431 organizing maps; applications to sunflower multi-environment data in Europe. *Theoretical and Applied*  
432 *Genetics*. <https://doi.org/10.1007/s00122-022-04098-9>

433 Bustos-Korts, D., Malosetti, M., Chapman, S., & van Eeuwijk, F. (2016). Chapitre 3—Modelling of  
434 Genotype by Environment Interaction and Prediction of Complex Traits across Multiple Environments as a  
435 Synthesis of Crop Growth Modelling, Genetics and Statistics. In X. Yin & P. C. Struik (Eds.), *Crop Systems*  
436 *Biology* (pp. 55–82). Springer International Publishing. [https://doi.org/10.1007/978-3-319-20562-5\\_3](https://doi.org/10.1007/978-3-319-20562-5_3)

437 Casadebaig, P., Gauffreteau, A., Landré, A., Langlade, N., Mestries, E., Sarron, J., Trépos, R., Vincourt,  
438 P., & Debaeke, P. (2021). *Optimized cultivar deployment improves the efficiency and stability of sunflower*  
439 *crop production at national scale*. <https://doi.org/10.1101/2020.09.21.306076>

440 Chapman, S. C., Cooper, M., Hammer, G. L., & Butler, D. G. (2000). Genotype by environment  
441 interactions affecting grain sorghum. II. Frequencies of different seasonal patterns of drought stress are  
442 related to location effects on hybrid yields. *Australian Journal of Agricultural Research*, 51(2), 209–222.  
443 <https://doi.org/10.1071/ar99021>

444 Chapman, S. C., Edmeades, G. O., & Crossa, J. (1997). Genotype by environment effects and selection  
445 for drought tolerance in tropical maize. I. Two mode pattern analysis of yield | SpringerLink. *Euphytica*, 95,  
446 1–9.

447 Chauhan, Y. S., Solomon, K. F., & Rodriguez, D. (2013). Characterization of north-eastern Australian  
448 environments using APSIM for increasing rainfed maize production. *Field Crops Research*, 144, 245–255.  
449 <https://doi.org/10.1016/j.fcr.2013.01.018>

450 Chenu, K. (2015). Chapter 13—Characterizing the crop environment – nature, significance and  
451 applications. In V. O. Sadras & D. F. Calderini (Eds.), *Crop Physiology (Second Edition)* (pp. 321–348).  
452 Academic Press. <https://doi.org/10.1016/B978-0-12-417104-6.00013-3>

453 Chenu, K., Cooper, M., Hammer, G. L., Mathews, K. L., Dreccer, M. F., & Chapman, S. C. (2011).  
454 Environment characterization as an aid to wheat improvement: Interpreting genotype–environment  
455 interactions by modelling water-deficit patterns in North-Eastern Australia. *Journal of Experimental Botany*,  
456 *62*(6), 1743–1755. <https://doi.org/10.1093/jxb/erq459>

457 Chenu, K., Dehifard, R., & Chapman, S. C. (2013). Large-scale characterization of drought pattern: A  
458 continent-wide modelling approach applied to the Australian wheatbelt – spatial and temporal trends. *New*  
459 *Phytologist*, *198*(3), 801–820. <https://doi.org/10.1111/nph.12192>

460 Chenu, K., Porter, J. R., Martre, P., Basso, B., Chapman, S. C., Ewert, F., Bindi, M., & Asseng, S. (2017).  
461 Contribution of Crop Models to Adaptation in Wheat. *Trends in Plant Science*, *22*(6), 472–490.  
462 <https://doi.org/10.1016/j.tplants.2017.02.003>

463 Cooper, M., & Byth, D. E. (1996). Understanding plant adaptation to achieve systematic applied crop  
464 improvement—A fundamental challenge. In *Plant Adaptation and Crop Improvement* (M. Cooper and G. L.  
465 Hammer, pp. 5–23). IRRI.

466 Cooper, M., & DeLacy, I. H. (1994). Relationships among analytical methods used to study genotypic  
467 variation and genotype-by-environment interaction in plant breeding multi-environment experiments.  
468 *Theoretical and Applied Genetics*, *88*(5), 561–572. <https://doi.org/10.1007/BF01240919>

469 Cooper, M., Powell, O., Gho, C., Tang, T., & Messina, C. (2023). Extending the breeder’s equation to  
470 take aim at the target population of environments. *Frontiers in Plant Science*, *14*.  
471 <https://www.frontiersin.org/articles/10.3389/fpls.2023.1129591>

472 Corlouer, E., Gauffreteau, A., Bouchet, A.-S., Bissuel-Bélaygue, C., Nesi, N., & Laperche, A. (2019).  
473 Envirotypes Based on Seed Yield Limiting Factors Allow to Tackle  $G \times E$  Interactions. *Agronomy*, *9*(12),  
474 Article 12. <https://doi.org/10.3390/agronomy9120798>

475 Couëdel, A., Edreira, J. I. R., Pisa Lollato, R., Archontoulis, S., Sadras, V., & Grassini, P. (2021).  
476 Assessing environment types for maize, soybean, and wheat in the United States as determined by spatio-  
477 temporal variation in drought and heat stress. *Agricultural and Forest Meteorology*, *307*, 108513.  
478 <https://doi.org/10.1016/j.agrformet.2021.108513>

479 Crossa, J., Fritsche-Neto, R., Montesinos-Lopez, O. A., Costa-Neto, G., Dreisigacker, S., Montesinos-  
480 Lopez, A., & Bentley, A. R. (2021). The Modern Plant Breeding Triangle: Optimizing the Use of Genomics,

481 Phenomics, and Enviromics Data. *Frontiers in Plant Science*, 12.  
482 <https://www.frontiersin.org/article/10.3389/fpls.2021.651480>

483 de la Vega, A. J., & Chapman, S. C. (2006). Defining Sunflower Selection Strategies for a Highly  
484 Heterogeneous Target Population of Environments. *Crop Science*, 46(1), 136.  
485 <https://doi.org/10.2135/cropsci2005.0170>

486 Döttinger, C. A., Hahn, V., Leiser, W. L., & Würschum, T. (2023). Do We Need to Breed for Regional  
487 Adaptation in Soybean? Evaluation of Genotype-by-Location Interaction and Trait Stability of Soybean in  
488 Germany. *Plants*, 12(4), Article 4. <https://doi.org/10.3390/plants12040756>

489 Dreccer, M. F., Borgognone, M. G., Ogonnaya, F. C., Trethowan, R. M., & Winter, B. (2007).  
490 CIMMYT-selected derived synthetic bread wheats for rainfed environments: Yield evaluation in Mexico and  
491 Australia. *Field Crops Research*, 100(2), 218–228. <https://doi.org/10.1016/j.fcr.2006.07.005>

492 Elmerich, C., Boulch, G., Faucon, M.-P., Lakhali, L., & Lange, B. (2023). Identification of Eco-Climatic  
493 Factors Driving Yields and Genotype by Environment Interactions for Yield in Early Maturity Soybean  
494 Using Crop Simulation. *Agronomy*, 13(2), 322. <https://doi.org/10.3390/agronomy13020322>

495 FAOSTAT. (2021). *Food and Agriculture Organization Corporate Statistical Database*.  
496 <https://www.fao.org/faostat/en/#data/QCL>

497 Holland, J. B., Nyquist, W. E., & Cervantes-Martínez, C. T. (2003). Estimating and interpreting  
498 heritability for plant breeding: An update. In *Plant breeding reviews* (Jules Janick, Vol. 22). John Wiley &  
499 Sons.

500 Hoogenboom, G., Porter, C. H., Boote, K. J., Shelia, V., Wilkens, P. W., Singh, U., White, J. W., Asseng,  
501 S., Lizaso, J. I., Moreno, L. P., Pavan, W., Ogoshi, R., Hunt, L. A., Tsuji, G. Y., & Jones, J. W. (2019). The  
502 DSSAT crop modeling ecosystem. In *Advances in Crop Modeling for a Sustainable Agriculture* (pp. 173–  
503 216). Burleigh Dodds Science Publishing. <https://doi.org/10.19103/AS.2019.0061.10>

504 IPCC. (1997). *Revised 1996 IPCC guidelines for national greenhouse gas inventories: Reference Manual*  
505 *(Volume 3)* (Houghton J.T., Meira L.G., Filho L.G., Lim B., Treanton K., Mamaty I., Bonduki Y., Griggs  
506 D.J. and Callender B.A.). Intergovernmental Panel on Climate Change.

507 IPCC. (2022). *Climate Change 2022: Impacts, Adaptation, and Vulnerability*. (Contribution of Working  
508 Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change).

509 Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., Wilkens, P. W.,  
510 Singh, U., Gijsman, A. J., & Ritchie, J. T. (2003). The DSSAT cropping system model. *European Journal*  
511 *of Agronomy*, *18*(3–4), 235–265. [https://doi.org/10.1016/S1161-0301\(02\)00107-7](https://doi.org/10.1016/S1161-0301(02)00107-7)

512 Karges, K., Bellingrath-Kimura, S. D., Watson, C. A., Stoddard, F. L., Halwani, M., & Reckling, M.  
513 (2022). Agro-economic prospects for expanding soybean production beyond its current northerly limit in  
514 Europe. *European Journal of Agronomy*, *133*, 126415. <https://doi.org/10.1016/j.eja.2021.126415>

515 Kurasch, A. K., Hahn, V., Leiser, W. L., Vollmann, J., Schori, A., Bétrix, C.-A., Mayr, B., Winkler, J.,  
516 Mechtler, K., Aper, J., Sudaric, A., Pejic, I., Sarcevic, H., Jeanson, P., Balko, C., Signor, M., Miceli, F.,  
517 Strijk, P., Rietman, H., ... Würschum, T. (2017). Identification of mega-environments in Europe and effect  
518 of allelic variation at maturity *E* loci on adaptation of European soybean: Soybean Adaptation to Europe.  
519 *Plant, Cell & Environment*, *40*(5), 765–778. <https://doi.org/10.1111/pce.12896>

520 Laffont, J.-L., Wright, K., & Hanafi, M. (2013). Genotype Plus Genotype × Block of Environments  
521 Biplots. *Crop Science*, *53*(6), 2332–2341. <https://doi.org/10.2135/cropsci2013.03.0178>

522 Marraccini, E., Gotor, A. A., Scheurer, O., & Leclercq, C. (2020). An Innovative Land Suitability Method  
523 to Assess the Potential for the Introduction of a New Crop at a Regional Level. *Agronomy*, *10*(3), 1–24.  
524 <https://doi.org/10.3390/agronomy10030330>

525 Messina, C. D., Technow, F., Tang, T., Totir, R., Gho, C., & Cooper, M. (2018). Leveraging biological  
526 insight and environmental variation to improve phenotypic prediction: Integrating crop growth models  
527 (CGM) with whole genome prediction (WGP). *European Journal of Agronomy*, *100*, 151–162.  
528 <https://doi.org/10.1016/j.eja.2018.01.007>

529 Millet, E. J., Kruijer, W., Coupel-Ledru, A., Alvarez Prado, S., Cabrera-Bosquet, L., Lacube, S.,  
530 Charcosset, A., Welcker, C., van Eeuwijk, F., & Tardieu, F. (2019). Genomic prediction of maize yield across  
531 European environmental conditions. *Nature Genetics*, *51*(6), 952–956. [https://doi.org/10.1038/s41588-019-](https://doi.org/10.1038/s41588-019-0414-y)  
532 [0414-y](https://doi.org/10.1038/s41588-019-0414-y)

533 Podlich, D. W., & Cooper, M. (1998). QU-GENE: A simulation platform for quantitative analysis of  
534 genetic models. *Bioinformatics*, *14*(7), 632–653. <https://doi.org/10.1093/bioinformatics/14.7.632>

535 Podlich, D. W., Cooper, M., Basford, K. E., & Geiger, H. H. (1999). Computer simulation of a selection  
536 strategy to accommodate genotype environment interactions in a wheat recurrent selection programme. *Plant*  
537 *Breeding*, 118(1), 17–28. <https://doi.org/10.1046/j.1439-0523.1999.118001017.x>

538 Rincent, R., Malosetti, M., Ababaei, B., Touzy, G., Mini, A., Bogard, M., Martre, P., Le Gouis, J., & van  
539 Eeuwijk, F. (2019). Using crop growth model stress covariates and AMMI decomposition to better predict  
540 genotype-by-environment interactions. *Theoretical and Applied Genetics*, 132(12), 3399–3411.  
541 <https://doi.org/10.1007/s00122-019-03432-y>

542 Sambatti, J. B. M., & Caylor, K. K. (2007). When is breeding for drought tolerance optimal if drought is  
543 random? *New Phytologist*, 175(1), 70–80. <https://doi.org/10.1111/j.1469-8137.2007.02067.x>

544 Schoving, C., Champolivier, L., Maury, P., & Debaeke, P. (2022). Combining multi-environmental trials  
545 and crop simulation to understand soybean response to early sowings under contrasting water conditions.  
546 *European Journal of Agronomy*, 133, 126439. <https://doi.org/10.1016/j.eja.2021.126439>

547 Sudarić, A. (2020). *Soybean for Human Consumption and Animal Feed*. IntechOpen.

548 Woollings, T. (2010). Dynamical influences on European climate: An uncertain future. *Philosophical*  
549 *Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1924), 3733–  
550 3756. <https://doi.org/10.1098/rsta.2010.0040>

551 Xu, Y. (2016). Envirotyping for deciphering environmental impacts on crop plants. *Theoretical and*  
552 *Applied Genetics*, 129(4), 653–673. <https://doi.org/10.1007/s00122-016-2691-5>

553 Yan, W., & Kang, M. S. (2003). *GGE biplot analysis: A graphical tool for breeders, geneticists, and*  
554 *agronomists*. CRC Press.

555