



**HAL**  
open science

# Impact of coupled input data source-resolution and aggregation on contributions of high-yielding traits to simulated wheat yield

Ehsan Eyshi Rezaei, Babacar Faye, Frank Ewert, Senthold Asseng, Pierre Martre, Heidi Webber

## ► To cite this version:

Ehsan Eyshi Rezaei, Babacar Faye, Frank Ewert, Senthold Asseng, Pierre Martre, et al.. Impact of coupled input data source-resolution and aggregation on contributions of high-yielding traits to simulated wheat yield. *Scientific Reports*, 2024, 14 (1), pp.23172. 10.1038/s41598-024-74309-4. hal-04802905

**HAL Id: hal-04802905**

**<https://hal.inrae.fr/hal-04802905v1>**

Submitted on 25 Nov 2024

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

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



Distributed under a Creative Commons Attribution 4.0 International License



## OPEN Impact of coupled input data source-resolution and aggregation on contributions of high-yielding traits to simulated wheat yield

Ehsan Eyshi Rezaei<sup>1,7</sup>✉, Babacar Faye<sup>1,2,7</sup>, Frank Ewert<sup>1,3</sup>, Senthold Asseng<sup>4</sup>, Pierre Martre<sup>5</sup> & Heidi Webber<sup>1,6</sup>

High-yielding traits can potentially improve yield performance under climate change. However, data for these traits are limited to specific field sites. Despite this limitation, field-scale calibrated crop models for high-yielding traits are being applied over large scales using gridded weather and soil datasets. This study investigates the implications of this practice. The SIMPLACE modeling platform was applied using field, 1 km, 25 km, and 50 km input data resolution and sources, with 1881 combinations of three traits [radiation use efficiency (RUE), light extinction coefficient (K), and fruiting efficiency (FE)] for the period 2001–2010 across Germany. Simulations at the grid level were aggregated to the administrative units, enabling the quantification of the aggregation effect. The simulated yield increased by between 1.4 and 3.1 t ha<sup>-1</sup> with a maximum RUE trait value, compared to a control cultivar. No significant yield improvement (< 0.4 t ha<sup>-1</sup>) was observed with increases in K and FE alone. Utilizing field-scale input data showed the greatest yield improvement per unit increment in RUE. Resolution of water related inputs (soil characteristics and precipitation) had a notably higher impact on simulated yield than of temperature. However, it did not alter the effects of high-yielding traits on yield. Simulated yields were only slightly affected by data aggregation for the different trait combinations. Warm-dry conditions diminished the benefits of high-yielding traits, suggesting that benefits from high-yielding traits depend on environments. The current findings emphasize the critical role of input data resolution and source in quantifying a large-scale impact of high-yielding traits.

Global food demand is estimated to increase by 62% by 2050 to meet the changing consumption needs of an expanding and more affluent population<sup>1</sup>. Achieving such a significant increase in food demand is not only challenging but also threatened by climate change<sup>2,3</sup>. The rising frequency and intensity of droughts, heatwaves, and floods observed worldwide in recent years have negatively impacted crop yields and are projected to persist<sup>4,5</sup>. Global wheat yield is projected to increase between 9% and 18%, depending on the greenhouse gas emission scenarios<sup>6</sup>. However, wheat yield has stagnated or even declined in recent decades across various regions<sup>7</sup>. A 1.88% per year yield gain is required to meet the projected demand to increase wheat production<sup>8</sup> which is currently varying between 0.5% and 1%<sup>9</sup>. Therefore, the development of climate-resilient and high-yielding cereal cultivars is essential to support future food security<sup>10</sup>.

Higher radiation use efficiency (RUE) and fruiting efficiency, homogenous light distribution across the canopy (using light extinction coefficient), and larger grain weight potential were defined as crucial traits that could improve wheat yield<sup>11</sup>. The majority of long-term wheat yield improvements are linked to optimizing the harvest index for greater carbon partitioning to yield without significantly altering total biomass accumulation<sup>12</sup>. Harvest index improvements might have reached its limits<sup>13</sup> as recent efforts to increase yield have primarily focused on increasing total biomass growth<sup>14</sup>. The potential of RUE for improving wheat biomass has not been fully utilized yet<sup>15</sup>. Improvement of RUE (as a complex trait) could be achieved by different processes such as

<sup>1</sup>Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany. <sup>2</sup>University of Sine Saloum EL-HÂDJ IBRAHIMA NIASS, Kaolack, Senegal. <sup>3</sup>Crop Science Group, University of Bonn, Institute of Crop Science and Resource Conservation (INRES), Bonn, Germany. <sup>4</sup>Department of Life Science Engineering, Technical University of Munich, Digital Agriculture, HEF World Agricultural Systems Center, Freising, Germany. <sup>5</sup>LEPSE, Université Montpellier, INRAE, Institut Agro Montpellier, Montpellier, France. <sup>6</sup>Institute of Environmental Sciences, Brandenburg University of Technology, Cottbus, Germany. <sup>7</sup>These authors contributed equally: Ehsan Eyshi Rezaei and Babacar Faye. ✉email: EhsanEyshi.Rezaei@zalf.de

a change in canopy photosynthesis efficiency, and enhanced carbon translocation<sup>16</sup>. Source limitations have been reported for modern wheat cultivars with a high harvest index<sup>17</sup>. However, higher biomass production by improved RUE would increase the relevance of the sink strength in keeping the source-sink balance<sup>18</sup>. An increment in biomass production should thus be accompanied by higher fruiting efficiency (FE) defined as grains set per unit of spike dry weight<sup>19</sup> to achieve a grain yield increase<sup>20</sup>. The light extinction coefficient ( $k$ ) determines the efficiency of light interception within the canopy profile<sup>21</sup>. More erect leaves lead to greater light penetration into the canopy profile increasing canopy RUE<sup>22</sup>. Disentangling the complex relationships between those traits (RUE, FE, and  $k$ ) and the environment is critical for guiding the breeding efforts in developing climate-resilient/high-yielding cultivars<sup>23</sup>. However, improving our mechanistic understanding by empirical experimentation only, is challenging due to costs, risks, and scaling issues.

Process-based crop models are robust tools to interpret genotype  $\times$  environment interactions from the results of experimental platforms<sup>24,25</sup>. The majority of these models are typically developed at the field scale, and upscaling their outcomes to regional or global scales amplifies uncertainty<sup>26</sup>. High-resolution model inputs are usually not available at larger scales, therefore, aggregated weather, soil, and crop data are often used for such simulation experiments<sup>27</sup>. Or outputs of crop models at high resolution are aggregated to a district, region, or a country, providing an overview for larger scales<sup>28</sup>, same as agricultural statistics obtained primarily from field scale and aggregate to a region or country. A few studies showed significant effects of weather data aggregation on crop model results, in spatially heterogeneous environments<sup>29</sup>. Most studies highlighted soil data as the primary source of aggregation error<sup>30–32</sup>. Weather input data source is another basis of uncertainty for large-scale modeling experiments<sup>33</sup>. Uncertainty of model results due to data source have been captured for both weather<sup>33</sup> and soil<sup>34</sup> inputs. However, the effects of data sources on crop model results have been less intensively studied than data aggregation impacts.

Crop models have already been implemented in genotype to phenotype pipelines at the field scale to explore the contribution of those traits to crop yield at larger spatial scales<sup>35</sup>. However, little is known about how the combined effect of input data sources and their associated resolutions (hereafter referred to as “resolution-sources”), along with data aggregation, influences the contribution of high-yielding traits to simulated yield. To narrow the knowledge gap, the presented research aims (i) to measure the effects of using different input data sources with different resolution-sources (field, 1 km  $\times$  1 km, 25 km  $\times$  25 km, and 50 km  $\times$  50 km) on simulated yield with variations of radiation use efficiency, fruiting efficiency, and light extinction coefficient for winter wheat, and (ii) to investigate the influence of output data aggregation on simulated yield with diverse trait combinations.

## Methods

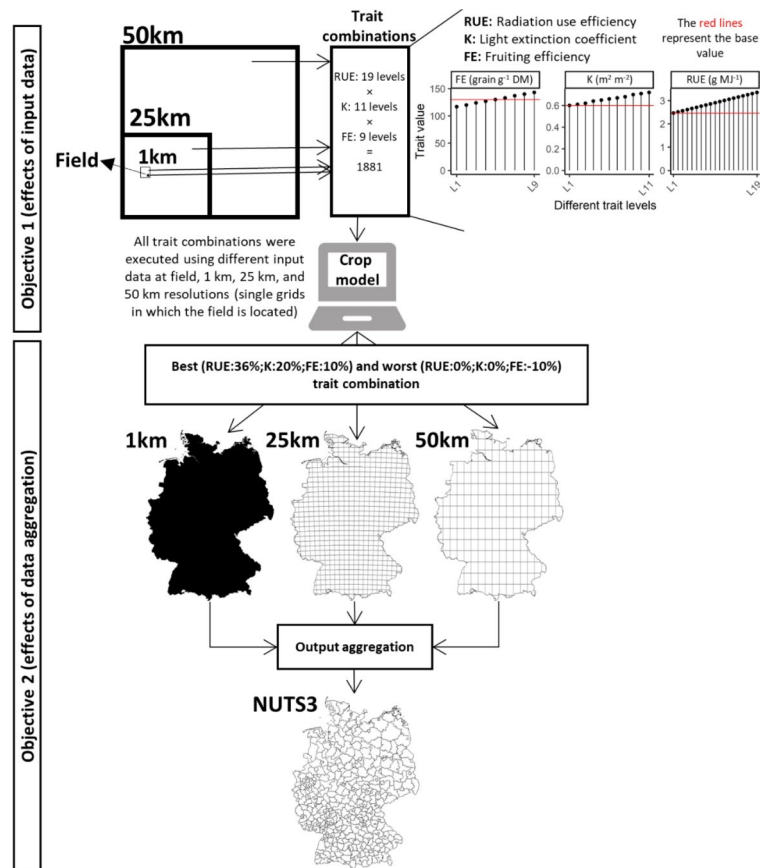
### General workflow of study

The study workflow was designed to address both objectives (see above) using a modeling platform (SIMPLACE)<sup>36</sup> and input data at different resolution-sources and trait combinations (Fig. 1). The weather and soil data were developed or extracted at 1 km, 25 km, and 50 km for Germany from different data sources. A factorial combination of traits (1,881 combinations) was developed<sup>24</sup> based on the observed potential of RUE, FE, and  $K$  of high-yielding wheat double haploid (DH) lines<sup>37</sup>. A modeling solution was executed using diverse scenarios under rainfed conditions. First, wheat growth and yield were simulated at the field scale and the single grids in which the field was located from 1 km, 25 km, and 50 km weather/soil datasets for all combinations of traits in the period 2001–2010. Second, the modeling solution was implemented for 1 km, 25 km, and 50 km across Germany using best and worst trait combinations (based on the results of the first step). The model outputs were aggregated to the Nomenclature of Territorial Units for Statistics Level 3 (NUTS3) scale to quantify the effects of data aggregation (Fig. 1).

### Weather and soil data

The crop model environmental inputs have been described in detail in various studies including<sup>38</sup>. They included daily weather (minimum and maximum temperatures, radiation, precipitation, and wind speed) and soil characteristics (volumetric (%) crop available water at permanent wilting point, field capacity, saturation, bulk density, soil organic carbon and rooting depth) at different resolutions were obtained from various sources in the period 2001–2010 across Germany. Field-scale model inputs (weather and soil) were obtained from an agricultural experimental station located in Northern Germany (54.53°N, 9.55°E). The weather data at 1 km were acquired by interpolating weather station data from the German weather service (DWD)<sup>29</sup>. The 25 km resolution weather data were extracted from the Joint Research Center’s (JRC) Agri4Cast database (version 1.0)<sup>39</sup>. Median of five CMIP6 global circulation models (IPSL-CM6A-LR, UKESM1-0-LL, GFDL-ESM4, MPI-ESM1-2-HR and MRI-ESM2-0) from Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) was employed as weather data inputs at 50 km resolution.

The soil data at 1 km resolution were derived from soil reconnaissance (BÜK1000N) by the Federal Institute for Geosciences and Natural Resources (BGR) based on the aggregation of the dominant soil type<sup>38</sup>. European Soil database from the JRC European Soil Data Portal (<http://eussoils.jrc.ec.europa.eu/>) was used to derive the soil data at 25 km resolution. The Corine Land Cover 2000 was used to mask out non-croplands (Webber et al., 2018<sup>39</sup>). The grid cells marked as non-cropland were masked out according to the Corine land cover 2006<sup>40</sup> for 1 km resolution. The soil data at 50 km resolution were derived from the Harmonized World Soil Database (ESDB) which was originally 30 arc-second ( $\approx$  1 km), aggregated at 50 km based on the procedure explained in Jägermeyr et al.<sup>6</sup>).



**Fig. 1.** Schematic overview of the progressing steps to address the study's objectives. The steps for objective 1 involved simulating 1,881 trait combinations with the crop model, using various field-scale data and grid resolutions of 1 km x 1 km, 25 km x 25 km, and 50 km x 50 km. For objective 2, the steps included running the crop model with both the best and worst trait combinations for all grid resolutions throughout Germany. The simulated yield was then aggregated to the NUTS3 level. The maps were generated with R package terra, version 1.7-78 using shapefiles from the federal states obtained from the Federal Agency for Cartography and Geodesy of Germany (<http://www.geodatenzentrum.de>).

### High-yielding traits

The high-yielding wheat traits were developed from the difference in RUE, K, and FE between a check cultivar (Bacanora) and the two highest yielding lines of the Bacanora × Weebil DH population<sup>37,41</sup>. The improvement in RUE, K and FE traits were up to +34%, +10% and -5%, respectively<sup>24,42</sup>. The ranges of each trait are defined based on those measurements, which resulted in 1,881 unique trait combinations including 19, 11 and 9 levels of RUE, K and FE, respectively (Fig. 1). Improved high-yielding traits were obtained under optimized nutritional (300 kg N ha<sup>-1</sup>, 300 kg P<sub>2</sub>O<sub>5</sub> ha<sup>-1</sup>, and 150 kg K<sub>2</sub>O ha<sup>-1</sup>) and agronomic management (including proper irrigation and controlling pest and disease) conditions<sup>37</sup>. The base values for RUE, K and FE were 2.46 g dry mass MJ<sup>-1</sup> PAR, 0.6 m<sup>-2</sup> ground m<sup>-2</sup> leaves and 130 grain number g<sup>-1</sup> spike dry mass, respectively<sup>42</sup>.

### Crop model setup

Scientific Impact assessment and modeling platform for advanced crop and ecosystem management (SIMPLACE) is a modeling framework based on the solution concept, including discrete and interchangeable modules<sup>36</sup>. It has been broadly employed in process understanding and impact assessment studies, particularly in capturing heat, drought, and nitrogen signals on crop growth<sup>30,39,43</sup>. SIMPLACE was developed by pairing the Lintul-5 coupled with SlimWater, the FAO-56 Penman-Monteith for evapotranspiration estimation, and a canopy temperature module model as input of the heat stress module described by Webber et al.<sup>38</sup>. The SIMPLACE modeling platform was chosen because it performed relatively similarly to the multi-model ensembles in previous studies<sup>44</sup>.

In SIMPLACE, crop development is driven by daily mean air temperatures modified by photoperiod and vernalization (for winter crops)<sup>45</sup>. Biomass accumulation is initially driven by light interception as a function of the green leaf area, then transformed into dry matter using the radiation use efficiency concept modified by the phenological stage. Heat stress impacts grain yield when the estimated canopy temperature exceeds the defined temperature threshold (31 °C for wheat) around anthesis. Water stress is modeled by using the ratio between actual and potential transpiration affecting leaf area expansion, carbon partitioning to different organs, and RUE. The crop model was not parametrized for yield, but wheat phenology and sowing/harvest dates were

calibrated using the German weather service database at grid scale<sup>46</sup>. Nitrogen stress was not included in these simulations, assuming in general well fertilized wheat crops. However, heat and drought stress were stressors affecting crop growth. Simulated yields are reported at 0% moisture.

### Simulation scenarios

To address the study's first objective, the modelling solution was executed for field and single grid cells from different input datasets (1 km, 25 km, and 50 km) where the field was located in Schleswig–Holstein, Germany using all trait combinations for the period 2001–2010. The worst (RUE:0%;K:0%;FE:-10%) and best (RUE:36%;K:20%;FE:10%) trait combinations<sup>24</sup> were selected based on the median yield at the first step and used for the second step of simulations at Germany scale. The country wide simulations were conducted using those two trait combinations for all study resolution-sources addressing the second objective of the study. The simulated yield for each resolution-source was aggregated (output aggregation) to the NUTS3 scale by averaging the yield for grids inside each polygon, quantifying the effects of data aggregation on simulated yield. The aggregation procedure for the 25 km and 50 km resolutions was performed using an area-weighted approach.

### Data analysis

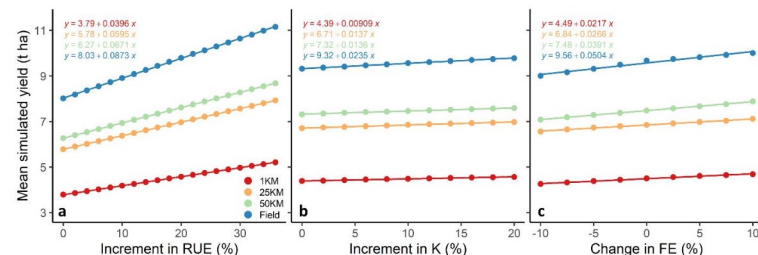
Summary statistics, frequency distributions, and kernel probability density of simulated yields at different input resolution-sources were calculated, quantifying the importance of each trait and data aggregation effects. Spatial patterns of simulated yields were compared between 1 km and other resolutions by calculating the absolute difference (AD) and the difference (D) at aggregated NUTS3 scale. In addition, the mean of absolute differences and the mean of differences across NUTS3 units were calculated. As AD is always positive, it is possible to measure the agreement of maps at the polygon scale by the mean of AD across all polygons. The mean of D was computed across all polygons to indicate systematic bias between aggregated maps. A segmented, piecewise linear regression<sup>47</sup> was used to explore the yield response to climate and soil variables. Through maximizing the combined coefficient of determination ( $R^2$ ) of the two linear regression estimates before and after the breakpoint, one single break point was detected<sup>48</sup>. All data analysis was conducted using RStudio (R version 4.0.4).

## Results and discussion

### Impact of high-yielding traits on simulated yield across different input data resolution-sources

As in the study of Stella et al. (2023)<sup>24</sup> across many environments, results here for RUE showed large impacts on simulated yield for the various resolution-sources input data (Fig. 2). Increments in RUE compared to the base cultivar increased the yield by up to 2.28 t ha<sup>-1</sup> (as mean over resolutions). However, increments in K and FE resulted in remarkably less (<0.4 t ha<sup>-1</sup>) yield improvement (Fig. 2). Field scale data resulted in the strongest response to change for all high-yielding traits (0.02 to 0.09 t ha<sup>-1</sup> increase per % of trait increment), while 1 km input data resulted in the least (0.01 to 0.04 t ha<sup>-1</sup> increase per % of trait value increment) yield improvement (Fig. 2). Yield response to high-yielding traits using 25 km and 50 km inputs were between field and 1 km (Supplementary Fig. 1).

Despite the large differences in simulated yield resulting from various input data resolution-sources, the ranking of high-yielding traits was not affected. Superior traits (and their sizes) consistently led to higher yields. Although, the impact of RUE increment as the most influential trait on yield (0.09 t ha<sup>-1</sup> increase per % of RUE increment) notably receded when the input data resolution-source shifted from field scale to 1 km scale (0.04 t ha<sup>-1</sup> increase per % of RUE increment). The differences in performance of high-yielding traits were also confirmed by testing their potential across diverse environments<sup>24</sup>. Some literature demonstrated the significant potential of increased RUE to enhance yield by promoting biomass growth<sup>14,49</sup>. The improvement in grain yield of the best yielding DH lines was 54% higher than that of the check cultivars (Invento-BAER), mainly explained by increase in grain number and RUE<sup>37</sup>. The results of the same experiments on high-yielding DH lines showed a curvilinear relationship between grain yield and grain number, which could lead to a decline in the potential of grain number for future yield improvement<sup>37</sup>. It is also possible that having more erect flag leaves may not improve overall light interception. The ideal canopy structure is conceptualized by a moderate leaf angle at the bottom and erect leaves at the top of the canopy<sup>50</sup>.

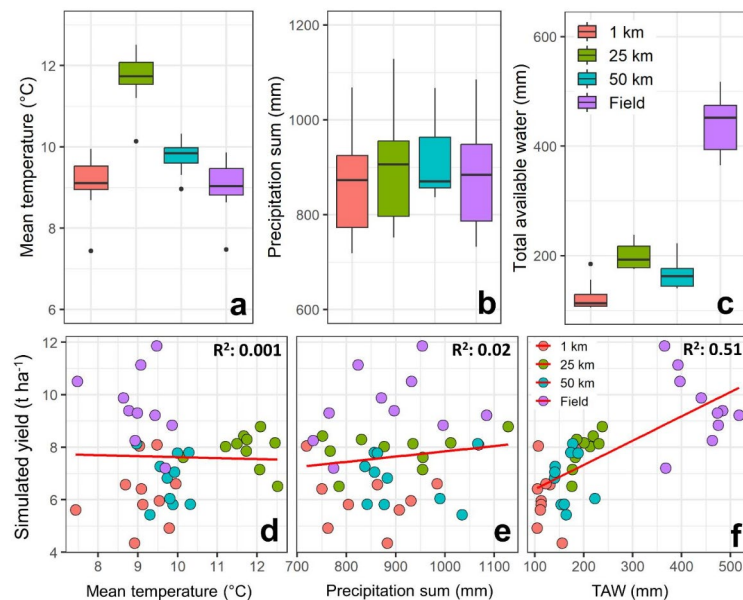


**Fig. 2.** Mean simulated yield (2001–2010) for various levels (increment from base cultivar) of radiation use efficiency (RUE) (a), light extinction coefficient (K) (b), and fruiting efficiency (FE) (c) executed at field, 1 km, 25 km and 50 km resolution-source (single grids in which the field is located) input data.

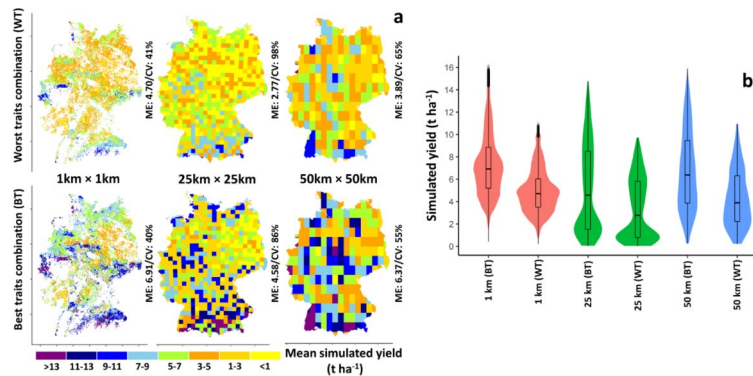


There was a significant variation in the simulated yield, with the highest yield resulting from field inputs and the lowest yield derived from 1 km resolution input data, regardless of the range of traits (Fig. 2). Comparison of the precipitation sum from various input resolutions and sources revealed no substantial difference, ranging from 867 mm per year to 910 mm per year (Fig. 3). However, there was a remarkable variation in mean temperature (from 9.1 °C to 11.7 °C) and total available soil water (125 mm to 437 mm) during the growth period among different input resolutions and sources (Fig. 3). Analysis of the relationships between model inputs and simulated yields revealed that total available water (TAW) was the main factor explaining yield variation ( $R^2=0.51$ ), considerably more than mean temperature ( $R^2=0.001$ ) and precipitation sum ( $R^2=0.02$ ; Fig. 3). Exploring the correlation between the difference in simulated yield (field minus other resolution-sources) and the difference in various input data (input from the field minus inputs from other resolution-sources) indicated that the difference in TAW showed the highest correlation with yield difference at 1 km (-0.61) and 25 km (-0.58) resolutions (Supplementary Fig. 2). However, the difference in mean temperature shows the highest correlation with yield difference for the 50 km (-0.55) resolution (Supplementary Fig. 2). The differences in TAW among resolutions, being higher with field data used as model input and much lower with 1 km data as input, may explain the large differences in simulated yields among resolutions shown in Fig. 2. This indicates the importance of soil data as the main driver of yield differences when altering input resolution and sources. Other scaling studies have also highlighted the significance of soil data on simulated yield compared to climate data<sup>30,51</sup>. Our analysis (Supplementary Fig. 2) reveals that yield-driving factors can be scale-specific. The correlation between yield differences (field vs. 50 km resolution) and TAW differences was notably lower (-0.27) compared to mean temperature differences (-0.55). This indicates that primary yield-driving factors can switch depending on the simulation scale and the response function to the input variable in the model. Such switches in yield-driving factors have been previously observed in different nutrition treatments<sup>52</sup>, supporting our findings. However, the response of simulated yield to changes in traits on was not considerably influenced by weather or soil data resolution and source. However, soil data uncertainty can potentially overshadow the response to climate on projected yields in impact assessment studies<sup>53</sup>.

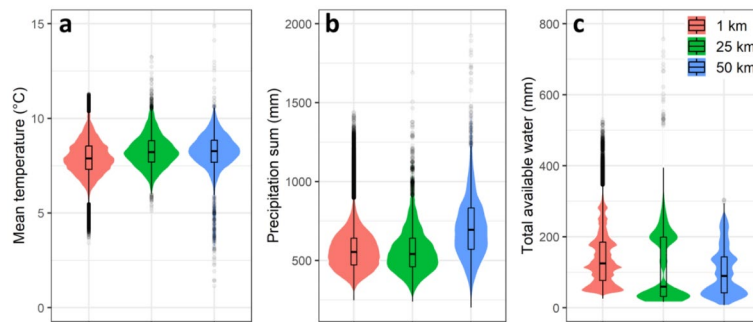
The results of large-scale simulations across Germany showed a large influence of input data resolution and sources on the simulated yield for the worst and best trait combinations (Fig. 4). Using 1 km resolution as input data resulted in the highest country-wide mean simulated yield ( $4.70 \text{ t ha}^{-1}$  and  $6.91 \text{ t ha}^{-1}$ ) with little difference in spatial variability (40% and 41%) for both trait combinations (Fig. 4a). On the other hand, implementing the crop model with 25 km resolution input data led to the lowest simulated mean yield ( $2.77 \text{ t ha}^{-1}$  and  $4.58 \text{ t ha}^{-1}$ ) and highest spatial yield variability (86% and 98%) (Fig. 4a). The variability of simulated yields (indicated by interquartile range) was higher for the best trait combinations (3.7, 7.0, and  $5.6 \text{ t ha}^{-1}$  for 1, 25 and 50 km, respectively) than for the worst trait combinations (2.5, 5.0, and  $4.1 \text{ t ha}^{-1}$ , for 1 km, 25, and 50 km, respectively) for all input data resolution-sources (Fig. 4b). Comparing the spatial patterns of simulated yield for both the best and worst trait combinations with the spatial pattern of observed yield (reported yield



**Fig. 3.** Boxplot comparing the mean temperature, precipitation sum during growth period, and total available water (TAW) for model inputs at field, 1 km, 25 km, and 50 km resolutions for the grid cell corresponding to the field plot considered in this study for the 2001–2010 period (a–c). (d–f) Relationships between yield and model inputs across various model resolutions and sources as input. Total available water is simulated in the SlimWater<sup>54</sup> component of the model, which considers the volumes of crop water uptake, soil evaporation, surface runoff, and percolation below the root zone.



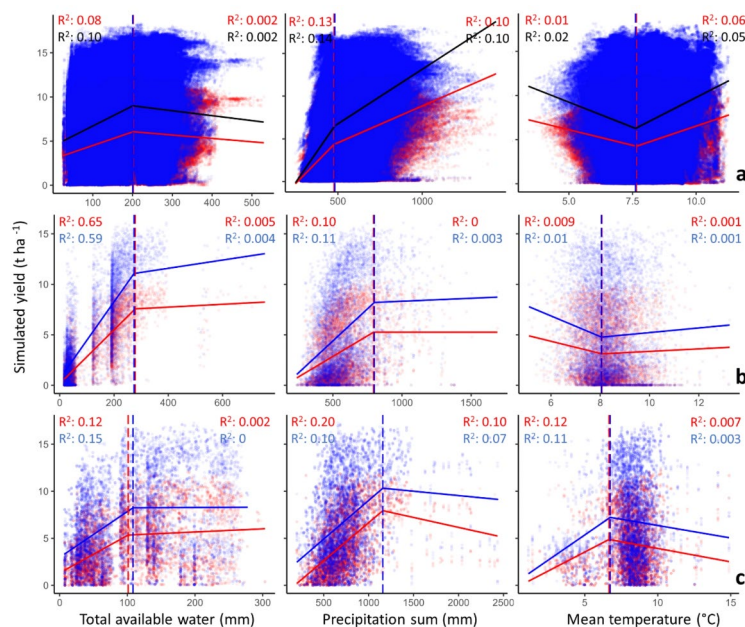
**Fig. 4.** The grid scale spatial pattern (a) and violin/box plot (b) of simulated yield (2001–2010) for worst (WT; RUE: 0%; K: 0%; FE: –10%) and best (BT; RUE: 36%; K: 20%; FE: 10%) traits combinations using 1, 25, and 50 km resolution-source input data. *ME* mean, *CV* coefficient of variation. The maps were generated with R package terra, version 1.7-78 using shapefiles from the federal states obtained from the Federal Agency for Cartography and Geodesy of Germany (<http://www.geodatenzentrum.de>).



**Fig. 5.** Violin/box plot of the mean temperature (a), annual precipitation sum (b), and total available water (c) during the growth period (emergence to maturity) for 1, 25, and 50 km input data resolution-source for all soil-weather grids nationwide.

statistics<sup>55</sup>) revealed that the model accurately captured the lower yields in eastern Germany and higher yields in the western parts. However, it overestimated the yield in the southern parts of the country (Supplementary Fig. 3). The model generally failed to account for the negative impacts of excessive rainfall events<sup>38</sup>, which are more prevalent in the southern parts of the country. The relatively low accuracy of the model in capturing the observed yield (Supplementary Fig. 3) could be attributed to the fact that the model was only calibrated for phenology. This calibration approach was chosen because the main objective of the study was to capture the yield response to changes in high-yielding traits, rather than to produce the best estimation of observed yield. Also, variability in management such as sowing data and in maturity type of varieties not considered in the simulations may have contributed to the deviation of model simulations from observations. Comparing the distribution of mean simulated yield for each NUTS3 unit using different source-resolutions and best and worst trait combinations across Germany with observed yield indicated a significant difference between simulated and observed yield distributions for all simulation options. However, the worst trait combination in 1 km and 50 km source-resolutions showed better overlap with the observed distribution (Supplementary Fig. 4). This suggests that the model overestimated yields when using the best trait combination, while the worst trait combination might better represent the average conditions across diverse NUTS3 units.

Seasonal mean temperatures at different spatial resolutions showed substantial overlap in their distribution with differences in medians (Fig. 5a). The median seasonal precipitation sum was 20% higher at a 50 km resolution than at other data resolutions (Fig. 5b). The cumulative available soil water at a 1 km resolution was 53% and 29% higher than those at 25 and 50 km, respectively (Fig. 5c). While the spatial patterns of those variables appear relatively similar, the value ranges differ notably (Supplementary Fig. 5). The processing of model outputs indicated that the responsiveness of simulated yield to weather and soil variables is related to the source-resolution of input data (Fig. 6). Precipitation sum during the growth period showed the largest impact on simulated yield variability ( $R^2 = 0.10$  and  $0.14$ ) when using 1 km data as model input (Fig. 6a). However, simulated yield at 25 km was more affected by soil total available water ( $R^2 = 0.65$  and  $0.59$ ) than by the variability of precipitation sum and mean temperature during the growth period for both the worst and best trait combinations (Fig. 6b). Soil water and precipitation sum indicated relatively similar impacts on simulated



**Fig. 6.** Segmented piecewise regression between crop yields simulated at 1 km (a), 25 km (b), and 50 km (c) source-resolution and total available water, seasonal precipitation sum, and mean temperature during the growing season for the worst (RUE: 0%; K: 0%; FE:  $-10\%$ —presented in red) and best (RUE: 36%; K: 20%; FE:  $10\%$ —presented in blue) trait combinations in the period 2001–2010. The red and blue lines show the estimated break points for the segmented regression. The  $R^2$  values represent the coefficient of determination before and after the estimated break point.

yield when the model was executed using 50 km inputs (Fig. 6c). The trait combinations did not remarkably influence the yield response to the soil and weather variables in all resolutions, and the estimated breakpoints were relatively similar between best and worst trait combinations (Fig. 6).

Experimental evidence has shown the adverse potential of terminal drought on significant suppression of the high-yielding traits of wheat without a remarkable impact of high temperature<sup>56</sup>. The yield advantage of high-yielding wheat genotypes (with higher biomass production as a driver) under optimal conditions was largely suppressed by drought stress<sup>57</sup>. Large-scale simulations have also highlighted the importance of water (soil characteristics and precipitation sum) as the primary drivers of input uncertainty<sup>53</sup> and yield variability<sup>58</sup>.

### Effect of output data aggregation on simulated yield

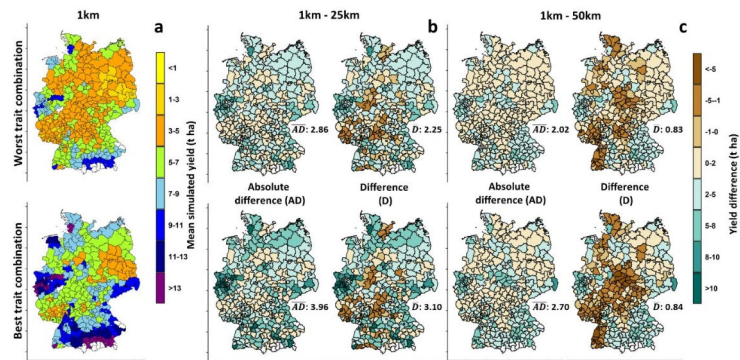
Model input resolution-source had a slight impact ( $0.09$  to  $0.68$   $t^{-1}$  ha) on average simulated yield (2001–2010) aggregated at NUTS3 level for both worst and best traits combination (Fig. 4 & supplementary Fig. 6). Yield variability (measured by the coefficient of variation) at NUTS3 level was comparable for the different input data resolution-sources. The mean absolute difference ( $\overline{AD}$ ) and difference ( $\overline{D}$ ) between 1 km and other resolutions were remarkably higher for 25 km ( $\overline{AD} = 2.86$ – $3.96$   $t^{-1}$  ha $^{-1}$  and  $\overline{D} = 2.25$ – $3.10$   $t^{-1}$  ha $^{-1}$ ) compared with 50 km, particularly for the best trait combination (Fig. 7). There was no systematic spatial yield pattern change observed between the 1 km and 50 km resolution sources, as both positive and negative differences were present, and the mean difference was remarkably smaller than the mean absolute difference (Fig. 7). However, using weather and soil datasets at 25 km as model input resulted in a systematic difference compared to 1 km input data in the spatial pattern of simulated yield because  $\overline{AD}$  was close to  $\overline{D}$  (Fig. 7). The greatest spatial yield difference (1–25 km) was obtained in the western and central parts of the country (Fig. 7).

There were large differences in simulated yield between the extreme low (2003) and best yielding year (2002) years (Fig. 8). Yield aggregated at NUTS3 level increased by 63% ( $4$   $t$  ha $^{-1}$ ) and 59% ( $5$   $t$  ha $^{-1}$ ) tags from extreme low to the best yielding year for the worst and best trait combinations, respectively (Fig. 8). The impact of high-yielding traits on simulated yield aggregated at NUTS3 level was limited to 34% ( $1.96$   $t$  ha $^{-1}$ ; Fig. 8). Nevertheless, data aggregation resulted in a slight increase in mean simulated yield of  $0.2$   $t$  ha $^{-1}$  (for the best yielding year) to  $0.5$   $t$  ha $^{-1}$  (for extreme low yield year; Fig. 8). The current results confirm the findings of previous studies on the minor effects of output data aggregation on mean simulated yield, particularly when soil data aggregation (different data sources were implemented in the current study) was not involved<sup>28,29,59</sup>.

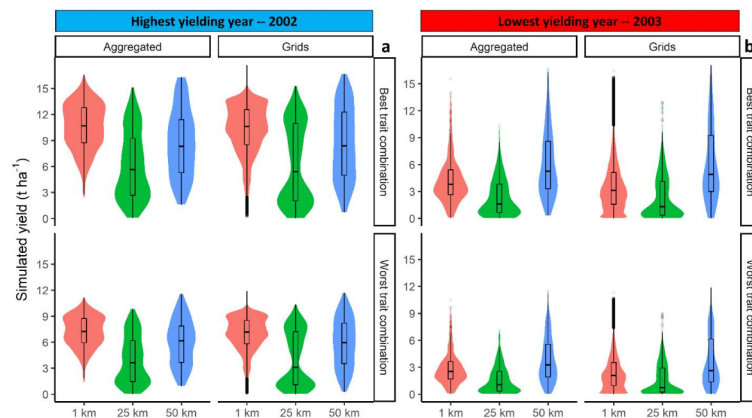
### Conclusion

The resolution and sources of input data did not alter the ranking of high-yielding trait impacts on grain yield, but they did influence the potential for yield increase achieved by implementing those traits. Nonetheless, it is





**Fig. 7.** Average simulated yield aggregated from 1 km to NUTS3 scale using worst (RUE: 0%; K: 0%; FE: – 10%) and best (RUE: 36%; K: 20%; FE: 10%) trait combinations (a). The maps in panels (b, c) illustrate the absolute (AD) and regular (D) differences in aggregated yield (grid to NUTS3 scale) between 1 km and 25 km (b)/50 km (c) resolutions for the period 2001–2010, using both the worst and best trait combinations. The maps were generated with R package terra, version 1.7-78 using shapefiles from the federal states obtained from the Federal Agency for Cartography and Geodesy of Germany (<http://www.geodatenzentrum.de>).



**Fig. 8.** Violin/box plot of aggregated and grid level simulated yield for the highest (a) and the lowest (b) yielding year, based for the worst (RUE: 0%; K: 0%; FE: – 10%) and best (RUE: 36%; K: 20%; FE: 10%) trait combinations simulated with input data at 1, 25, and 50 km resolution-source. Extreme low and the highest yielding years were selected based on the median simulated yield at 1 km resolution.

essential to acknowledge that the outcomes might have been substantially dissimilar for adaptive traits, such as RUE response to water deficit or change in stomatal conductance to vapor pressure deficit. The impact on yield of RUE was much more substantial than that of K or FE for all input data resolution and sources. However, the yield improvement rate per percent increase in RUE was resolution-specific. Our results suggest that water related inputs affecting plant water availability is the primary driver of simulated yield variability (compared to temperature) among the various model input datasets. However, it did not change the contribution of high-yielding traits to increasing yield, which means that yield-increasing trait combinations always results in higher yield simulations regardless of input data resolution and source. Data aggregation had a minor effect on the simulated yield for different trait combinations. Finally, warm-dry conditions (such as year 2003) can mostly suppress a positive impact of high-yielding traits regardless of input data resolution-sources. Changing the model input data resolution-source from field to 50 km affects the potential size of yield improvement from high-yielding traits, highlighting the need for high-quality input data and a tailored approach and careful interpretation when upscaling simulation results from the field to regional scales. Future research should aim to test the potential of high-yielding traits under more extreme seasonal growing conditions and under sub-optimal management, where yield benefits might be low, using various model input sources and resolutions. In addition, it should be noted that current crop models cannot fully address the interactions among RUE, the K, and FE due to the complexity of these interactions and the lack of process understanding of the dynamics of source-sink relationships<sup>60</sup>. This limitation suggests the need for repeating such modeling experiments using a new generation of models that better consider these interactions.

## Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Received: 19 June 2024; Accepted: 25 September 2024

Published online: 05 October 2024

## References

- van Dijk, M., Morley, T., Rau, M. L. & Saghai, Y. A meta-analysis of projected global food demand and population at risk of hunger for the period 2010–2050. *Nat. Food*. **2**, 494–501. <https://doi.org/10.1038/s43016-021-00322-9> (2021).
- Tian, X. et al. Will reaching the maximum achievable yield potential meet future global food demand? *J. Clean. Prod.* **294**, 126285. <https://doi.org/10.1016/j.jclepro.2021.126285> (2021).
- Beltran-Peña, A., Rosa, L. & D'Odorico, P. Global food self-sufficiency in the 21st century under sustainable intensification of agriculture. *Environ. Res. Lett.* <https://doi.org/10.1088/1748-9326/ab9388> (2020).
- Lesk, C. et al. Stronger temperature–moisture couplings exacerbate the impact of climate warming on global crop yields. *Nat. Food*. **2**, 683–691. <https://doi.org/10.1038/s43016-021-00341-6> (2021).
- Wang, X. et al. Emergent constraint on crop yield response to warmer temperature from field experiments. *Nat. Sustain.* **3**, 908–916. <https://doi.org/10.1038/s41893-020-0569-7> (2020).
- Jägermeyr, J. et al. Climate impacts on global agriculture emerge earlier in new generation of climate and crop models. *Nat. Food*. **2**, 873–885. <https://doi.org/10.1038/s43016-021-00400-y> (2021).
- Ray, D. K., Ramankutty, N., Mueller, N. D., West, P. C. & Foley, J. A. Recent patterns of crop yield growth and stagnation. *Nat. Commun.* **3**, 1293. <https://doi.org/10.1038/ncomms2296> (2012).
- Sharma, R. C. et al. Genetic gains for Grain Yield in CIMMYT Spring Bread Wheat across International environments. *Crop Sci.* **52**, 1522–1533. <https://doi.org/10.2135/cropsci2011.12.0634> (2012).
- Tshikunde, N. M., Mashilo, J., Shimelis, H. & Odindo, A. Agronomic and physiological traits, and Associated Quantitative Trait Loci (QTL) affecting yield response in wheat (*Triticum aestivum* L.): A review. *Front. Plant Sci.* **10**, 1428. <https://doi.org/10.3389/fpls.2019.01428> (2019).
- Bailey-Serres, J., Parker, J. E., Ainsworth, E. A., Oldroyd, G. E. D. & Schroeder, J. I. Genetic strategies for improving crop yields. *Nature*. **575**, 109–118. <https://doi.org/10.1038/s41586-019-1679-0> (2019).
- Reynolds, M. et al. Raising yield potential of wheat. I. Overview of a consortium approach and breeding strategies. *J. Exp. Bot.* **62**, 439–452. <https://doi.org/10.1093/jxb/erq311> (2011).
- Foulkes, M. J. et al. Raising yield potential of wheat. III. Optimizing partitioning to grain while maintaining lodging resistance. *J. Exp. Bot.* **62**, 469–486. <https://doi.org/10.1093/jxb/erq300> (2011).
- Brancourt-Hulmel, M. et al. Genetic improvement of agronomic traits of Winter Wheat cultivars Released in France from 1946 to 1992. *Crop Sci.* **43**, 37. <https://doi.org/10.2135/cropsci2003.3700> (2003).
- Molero, G. et al. Elucidating the genetic basis of biomass accumulation and radiation use efficiency in spring wheat and its role in yield potential. *Plant Biotechnol. J.* **17**, 1276–1288. <https://doi.org/10.1111/pbi.13052> (2019).
- Acreche, M. M. & Slafer, G. A. Grain weight, radiation interception and use efficiency as affected by sink-strength in Mediterranean wheats released from 1940 to 2005. *Field Crops Res.* **110**, 98–105. <https://doi.org/10.1016/j.fcr.2008.07.006> (2009).
- Asseng, S. et al. Model-driven multidisciplinary global research to meet future needs: The case for improving radiation use efficiency to increase yield. *Crop Sci.* **59**, 843–849. <https://doi.org/10.2135/cropsci2018.09.0562> (2019).
- Álvaro, F., Royo, C., García Moral, D., Villegas, D. & L. F. & Grain Filling and Dry Matter translocation responses to Source-Sink modifications in a historical series of Durum Wheat. *Crop Sci.* **48**, 1523–1531. <https://doi.org/10.2135/cropsci2007.10.0545> (2008).
- Reynolds, M., Calderini, D., Condon, A. & Vargas, M. Association of source/sink traits with yield, biomass and radiation use efficiency among random sister lines from three wheat crosses in a high-yield environment. *J. Agric. Sci.* **145**, 3–16. <https://doi.org/10.1017/S0021859607006831> (2007).
- Slafer, G. A. et al. Fruiting efficiency: An alternative trait to further rise wheat yield. *Food Energy Secur.* **4**, 92–109. <https://doi.org/10.1002/fes3.59> (2015).
- Acreche, M. M., Briceño-Félix, G., Martín Sánchez, J. A. & Slafer, G. A. Radiation interception and use efficiency as affected by breeding in Mediterranean wheat. *Field Crops Res.* **110**, 91–97. <https://doi.org/10.1016/j.fcr.2008.07.005> (2009).
- Zhang, L., Hu, Z., Fan, J., Zhou, D. & Tang, F. A meta-analysis of the canopy light extinction coefficient in terrestrial ecosystems. *Front. Earth Sci.* **8**, 599–609. <https://doi.org/10.1007/s11707-014-0446-7> (2014).
- Furbank, R. T., Quick, W. P. & Sirault, X. R. Improving photosynthesis and yield potential in cereal crops by targeted genetic manipulation: Prospects, progress and challenges. *Field Crops Res.* **182**, 19–29. <https://doi.org/10.1016/j.fcr.2015.04.009> (2015).
- de Leon, N., Jannink, J. L., Edwards, J. W. & Kaeppeler, S. M. Introduction to a special issue on genotype by Environment Interaction. *Crop Sci.* **56**, 2081–2089. <https://doi.org/10.2135/cropsci2016.07.0002in> (2016).
- Stella, T. et al. Wheat crop traits conferring high yield potential may also improve yield stability under climate change. *In Silico Plants*. <https://doi.org/10.1093/insilicoplants/diad013> (2023).
- Senapati, N. et al. Global wheat production could benefit from closing the genetic yield gap. *Nat. Food*. **3**, 532–541. <https://doi.org/10.1038/s43016-022-00540-9> (2022).
- Manivasagam, V. S. & Rozenstein, O. Practices for upscaling crop simulation models from field scale to large regions. *Comput. Electron. Agric.* **175**, 105554. <https://doi.org/10.1016/j.compag.2020.105554> (2020).
- Grassini, P. et al. How good is good enough? Data requirements for reliable crop yield simulations and yield-gap analysis. *Field Crops Res.* **177**, 49–63. <https://doi.org/10.1016/j.fcr.2015.03.004> (2015).
- Ojeda, J. J. et al. Implications of data aggregation method on crop model outputs—The case of irrigated potato systems in Tasmania, Australia. *Eur. J. Agron.* **126**, 126276. <https://doi.org/10.1016/j.eja.2021.126276> (2021).
- Zhao, G. et al. Demand for multi-scale weather data for regional crop modeling. *Agric. For. Meteorol.* **200**, 156–171. <https://doi.org/10.1016/j.agrformet.2014.09.026> (2015).
- Rezaei, E. E., Siebert, S. & Ewert, F. Impact of data resolution on heat and drought stress simulated for winter wheat in Germany. *Eur. J. Agron.* **65**, 69–82. <https://doi.org/10.1016/j.eja.2015.02.003> (2015).
- Grosz, B. et al. The implication of input data aggregation on up-scaling soil organic carbon changes. *Environ. Model. Softw.* **96**, 361–377. <https://doi.org/10.1016/j.envsoft.2017.06.046> (2017).
- Ojeda, J. J. et al. Impact of crop management and environment on the spatio-temporal variance of potato yield at regional scale. *Field Crops Res.* **270**, 108213. <https://doi.org/10.1016/j.fcr.2021.108213> (2021).
- van Wart, J., Grassini, P. & Cassman, K. G. Impact of derived global weather data on simulated crop yields. *Glob. Change Biol.* **19**, 3822–3834. <https://doi.org/10.1111/gcb.12302> (2013).
- Kværnø, S. H. & Stolte, J. Effects of soil physical data sources on discharge and soil loss simulated by the LISEM model. *CATENA*. **97**, 137–149. <https://doi.org/10.1016/j.catena.2012.05.001> (2012).

35. Cooper, M., Voss-Fels, K. P., Messina, C. D., Tang, T. & Hammer, G. L. Tackling  $G \times E \times M$  interactions to close on-farm yield-gaps: Creating novel pathways for crop improvement by predicting contributions of genetics and management to crop productivity. *TAG. Theor. Appl. Genet. Theoretische Und Angewandte Genetik*. **134**, 1625–1644. <https://doi.org/10.1007/s00122-021-03812-3> (2021).
36. Enders, A. et al. SIMPLACE—A versatile modelling and simulation framework for sustainable crops and agroecosystems. *In Silico Plants*. <https://doi.org/10.1093/insilicoplants/diad006> (2023).
37. Bustos, D. V., Hasan, A. K., Reynolds, M. P. & Calderini, D. F. Combining high grain number and weight through a DH-population to improve grain yield potential of wheat in high-yielding environments. *Field Crops Res.* **145**, 106–115. <https://doi.org/10.1016/j.fcr.2013.01.015> (2013).
38. Webber, H. et al. No perfect storm for crop yield failure in Germany. *Environ. Res. Lett.* <https://doi.org/10.1088/1748-9326/aba2a4> (2020).
39. Webber, H. et al. Diverging importance of drought stress for maize and winter wheat in Europe. *Nat. Commun.* <https://doi.org/10.1038/s41467-018-06525-2> (2018).
40. CORINE. CORINE Land Cover. (2006). <https://land.copernicus.eu/pan-european/corine-land-cover/clc-2006>
41. García, G. A. et al. Wheat grain number: identification of favourable physiological traits in an elite doubled-haploid population. *Field Crops Res.* **168**, 126–134. <https://doi.org/10.1016/j.fcr.2014.07.018> (2014).
42. Guarín, J. R. et al. Evidence for increasing global wheat yield potential. *Environ. Res. Lett.* <https://doi.org/10.1088/1748-9326/aca77c> (2022).
43. Gabaldón-Leal, C. et al. Modelling the impact of heat stress on maize yield formation. *Field Crops Res.* **198**, 226–237. <https://doi.org/10.1016/j.fcr.2016.08.013> (2016).
44. Asseng, S. et al. Rising temperatures reduce global wheat production. *Nat. Clim. Change.* **5**, 143–147. <https://doi.org/10.1038/nclimate2470> (2015).
45. Rezaei, E. E., Siebert, S. & Ewert, F. in *Temperature Routines in SIMPLACE <LINTUL2-CC-HEAT>*. (eds Alderman, P., Quilligan, E., Asseng, S., Ewert, F. & Reynolds, M. P.) (El Batán, 2013).
46. Kaspar, F., Zimmermann, K. & Polte-Rudolf, C. An overview of the phenological observation network and the phenological database of Germany's national meteorological service (Deutscher Wetterdienst). *Adv. Sci. Res.* **11**, 93–99. <https://doi.org/10.5194/asr-11-93-2014> (2014).
47. Muggeo, V., Atkins, D. C., Gallop, R. J. & Dimidjian, S. Segmented mixed models with random change-points: A maximum likelihood approach with application to treatment for depression study. *Stat. Modelling.* **14**, 293–313. <https://doi.org/10.1177/1471082X13504721> (2014).
48. Rezaei, E. E., Siebert, S. & Ewert, F. Climate and management interaction cause diverse crop phenology trends. *Agric. For. Meteorol.* **233**, 55–70. <https://doi.org/10.1016/j.agrformet.2016.11.003> (2017).
49. Reynolds, M. P. et al. Strategic crossing of biomass and harvest index—Source and sink—Achieves genetic gains in wheat. *Euphytica*. <https://doi.org/10.1007/s10681-017-2040-z> (2017).
50. Li, P. F. et al. Dryland wheat domestication changed the development of aboveground architecture for a well-structured canopy. *PLoS One.* **9**, e95825. <https://doi.org/10.1371/journal.pone.0095825> (2014).
51. Maharjan, G. R. et al. Effects of input data aggregation on simulated crop yields in temperate and Mediterranean climates. *Eur. J. Agron.* **103**, 32–46. <https://doi.org/10.1016/j.eja.2018.11.001> (2019).
52. Eyshi Rezaei, E., Gaiser, T., Siebert, S., Sultan, B. & Ewert, F. Combined impacts of climate and nutrient fertilization on yields of pearl millet in Niger. *Eur. J. Agron.* **55**, 77–88. <https://doi.org/10.1016/j.eja.2014.02.001> (2014).
53. Folberth, C. et al. Uncertainty in soil data can outweigh climate impact signals in global crop yield simulations. *Nat. Commun.* **7**, 11872. <https://doi.org/10.1038/ncomms11872> (2016).
54. Addiscott, T. M. & Whitmore, A. P. Simulation of solute leaching in soils of differing permeabilities. *Soil Use Manag.* **7**, 94–102. <https://doi.org/10.1111/j.1475-2743.1991.tb00856.x> (1991).
55. Duden, C., Nacke, C. & Offermann, F. German yield and area data for 11 crops from 1979 to 2021 at a harmonized spatial resolution of 397 districts. *Sci. data.* **11**, 95. <https://doi.org/10.1038/s41597-024-02951-8> (2024).
56. Dias de Oliveira, E. A., Siddique, K. H. M., Bramley, H., Stefanova, K. & Palta, J. A. Response of wheat restricted-tillering and vigorous growth traits to variables of climate change. *Glob. Change Biol.* **21**, 857–873. <https://doi.org/10.1111/gcb.12769> (2015).
57. Dodig, D., Zoric, M., Knezevic, D., King, S. R. & Surlan-Momirovic, G. Genotype  $\times$  environment interaction for wheat yield in different drought stress conditions and agronomic traits suitable for selection. *Aust J. Agric. Res.* **59**, 536. <https://doi.org/10.1071/AR07281> (2008).
58. Ojeda, J. J. et al. Effects of soil- and climate data aggregation on simulated potato yield and irrigation water requirement. *Sci. Total Environ.* **710**, 135589. <https://doi.org/10.1016/j.scitotenv.2019.135589> (2020).
59. Angulo, C. et al. Characteristic 'fingerprints' of crop model responses to weather input data at different spatial resolutions. *Eur. J. Agron.* **49**, 104–114. <https://doi.org/10.1016/j.eja.2013.04.003> (2013).
60. Reynolds, M. P. et al. A wiring diagram to integrate physiological traits of wheat yield potential. *Nat. Food.* **3**, 318–324. <https://doi.org/10.1038/s43016-022-00512-z> (2022).

## Acknowledgements

The study was supported by the International Wheat Yield Partnership (IWYP, grant IWYP115) through funding from the International Maize and Wheat Improvement Center (CIMMYT) and from the National Institut for Agricultural Research and Environment (INRAE) to PM. EER acknowledges support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, project no. 520102751). FE also acknowledges support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2070–390732324. Open Access funding enabled and organized by Projekt DEAL.

## Author contributions

E.E.R: investigation, formal analysis, methodology, visualization, writing – original draft. B.F: investigation, formal analysis, methodology, writing – editing. F.E, S.A and P.M: conceptualization, writing – editing. H.W: conceptualization, supervision, methodology, writing – editing.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Declarations

### 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/s41598-024-74309-4>.

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

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024