The Journal of Agricultural Science

cambridge.org/ags

Crops and Soils Research Paper

Cite this article: Anwar MR, Emebiri L, Ip RHL, Luckett DJ, Chauhan YS, Zeleke KT (2024). Least absolute shrinkage and selection operator regression used to select important features when predicting wheat yield from various genotype groups. *The Journal of Agricultural Science* **162**, 245–259. https://doi.org/10.1017/S0021859624000479

Received: 11 March 2024
Revised: 9 July 2024
Accepted: 11 July 2024
First published online: 6 November 2024

Keywords:
crop modelling; heat stress; shrinkage estimator; *Triticum aestivum*; water stress

Corresponding author:
Muhuddin Rajin Anwar;
Email: muhuddin.anwar@dpi.nsw.gov.au

Least absolute shrinkage and selection operator regression used to select important features when predicting wheat yield from various genotype groups

Muhuddin Rajin Anwar[1,2], Livinus Emebiri[1], Ryan H. L. Ip[3,4], David J. Luckett[2], Yashvir S. Chauhan[5] and Ketema T. Zeleke[2,6]

[1]NSW Department of Primary Industries, Wagga Wagga Agricultural Institute, Wagga Wagga, NSW, Australia; [2]Gulbali Institute (Agriculture, Water and Environment), Charles Sturt University, Wagga Wagga, NSW, Australia; [3]Department of Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand; [4]School of Computing and Mathematics, Charles Sturt University, Wagga Wagga, NSW, Australia; [5]Department of Agriculture and Fisheries (DAF), Kingaroy, QLD, Australia and and [6]School of Agricultural, Environmental and Veterinary Sciences, Charles Sturt University, Wagga Wagga, NSW, Australia

Abstract

Bread wheat and durum wheat genotypes were grown in field experiments at two locations in New South Wales, Australia across several years and using two sowing times ('early' *v.* 'late'). Genotypes were grouped based on genetic similarity. Grain yield, grain size, soil characteristics and daily weather data were collected. The weather data were used to calculate water and heat stress indices for four key growth periods around flowering. Least absolute shrinkage and selection operator (LASSO) was used to predict grain yield and to identify the most influential features (a combination of index and growth period). A novel approach involving the crop water supply–demand ratio effectively summarized water relations during growth. LASSO predicted grain yield quite well (adjusted R^2 from 0.57 to 0.98), especially in a set of durum genotypes. However, the addition of other important variables such as lodging score, disease incidence, weed incidence and insect damage could have improved modelling results. Growth period 2 (30 days pre-flowering up to flowering) was the most sensitive for yield loss from heat stress and water stress for most features. Although one group of bread wheat genotypes was more sensitive to water stress (drought) in period 3 (20 days pre-flowering to 10 days post-flowering). Evapotranspiration was a significant positive feature but only in the vegetative phase (pre-flowering, period 1). This study confirms the usefulness of LASSO modelling as a technique to make predictions that could be used to identify genotypes that are suitable candidates for further investigation by breeders for their stress-tolerance ability.

Introduction

Wheat (*Triticum aestivum* L.) is the world's third most abundant staple cereal food crop, after maize and rice in terms of production. The Organization for Economic Co-operation and Development and Food and Agriculture Organization of the United Nations (OECD/FAO, 2023) report that its annual output for the 2022–23 period was approximately 800 million tonnes. In Australia, wheat is the fifth most exported commodity (OEC, 2023). However, climate change-induced abiotic stresses, such as drought and increased temperatures, pose significant challenges to wheat production (Collins and Chenu, 2021). Drought occurs when the potential evapotranspiration (ET) is higher than usual for a particular production environment (Langridge and Reynolds, 2021), and these areas are expanding due to climate change impacting wheat yields globally (Lobell *et al.*, 2011). In the last 40 years, drought-sensitive areas have ballooned by an area roughly equivalent to the size of South Africa (1.2 million km^2), according to a new study by Li *et al.* (2024). Additionally, water scarcity significantly hinders wheat cultivation in Australia, prompting growers, breeders and agronomists to focus on improving water-use efficiency (WUE) (Sadras and McDonald, 2012). Drought conditions, particularly those experienced in eastern Australia from 2017 to 2020, can severely restrict wheat yields (NSW Government Water, 2020; PDI, 2022). Durum wheat (*Triticum turgidum* L. subsp. *durum* (Desf.) Husn.) is a secondary wheat crop in Australia – grown for use in pasta production (GRDC GrowNote, 2017) with an annual production in Australia of around 0.5 million tonnes – much less than the total bread wheat production of about 36 million tonnes in 2022 (ABS 2024).

High temperatures during critical crop development stages, such as flowering, can reduce grain yield by directly affecting grain number and grain weight (Stone and Nicolas, 1994; Talukder *et al.*, 2014; Trethowan, 2022). Even a short period of high temperature during flowering can significantly reduce grain weight and set, especially in sensitive cultivars (Talukder *et al*., 2013). For example, a field experiment by Nuttall *et al*. (2012) showed that a temperature of 36–38°C for 6 days after flowering resulted in a 12% reduction in grain number and a 13% loss in grain yield. Additionally, environmental factors like water shortages and high temperatures significantly impact global wheat production through plant phenotypic and physiological changes (Abhinandan *et al*., 2018). Studies indicate that for each additional degree of global mean temperature increase, wheat yields could decline by up to 6%, under the assumption of no CO_2 fertilization, continued effective management practices and no changes in crop genetics (Asseng *et al*., 2015; Zhao *et al*., 2017). This impact has been already evident in Australia, where simulations suggest a huge and concerning 27% decline in water-limited potential wheat yield from 1990 to 2015 (Hochman *et al*., 2017). This decrease is likely attributable to a combination of stressors such as seasonal rainfall and increased temperatures, coupled with the limited ability of increased atmospheric concentration (CO_2) to fully compensate for these negative factors (Wang *et al*., 2017; Li *et al*., 2022). While climate change and climate variability present major hurdles, analysing the connections between climate, soil and wheat yield empowers the scientific community to design actionable strategies to mitigate yield losses. By unravelling the relative importance of various variables, we can guide future research towards developing climate-resilient wheat cultivars and innovative management practices, ultimately transforming vulnerability into opportunity.

In any individual wheat crop, in addition to climatic and edaphic influences, many other biotic and abiotic stresses will affect crop growth and yield. In this work, we were particularly interested in abiotic soil water deficit and heat stresses, especially during the reproductive period of the crop's growth. We explored how we can use soil characteristics and weather variables to improve the prediction of phenology and yield via modelling. In addition, we wanted to identify the most influential variables and the most sensitive crop growth period upon which these variables act. Least absolute shrinkage and selection method (LASSO) has shown good utility for identifying the most influential variable in a multiple-regression situation of this type (Didari *et al*., 2023). This methodology may assist wheat breeders by identifying different influential variables depending on the type of wheat genotype being examined. Lohithaswa *et al*. (2022) used LASSO for genetic selection in maize breeding against abiotic and biotic stresses. In addition, if other traits are of interest, the same approach can be used to dissect the influential variables (Shafiee *et al*., 2021). This study aimed to achieve the following objectives:

- Utilize an existing wheat data set encompassing six sets of bread wheat and durum wheat genotypes (with varying genotype numbers) grown in multi-year experiments across two sites and up to two sowing times. These data were used to investigate the possibility of predicting grain yield from a suite of weather- and soil-based climatic variables, particularly focusing on crop water-use, crop water stress, and crop heat stress.
- Leverage the relatively new LASSO regression technique to identify the most influential and effective variables for yield prediction. Additionally, this analysis aimed to determine if the genotype groups exhibited differential responses to these variables.
- Calculate weather and soil-based variables across four distinct crop growth periods (overlapping developmental stages) and assess their influence on yield prediction.
- By integrating the findings, this study sought to identify the most critical growth period for crop damage (yield loss) caused by water stress and/or heat stress within the different genotype groups. This information can inform wheat breeders on which crop traits require focus to minimize potential yield losses due to water and heat stress.

Materials and methods

Study area and soil data

Two typical rainfed crop-livestock growing locations (Leeton and Wagga Wagga) in southeast Australia, encompassing different climatic conditions, were selected for analysis (Table 1) utilizing an existing data set previously published to determine suitable planting windows that minimize the impact of environmental stress (Sissons *et al*., 2018; Zeleke *et al*., 2023). The soils across the sites are predominantly Wunnamurra clay (Leeton) and kandosol (Wagga Wagga) according to the Australian Soil Classification (Isbell and National Committee on Soil and Terrain, 2021). The soil properties at these sites have been summarized by others (Wang *et al*., 2017; Xing *et al*., 2017). Briefly, at Leeton, the plant available water capacity (PAWC) was 293 mm, to a total soil depth of 1.8 m, pH (1:5 water) ranged from 7.2 to 8.9, bulk density (g/cm^3) of 1.20–1.40 and initial nitrate (NO_3) was 81 kg/ha. At Wagga Wagga, the PAWC was 128 mm, to a total soil depth of 1.25 m, pH (1:5 water) ranged from 6.2 to 6.9, bulk density was 1.37–1.56 g/cm^3 and initial NO_3 was 69 kg/ha. These sites have been the subject of variable verification in wheat cropping systems (Anwar *et al*., 2015, 2022) especially for the numerous initial values and variables required for running the Agricultural Production Systems sIMulator (APSIM) crop growth model (https://www.apsim.info/).

Field experiments and agronomy

The layout of the field experiments, the details of the genotypes used and the agronomy used during each year are detailed in previous publications (Sissons *et al*., 2018; Zeleke *et al*., 2023). Briefly, the experiments were conducted at Leeton in 2011 and 2015 and at Wagga Wagga in 2012, 2018 and 2019. Two sowing times were used at each site/year: 'early' and 'late', in order to the maximize the differences in the weather experienced by the crops. Following standard rates used for local irrigated wheat, 100 mm of irrigation was applied to minimize drought stress and/or to facilitate timely sowing. Not all genotype sets were grown in every site/year. Here we note, in addition, that some lodging occurred in the field along with some fungal disease; the genotypes were variously affected but the scoring used was unfortunately inconsistent. Consequently, in this analysis these factors were not included in the LASSO modelling (see below) and may have contributed to some imprecision in the predicted values. We note that lodging may also have negatively impacted the recovery of grain due to the use of mechanical plot harvesting.

We note that the number of genotypes is not the same between the two sowing times within a genotype group (category), although there was a considerable overlap. The frequencies of concurrence of genotypes across 'site_year_sowing-time' are given in Table S1 in the Supplementary materials. This was due to practical issues, such as the lack of seed supply. The 'BreadWheat_NILines' group was only sown once, while the

Table 1. Characterization of the soils, site descriptions and long-term (1950–2022) average climate variables for the two sites used in the field experiments described in this paper

Site	Latitude Longitude	Soil type	Soil profile: Soil depth (m)	Soil profile: Total PAWC (mm)	Soil profile: NO_3 (kg/ha)	Top soil layer[a]: BD (g/cm^3)	Top soil layer[a]: pH (1:5 water)	Top soil layer[a]: OC (%)	Rain (mm)	GS rain (mm)	Solar (MJ/m^2)	GS Solar (MJ/m^2)	maxT GS (°C)	minT GS (°C)	avT GS (°C)	Frost GS (days)
Leeton	34.73 S 146.55 E	Wunnamurra clay (clay loam)	1.80	293	81	1.30	7.20	1.75	478 (32.5)[b]	293 (37.3)	17.7 (5.25)	13.2 (6.78)	18.5 (4.39)	6.3 (9.57)	12.4 (3.71)	22 (44.4)
Wagga Wagga	35.05 S 147.35 E	Kandosol	1.25	128	69	1.45	6.43	1.69	560 (29.4)	339 (34.6)	17.3 (5.25)	12.8 (6.79)	17.4 (5.17)	5.7 (11.89)	11.6 (4.02)	28 (42.6)

PAWC, plant available water capacity; BD, bulk density; OC, soil organic carbon; GS, growing season (April–October); maxT, mean annual maximum temperature; minT, mean annual minimum temperature; avT, mean annual average temperature; Frost, mean annual number of days where minT ≤ 0°C.
[a]Top soil layer equals 0–10 cm.
[b]Coefficient of variation (%) in parentheses.

'Durum_Elite' category had only a small number of genotypes. Both of these groups were excluded from the LASSO analysis.

Wheat genotype groups

ABD lines

These are advanced breading lines of wheat (*T. aestivum*) produced at the International Maize and Wheat Improvement Centre, Mexico. They are comprised of the line selections from the high-temperature wheat yield trials, and selections made for their large grain size (Sissons *et al.*, 2024). Hereafter referred to as 'BreadWheat_ABDLines'.

Elite wheat

These are bread wheat varieties of historical significance, recently released cultivars and parents used in breeding programmes by the major private breeding companies (InterGrain, LongReach and Australian Grain Technologies) in Australia. These varieties have been bred to meet the specific needs of Australian growers, such as resistance to diseases and pests, tolerance to heat and drought, good grain quality and high yield, and were chosen based on being potentially heat tolerant (or in some cases intolerant) according to Australian breeder recommendations and the literature (Sissons *et al.*, 2024). Hereafter referred to as 'BreadWheat_Elite'.

Landrace wheat

The bread wheat landraces were sourced from heat-prone areas in Afghanistan, Iran, Iraq and India. They were identified using the focused identification of genotype strategy (FIGS), an approach that uses environmental variables described in plant genotype collection sites as selection criteria to identify materials that most likely have undergone selection pressures for the target variables (Sissons *et al.*, 2018). Hereafter referred to as 'BreadWheat_Landraces'.

Tamaroi × Saintly durum bi-parent density

These are durum wheat double-haploids, which were produced from F1 plants of a cross between the SA-bred variety, Saintly and the NSW-bred variety, Tamaroi. Saintly has a reputation for performing well in seasons with terminal drought stress, while the variety Tamaroi has a very high inherent 1000-kernel weight but is susceptible to heat stress. Hereafter referred to as 'Durum_Biparent'.

Durum elite

The durum wheat genotype comprised of a worldwide collection trialled for heat tolerance in southern Australia (Collins *et al.*, 2017). They included commercial durum varieties and breeding lines, along with tetraploid wheat landraces sourced from heat-prone regions by using the FIGS (Street *et al.*, 2016). They have been shown to exhibit significant variability for tolerance/intolerance to late-sown heat stress (Sissons *et al.*, 2018) and natural heat waves (Emebiri *et al.*, 2024). Hereafter referred to as 'Durum_Elite'.

Bread wheat near-isogenic lines

The near-isogenic (NI) lines were created from a cross of wheat varieties Drysdale and Waagan. Both parents are semi-dwarf varieties and carry genetic loci for intolerance and tolerance, respectively, to both booting and grain filling stage heat stress (Shirdelmoghanloo *et al.*, 2016; Erena *et al.*, 2021). The NI lines were created by using molecular markers to identify single Drysdale × Waagan $F_{2:8}$ plants that were heterozygous for genetic loci located on wheat chromosomes 2B, 3B and 6B; then the

https://doi.org/10.1017/S0021859624000479 Published online by Cambridge University Press

progeny of these plants was screened to identify plants homozygous for each allele at the respective loci (Erena *et al.*, 2021). Hereafter referred to as 'BreadWheat_NILines'.

Soil water balance

Rainfall, ET, runoff and drainage are key factors that affect how much water is available to crops (Unkovich *et al.*, 2018, 2023). In this study, temperature, rainfall, irrigation, simulated initial soil water content and simulated soil water content at harvest were used to calculate the soil water balance following the procedure of He and Wang (2019). We used the pre-validated APSIM (version 7.10) that simulates the key biophysical processes related to crop growth and production, water, carbon and N cycling in the soil–plant system (Holzworth *et al.*, 2014). Published studies have also used the APSIM model to calculate hydraulic variables for wheat cropping systems (such as, soil water content at sowing and harvest, water use [WU], runoff, drainage and soil evaporation). The variables in the Soil Water module of APSIM were the same for our sites as those used in other published work (Liu *et al.*, 2014; Wang *et al.*, 2017; Xing *et al.*, 2017; Zeleke and Nendel, 2019). To estimate the initial soil water content at the start of the experimental period (2011), we assumed that the starting soil water on 1 January 2002 was equal to LL15 (water content at 15 bar suction). Then by running APSIM for the 2002–11 period using actual weather data we simulated the initial soil water in 2011 (He and Wang, 2019). The LL15 value was determined in the Wagga Wagga Agricultural Institute soil moisture analysis laboratory (Anwar *et al.*, 2022). The APSIM model was then run continuously until the end of the experimental period (31 December 2019), without resetting soil water conditions, to obtain the 'initial soil water at sowing' and the 'soil water at harvest' for each of the wheat experiments (Sissons *et al.*, 2018). The APSIM crop sequence used in the 10-year run-up period before the wheat experiments commenced was a typical one used in the wheat growing regions in Australia: wheat(W)-canola(C)-chickpea (CP)-W-C-CP-W-C-CP-W.

Total crop WU expressed as ET was calculated by subtracting the final soil water content at harvest from the initial soil water content at sowing and adding the amount of irrigation and rainfall received during the growing season:

$$\mathrm{ET} = \mathrm{P} + \mathrm{I} + \mathrm{SWs} - \mathrm{SWh} - \mathrm{R} - \mathrm{D} \tag{1}$$

where P, I, R and D are cumulative rainfall, irrigation, runoff and deep drainage from the day of sowing to harvest, and SWs and SWh are soil water at the sowing and harvest dates, respectively (Yang *et al.*, 2016).

In contrast, transpiration (T), which does not include soil evaporation (E) (Eqn (1)), was calculated using the following soil water balance equation (Yang *et al.*, 2016):

$$\mathrm{T} = \mathrm{P} + \mathrm{I} + \mathrm{SWs} - \mathrm{SWh} - \mathrm{R} - \mathrm{D} - \mathrm{E} \tag{2}$$

The APSIM soil water module also calculates daily potential ET using the Priestley–Taylor method (Priestly and Taylor, 1972; APSIM, 2023), which is based on the physiological relationship between crop yield and ET (Paredes *et al.*, 2014; Trout and DeJonge, 2017; Akumaga and Alderman, 2019).

Water supply–demand ratio (SDR)

The APSIM model calculates a water-deficit index (Chapman *et al.*, 1993; Chenu *et al.*, 2011), also known as the 'water supply' and 'water demand' ratio, which indicates how well the water extractable by the crop's roots (water supply) meets the crop's potential transpiration (water demand). The crop water supply is calculated for each layer of the soil where roots are present and depends on the root growth and soil property of each layer. The water demand is the amount of water the crop would have transpired in the absence of soil water constraint. It is estimated daily based on the amount of crop growth on that day and the atmospheric saturation vapour pressure deficit.

Water SDR is the ratio between water supply and water demand, bounded between 0 and 1, which indicates if the plant is water-stressed:

$$\mathrm{SDR} = \begin{cases} \min\left(\dfrac{\mathrm{Supply}}{\mathrm{Demand}}, 1\right), & \mathrm{Demand} > 0 \\ 1, & \mathrm{Demand} = 0 \end{cases} \tag{3}$$

When SDR = 1, there is no water stress. Otherwise, the plant is stressed. Based on SDR, we define water deficiency (D) such that:

$$D = 1 - \mathrm{SDR} \tag{4}$$

The interpretation is the opposite of SDR. When $D = 0$, there is no water stress. Positive D indicates stress. Daily deficiency values were calculated and were accumulated within the following four crop development periods, each spanning approximately 30 days (see below).

Wheat developmental period

Abiotic stress during the reproductive stage of plants (anthesis and grain filling) has a significant effect on grain yield and quality. The critical period for abiotic stress is the time when plants are most sensitive to these stresses. Some previous studies have defined the critical period as 30 or 45 days before to 0 days after 50% anthesis (Fischer, 1985). Other studies have found that the critical period is narrower, spanning only about 20 days before to 10 days after anthesis (Ortiz-Monasterio *et al.*, 1994; Abbate *et al.*, 1995). More recently, Slafer *et al.* (2023) found that the critical period for wheat is from 30 days before to 10 days after anthesis.

In this study, we defined and examined four contrasting crop growth periods based on previously published studies:

Period 1: from sowing to the day of flowering (varying lengths)
Period 2: from 30 days before flowering to the day of flowering (30 days total)
Period 3: from 20 days before flowering to 10 days after flowering (30 days total)
Period 4: from 15 days before flowering to 15 days after flowering (30 days total)

We chose these 30-day intervals based on the findings of previous studies (Fischer, 1985; Slafer *et al.*, 2023). There is considerable chronological overlap between these periods, but we wanted to test which of these periods might be most sensitive to stress effects on grain yield. By definition, period 2 overlaps with period 3 by 67%; period 2 overlaps with period 4 by 50% and period 4 overlaps with period 3 by 83%. The degree to which period 1 overlaps with the others depends on the interval from sowing to flowering (in days). The means and ranges for the sowing-to-flowering interval for each genotype group across each site/year/sowing-time combination are given in Table S2 in the Supplementary materials. The overall mean of this duration was 106.0 days. The mean overlap (and the ranges) between periods 1 and 2 was 28.8% (21.5–38.8%). For periods 1 and 3 the corresponding data were 19.2% (14.3–25.9%). For periods 1 and 4 they were 14.4% (10.8–19.4%).

Statistical techniques and LASSO

First, we examined several summary statistics for each genotype set and each sowing time (across both sites and all years). Since the 'early' (coded '1') and 'late' (coded '2') sowing times were designed to present the crops with contrasting stress environments, we expected to see quite large differences in means and ranges for the traits of interest.

Second, to investigate the impact of daily abiotic stress indices (heat stress, water deficit and ET) on wheat yield, accumulated over four key growth stages, we used the following approach.

Pearson correlation coefficients were calculated to assess the relationship between yield, 1000-grain weight (TGW), ET, transpiration (T), accumulated water deficit (water SDR) and heat stress (number of days with temperatures >30°C). Correlation analysis was restricted to period 3 only (20 days before flowering to 10 days after flowering, see above) because this flowering period has proven to be the most important with respect to yield in other published papers (see above). All data were normalized to zero mean and unit variance prior to analysis.

Third, to study the relationship between wheat yield (the target trait) and daily stress indices (the explanatory variables) accumulated over critical periods of growth (the four 'periods') and across different sets of genotypes groups, we undertook LASSO regression analysis. In linear models such as multiple linear regression models, it is often assumed that the explanatory variables are independent (Monahan, 2011). When explanatory variables are correlated, multicollinearity is said to exist (Kutner *et al*., 2005). As a result of multicollinearity, the estimation of coefficients can become unstable, leading to unreliable estimates. In some extreme cases, the regression coefficients do not reflect the inherent relationship between the explanatory variable and the response variable. For example, a negative coefficient may be obtained although the relationship should be positive.

For better interpretability, many statistical methods have been proposed to deal with multicollinearity, many of which are aimed at minimizing the prediction error while forcing (i.e. 'shrinking') some of the regression coefficients to zero, hence effectively removing some of the explanatory variables and highlighting the most influential ones (Dormann *et al*., 2013). Among these methods, LASSO is a popular choice. In this work, we used LASSO to find the best subset of explanatory variables from the large initial number. To obtain scientifically sensible regression coefficients, constrains were imposed on them in the estimation procedure. Specifically, the coefficients of variables related to heat and water stresses were set to be non-positive. The computations were performed using the 'glmnet' package in R (Friedman *et al*., 2010).

The 'tidyverse' R package (Wickham *et al*., 2019), the RStudio GUI (RStudio Team, 2023) and the R software suite (R Core Team, 2023) were used for data preparation, summarization and graphics.

Results

Data summaries across sites, genotype groups and sowing time

The Wagga Wagga soil, compared to Leeton, is a shallower and more dense soil, with lower pH, which holds much less water than Leeton (Table 1). Both sites face sizeable year-to-year variations in the climate variables (rain, solar and temperatures). Wagga Wagga gets more rain, both overall and during the growing season, with Wagga Wagga's temperatures being slightly cooler than Leeton.

While there was a large variation within each genotype category, the grain yield (Table 2) was always substantially lower in sowing-time_2 due to higher stress levels with an overall range of nearly 9 t/ha to less than 0.2 t/ha. Grain size was similarly reduced in sowing-time_2 except for the 'Durum_Biparent' category (Table 2).

The mean total transpiration and mean total ET were always higher in sowing-time_2 (Table 3) due to the crops growing in a hotter and drier period of the year, with ET always greater than T (as expected). WUE, both for transpiration (WUE_T) and evapotranspiration (WUE_ET), was much reduced in sowing-time_2 compared to sowing-time_1, often by more than 50%. The WUE_T ranged overall for individual genotypes from 57.5 to 0.92 kg of grain/ha/mm, whereas WUE_ET ranged from 33.5 to 0.6 kg of grain/ha/mm (Table 3).

Correlation analysis

Correlation coefficients between yield, TGW, ET, transpiration (T), accumulated water deficit (SDR = supply–demand ratio) and heat stress ($H > 30$ = number of days with temperatures >30°C) (see Tables 4–9). Generally, the correlations between traits were highly significant (either positively or negatively) within each genotype category but significance levels were much lower (or non-existent) in the 'BreadWheat_Landraces' and the two durum categories. The T and ET variables were always highly positively correlated (as expected). The $H > 30$ (index of heat stress) was usually highly positively correlated with both T and ET but was not significant in the 'BreadWheat_Landraces' category (Table 6) nor for ET in the 'Durum_Elite' category (although the number of values was small, $n = 10$, Table 9).

The TGW and grain yield were generally positively correlated but again not within the 'BreadWheat_Landraces' group (Table 6), and were strongly negative in the 'Durum_Biparent' material (Table 8). The SDR (index of water stress) was usually negatively correlated (when significant) with the other traits but a contrasting positive correlation was seen in the 'BreadWheat_ABDLines' material with TGW (Table 4), and with T and ET in the 'Durum_Biparent' material (Table 8).

T and ET were generally significantly negatively correlated with both yield and TGW, except in the 'BreadWheat_Landraces' group (Table 6) and in 'Durum_Biparent' (Table 8). As expected, the numerous genotypes in the 'BreadWheat_Landraces' category provided the greatest range in performance (yield and TGW), water use (T and ET) and water use efficiency (WUE_T and WUE_ET), plus less rigid inter-trait correlations.

Distributional characteristics of water and heat stress and ET

For each of the four growing periods in this study, the intensity and frequency of water stress (calculated via SDR) and heat stress including ET are summarized below for individual genotypes within six genotype groups.

Boxplots show the distribution of accumulated values of SDR for a single genotype group and growing period combination (Fig. 1). There is a considerable amount of variability in SDR within each genotype group and growing period. The boxes,

https://doi.org/10.1017/S0021859624000479 Published online by Cambridge University Press

Table 2. Mean grain yield and grain weight of six wheat genotype groupings (across two wheat species, bread wheat and durum) sown at one or two different times

Category	Sowing time	Number of genotypes	Mean sowing date (Julian day)	Mean yield (t/ha)	Range in yield (t/ha)	Mean TGW (g)	Range in TGW (g)
BreadWheat_ABDLines	1	72	156	5.66	3.17–8.96	41.9	26.80–50.38
BreadWheat_ABDLines	2	61	217	3.85	1.86–5.75	33.3	21.38–42.65
BreadWheat_Elite	1	217	156	5.33	2.13–8.73	39.1	27.92–52.25
BreadWheat_Elite	2	219	217	3.33	0.19–5.46	31.2	na
BreadWheat_Landraces	1	196	157	3.74	2.00–6.35	40.3	30.68–58.0
BreadWheat_Landraces	2	201	218	2.07	0.44–4.23	31.9	22.28–48.40
BreadWheat_NILines	1	61	137	2.62	2.30–2.92	33.6	25.52–39.30
Durum_Biparent	1	232	152	2.78	2.25–3.43	35.1	26.51–44.20
Durum_Biparent	2	322	216	1.48	0.88–2.03	40.4	31.82–47.76
Durum_Elite	1	4	152	4.73	2.59–7.22	42.0	38.34–45.30
Durum_Elite	2	6	215	2.49	0.86–4.61	39.8	32.05–44.95

The number of genotypes of each genotype category and the range of genotype means for grain yield (t/ha) and 1000-grain weight (TGW, g) are also presented.
na, not available.

which represent the interquartile range (IQR), span a wide range of values in most cases. The whiskers, which extend to the most extreme data points not considered outliers, also show a wide range of values for many of the groups and periods. The medians (represented by the horizontal lines within the boxes) are generally lower for groups and periods with higher SDR. For example, in the first growing period, the median SDR for the BreadWheat_Elite group is around 10, while the median SDR for the same group at third growing period decreased to about 4.8. The dispersion, or spread, of the data is also influenced by SDR. The boxes tend to be wider for groups and periods with higher SDR, indicating that there is a greater range of SDR values within those groups. For example, in the second growing period, the box for the BreadWheat_Elite group is wider than the third growing period. In contrast, the BreadWheat_NILines group didn't show dispersion but higher SDR values in the first and fourth growing periods compared to the second and third growing periods. SDR decreases from growing period 1 to growing period 4 for the genotypes in all the groups (Fig. 1). For three of the genotype groups (BreadWheat_ABDLines, BreadWheatNILines, Durum_Wheat), the lowest SDR is in growing period 2.

The potential ET (Fig. 2) exhibits considerable variability within each genotype group and growing period, as evidenced by the IQRs and whiskers of the boxes. The degree of dispersion in ET values differs across genotype groups and growing periods. For instance, Durum_Elite lines generally demonstrate a more compact distribution of ET values compared to BreadWheat_ABDLines, suggesting greater consistency in WU within the Durum_Elite group. The

Table 3. Water use and water-use efficiency based on transpiration (T and WUE_T) and evapotranspiration (ET and WUE_ET), respectively for six wheat genotype groupings sown at one or two different times

Category	Sowing time	Mean T (mm)	Range in T (mm)	Mean ET (mm)	Range in ET (mm)	Mean WUE_T (kg grain/ ha/mm)	Range in WUE_T (kg grain/ ha/mm)	Mean WUE_ET (kg grain/ ha/mm)	Range in WUE_ET (kg grain/ ha/mm)
BreadWheat_ABDLines	1	155	136–187	245	219–277	36.6	20.1–49.9	23.0	13.7–33.5
BreadWheat_ABDLines	2	222	194–255	308	272–346	17.3	7.8–23.1	12.5	5.8–16.8
BreadWheat_Elite	1	158	136–188	249	219–299	33.9	15.6–57.5	21.4	9.1–34.9
BreadWheat_Elite	2	222	187–258	310	262–363	15.2	0.92–24.6	10.9	0.6–17.6
BreadWheat_Landraces	1	157	136–188	260	219–313	24.5	13.7–46.6	14.6	7.9–28.2
BreadWheat_Landraces	2	220	191–258	313	268–370	9.7	1.69–21.7	6.8	1.2–15.2
BreadWheat_NILines	1	159	158–159	269	265–270	16.5	14.4–18.4	9.7	8.5–10.8
Durum_Biparent	1	146	138–156	257	239–287	18.9	15.8–21.9	10.8	9.1–12.1
Durum_Biparent	2	189	141–232	279	223–332	7.8	5.7–9.9	5.3	3.7–6.4
Durum_Elite	1	163	142–186	258	245–275	27.8	18.3–38.9	17.9	10.6–26.3
Durum_Elite	2	211	151–254	299	236–342	11.1	5.7–18.2	7.9	3.6–13.6

Overall mean values are presented along with the corresponding ranges for individual genotype means.

Table 4. Pearson correlation coefficients between yield, TGW, ET, transpiration (T), accumulated water deficit (SDR = supply–demand ratio) and heat stress ($H > 30$ = number of days with temperatures >30°C) during growth period 3 (see text for explanation) in the 'BreadWheat_ABDLine' genotype category for both sites (Wagga Wagga and Leeton) combined, including all sowing times

	BreadWheat_ABDLines genotype group					
	Yield	TGW	ET	T	SDR	$H > 30$
Yield		0.52***	−0.38***	−0.37***	−0.15**	−0.62***
TGW			−0.61***	−0.62***	0.16**	−0.60***
ET				0.99***	−0.59***	0.70***
T					−0.60***	0.68***
SDR						−0.23***

Significance levels are indicated as follows: *0.01 < P < 0.05, **0.001 <P < 0.01, ***P < 0.001.

Table 5. Pearson correlation coefficients between yield, TGW, ET, transpiration (T), accumulated water deficit (SDR = supply–demand ratio) and heat stress ($H > 30$ = number of days with temperatures >30°C) during growth period 3 (see text for explanation) in the 'BreadWheat_Elite' genotype category for both sites combined, including all sowing times

	BreadWheat_Elite genotype group					
	Yield	TGW	ET	T	SDR	$H > 30$
Yield		0.59***	−0.62***	−0.66***	0.20***	−0.73***
TGW			−0.28***	−0.37***	−0.11*	−0.49***
ET				0.97***	−0.61***	0.82***
T					−0.55***	0.77***
SDR						−0.29***

Significance levels are indicated as follows: *0.01 < P < 0.05, **0.001 < P < 0.01, ***P < 0.001.

Table 6. Pearson correlation coefficients between yield, TGW, ET, transpiration (T), accumulated water deficit (SDR = supply–demand ratio) and heat stress ($H > 30$ = number of days with temperatures >30°C) during growth period 3 (see text for explanation) in the 'BreadWheat_Landraces' genotype category for both sites combined, including all sowing times

	BreadWheat_Landraces genotype group					
	Yield	TGW	ET	T	SDR	$H > 30$
Yield		−0.10	0.24	0.27	−0.93***	−0.85**
TGW			−0.59	−0.56	0.25	0.26
ET				0.99***	−0.33	−0.18
T					−0.32	−0.13
SDR						0.97***

Significance levels are indicated as follows: *0.01 < P < 0.05, **0.001 < P < 0.01, ***P < 0.001.

Table 7. Pearson correlation coefficients between yield, TGW, ET, transpiration (T), accumulated water deficit (SDR = supply–demand ratio) and heat stress ($H > 30$ = number of days with temperatures >30°C) during growth period 3 (see text for explanation) in the 'BreadWheat_NILines' genotype category for both sites combined, including all sowing times

	BreadWheat_NILines genotype group					
	Yield	TGW	ET	T	SDR	$H > 30$
Yield		0.57***	−0.18*	−0.20*	−0.40***	−0.30***
TGW			−0.31***	−0.34***	−0.34***	−0.07
ET				0.99***	−0.50***	0.36**
T					−0.48***	0.30***
SDR						−0.40***

Significance levels are indicated as follows: *0.01 < P < 0.05, **0.001 < P < 0.01, ***P < 0.001.

Table 8. Pearson correlation coefficients between yield, TGW, ET, transpiration (T), accumulated water deficit (SDR = supply–demand ratio) and heat stress ($H > 30$ = number of days with temperatures >30°C) during growth period 3 (see text for explanation) in the 'Durum_Biparent' genotype category for both sites combined, including all sowing times

	Durum_Biparent genotype group					
	Yield	TGW	ET	T	SDR	$H > 30$
Yield		0.16	−0.06	−0.06	0.13	
TGW			0.06	0.03	0.05	
ET				0.90***	0.33**	
T					0.34**	
SDR						

Significance levels are indicated as follows: *0.01 < P < 0.05, **0.001 < P < 0.01, ***P < 0.001.

median ET values vary across genotype groups and growing periods. Notable trends include BreadWheat_ABDLines tend to have higher median ET values than other groups across most growing periods. Durum Elite lines generally exhibit lower median ET values, particularly in growing periods 2 and 3. BreadWheat_Landraces display a wider range of median ET values across growing periods. The potential ET of growing period 1 > 4 > 3 > 2. Growing period 1 has the highest potential ET and growing period 2 has the lowest potential ET. Compared to the other genotypes, BreadWheat_Landraces has the highest potential ET for each of the respective growing periods and high variability in ET across growing periods, suggesting that water-use strategies of genotypes may vary depending on environmental conditions and crop developmental stages.

There was a wide range of variability in heat stress within each genotype group and growing period (Fig. 3). The boxes show the middle 50% of the data, with the whiskers extending to the 10th and 90th percentiles. For example, in the BreadWheat_ABDLines group, the heat stress ranges from 0 to 15 days across the growing period. The median heat stress is also different for each genotype group and growing period. For example, the median heat stress for the BreadWheat_ABDLines group is about 7 days in the third growing period, while the median heat stress for the Durum_Biparent group is higher (about 15 days) in the same growing period. Figure 3 shows that heat stress growing period 1 < 2 < 3 < 4. Growing period 1 has the lowest heat stress and growing period 4 has the lowest potential ET. Durum_Biparent had the highest stress for a given growing period compared to the other genotypes.

LASSO feature selection

Our wheat data consisted of six genotype categories (Table S2 in the Supplementary materials); however, when fitting a LASSO model, specific criteria must be met. In our case, the 'BreadWheat_NILines' category only had one sowing time at one location; hence there was no variation in the explanatory variables, and this category was excluded from the final modelling. Similarly, the 'Durum_Elite' category had too few observations (two genotypes only) to allow the fitting of the explanatory variables. The interpretation of coefficients from LASSO is almost the same as in multiple regression models. The only difference is that LASSO 'forces' some of the coefficients to zero. Table 10 shows the estimated coefficients from LASSO and overall model performance.

In all four major genotype categories, the effect of ET on yield was effectively zero except in period 1 (where it presumably influenced vegetative biomass, which led to more yield), and in period 4 for 'Durum_Biparent' genotypes. Notably, in period 1, the effect of ET on yield was higher for 'BreadWheat_ABDLine' and 'BreadWheat_Elite' compared to the other two genotype groups. Heat stress (H) was damaging in all periods for the first two bread wheat categories and the 'Durum_Biparent' set, but less so for the 'BreadWheat_Landraces' set in period 3. Yet, heat stress in period 2 was found to be highly damaging for the 'BreadWheat_Landraces' group. The results were more mixed for the water stress index (D), particularly detrimental in 'BreadWheat_ABDlines' and in period 3.

For BreadWheat_ABDLines (Table 10), wheat grain yield was found to be most severely affected by water stress in period 3, followed by heat stress in period 1. For each unit increase in water stress in period 3, yield is expected to decrease by 0.789 t/ha, assuming all other factors remain unchanged. Water stress in period 4 and ET in periods 2–4 were found to be relatively less influential to the grain yield. In the 'BreadWheat_Landraces' genotype category, heat stress during period 2 had the strongest negative impact on wheat grain yield, reducing it by 0.725 t/ha. In contrast, water stress and ET in all periods had minimal to no effect on grain yield. Among the genotype categories in period 2, BreadWheat_Elite experienced the greatest yield reduction due to water stress, with an expected decrease of 0.501 t/ha and heat stress followed closely (yield decline of about 0.449 t/ha). Notably, ET had no impact on grain yield for BreadWheat_Elite in periods 2–4. Durum_Biparent appears to be less sensitive to water stress and heat stress than BreadWheat_ABDLine. In Durum_Biparent, heat stress had the greatest impact in period 2, with an expected

Table 9. Pearson correlation coefficients between yield, TGW, ET, transpiration (T), accumulated water deficit (SDR = supply–demand ratio) and heat stress ($H > 30$ = number of days with temperatures >30°C) during growth period 3 (see text for explanation) in the 'Durum_Elite' genotype category for both sites combined, including all sowing times

	Durum_Elite genotype group					
	Yield	TGW	ET	T	SDR	$H > 30$
Yield	1.00	−0.75***	0.18***	−0.16***	−0.22***	−0.87***
TGW		1.00	−0.17***	0.11*	0.08	0.62***
ET			1.00	0.93***	−0.07	0.05
T				1.00	−0.10*	0.31***
SDR					1.00	0.57***

Significance levels are indicated as follows: *0.01< P < 0.05, **0.001 < P < 0.01, ***P < 0.001.

https://doi.org/10.1017/S0021859624000479 Published online by Cambridge University Press

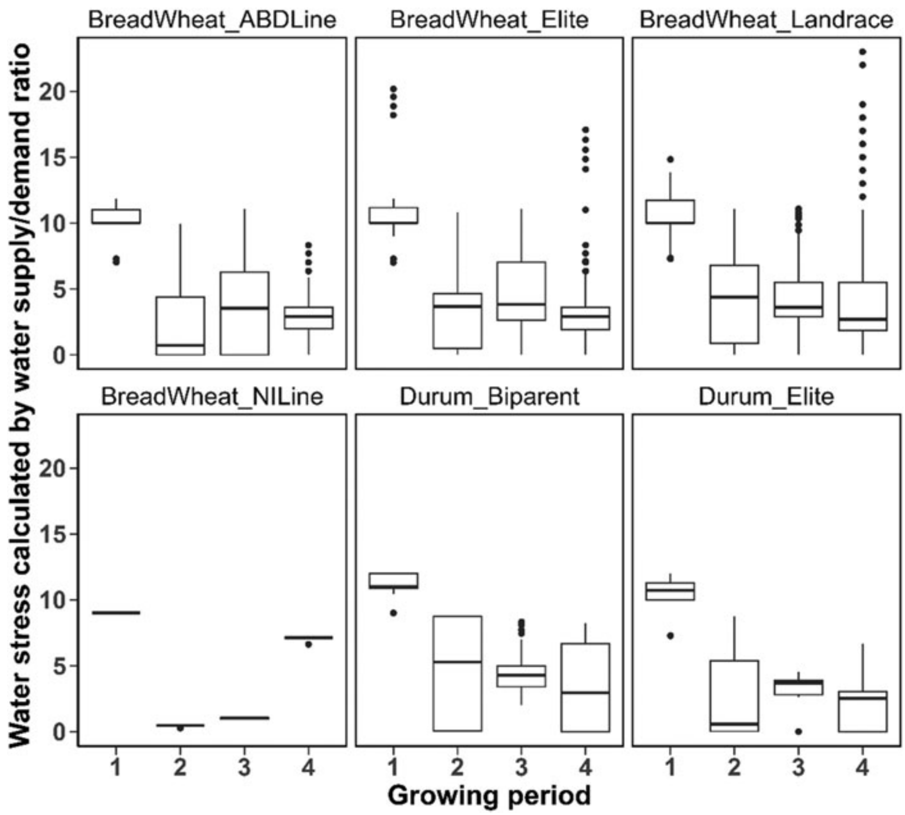

Figure 1. Boxplots of mean SDR (supply–demand ratio = water stress) for individual genotypes within six genotype groups and for each of the growing periods (period 1 = from sowing to the day of flowering; period 2 = from 30 days before flowering to the day of flowering; period 3 = from 20 days before flowering to 10 days after flowering and period 4 = from 15 days before flowering to 15 days after flowering, see text for details). Data include all sowing times. The number of genotypes in each group and sowing time is given in Table 3.

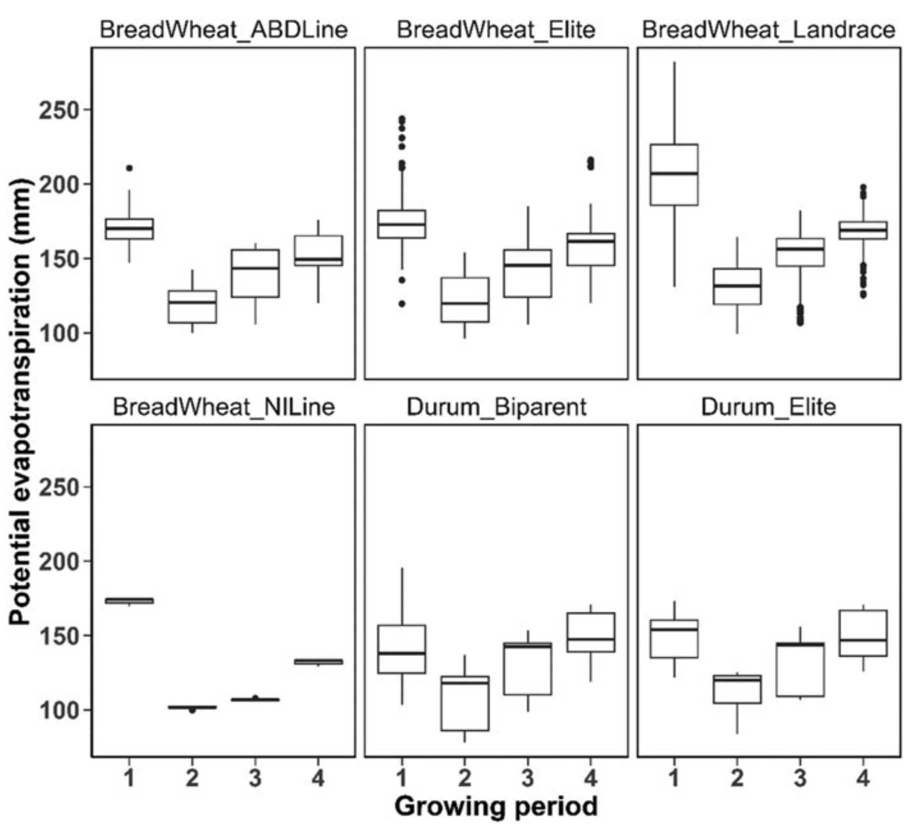

Figure 2. Boxplots of potential ET for individual genotypes within six genotype groups and for each of the growing periods (period 1 = from sowing to the day of flowering; period 2 = from 30 days before flowering to the day of flowering; period 3 = from 20 days before flowering to 10 days after flowering and period 4 = from 15 days before flowering to 15 days after flowering, see text for details). Data include all sowing times. The number of genotypes in each group and sowing time is given in Table 3.

https://doi.org/10.1017/S0021859624000479 Published online by Cambridge University Press

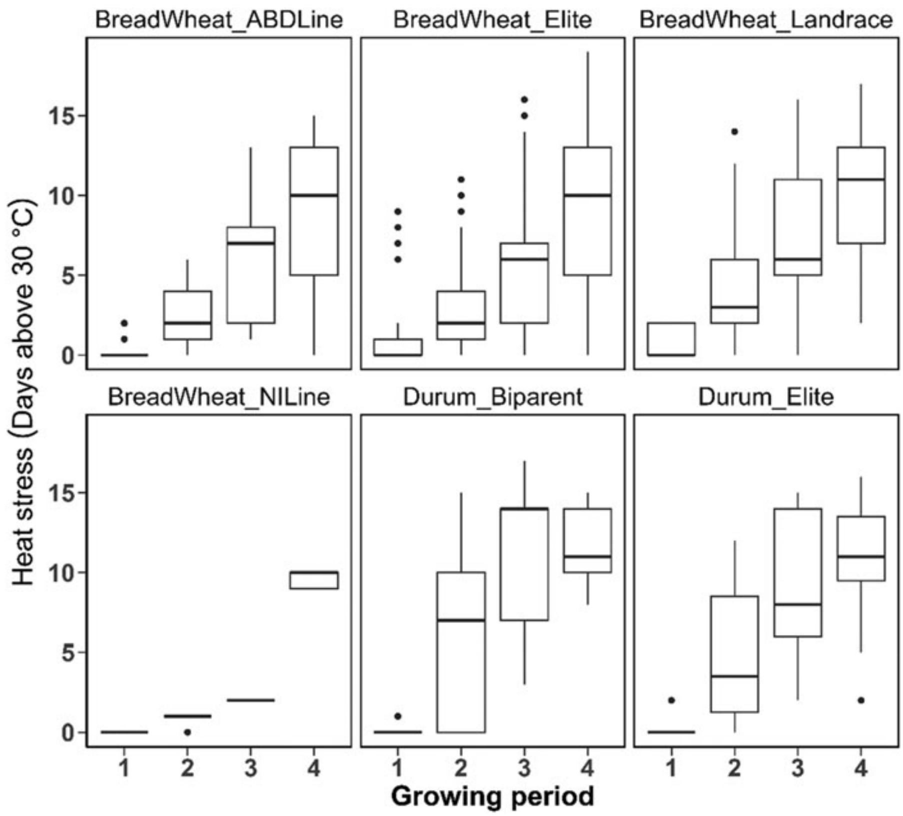

Figure 3. Boxplots of heat stress for individual genotypes within six genotype groups and for each of the growing periods (period 1 = from sowing to the day of flowering; period 2 = from 30 days before flowering to the day of flowering; period 3 = from 20 days before flowering to 10 days after flowering and period 4 = from 15 days before flowering to 15 days after flowering, see text for details). Data include all sowing times. The number of genotypes in each group and sowing time is given in Table 3.

yield decrease of 0.361 t/ha. This was followed by water stress with a decrease of 0.241 t/ha. ET had a positive effect on grain yield in periods 1 and 4, with increases ranging from 0.224 to 0.278 t/ha. However, it had no impact on yield in periods 2 and 3.

Yield prediction

LASSO modelling predicted yield reasonably well (Fig. 4) with highly significant positive regression between the observed and predicted values: the 'Durum_Biparent' relationship being particularly strong. There are some outlying groups of genotypes, for example in the 'BreadWheat_Elite' category but these were very low yielding genotypes. As shown in Table 4, the root mean squared errors (RMSE) ranged from 0.119 to 0.976 t/ha across the four genotypes and the adjusted R^2 ranged from 0.57 to 0.98. So, overall, the LASSO approach worked well at predicting crop outcomes, especially for 'Durum_Biparent', from weather-based and soil-based indices. Some other explanatory variables (not considered in this study) are required to improve further the goodness of fit, such as disease scores, lodging scores, weed measurements and crop plant density.

Discussion

Climate change throws a complex web of challenges at crop production, weaving together water deficits, scorching heat and fluctuating evaporative demands (Anwar *et al.*, 2015; Kerr *et al.*, 2022). These interwoven environmental stresses act like a multi-pronged attack, inflicting far more damage on plant growth and yield than individual stressors do in isolation (Pandey *et al.*, 2017). This 'synergistic effect' can significantly cripple crop production, exceeding initial projections, as evidenced by numerous studies (Mittler, 2006; Prasad *et al.*, 2011). Additionally, environmental factors like water shortages and high temperatures significantly impact global wheat production through plant

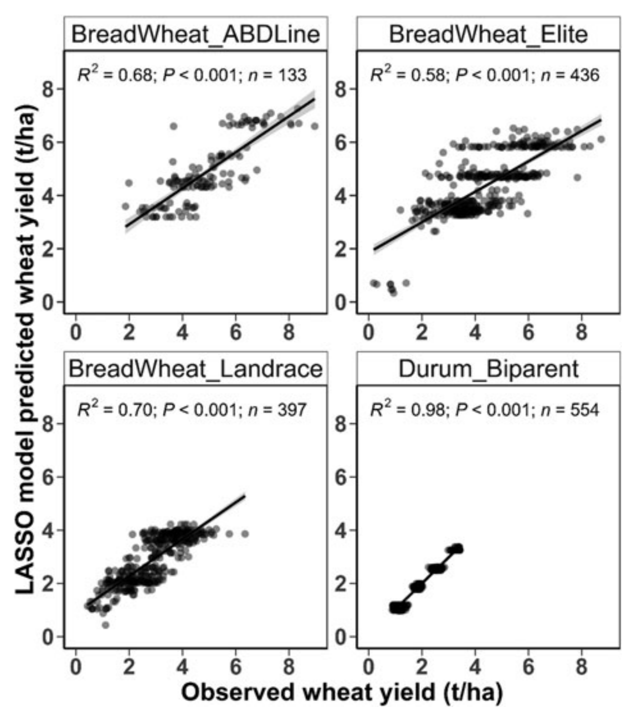

Figure 4. Plots of predicted wheat yield (t/ha) derived from LASSO modelling against observed yield for four wheat genotype categories. The best fitted straight line, together with the R^2 and P-value of the slope term, are provided for better visual assessment.

https://doi.org/10.1017/S0021859624000479 Published online by Cambridge University Press

Table 10. Estimated coefficients from LASSO and model performance for Leeton and Wagga Wagga (all sowing times pooled), as measured by estimated coefficients, RMSE, adjusted (Adj.) R^2 and lambda (λ)

BreadWheat_ABDLines		BreadWheat_Elite	
Item	LASSO coefficients	Item	LASSO coefficients
Intercept	4.834	Intercept	4.327
D_period1[a]	−0.498		−0.036
D_period2	−0.341		−0.501
D_period3	−0.789		0
D_period4	0		−0.261
H_period1	−0.575		−0.227
H_period2	−0.432		−0.449
H_period3	−0.073		−0.337
H_period4	−0.312		−0.227
ET_period1	0.431		0.344
ET_period2	0		0
ET_period3	0		0
ET_period4	0		0
RMSE	0.846		0.976
Adj. R^2	0.66		0.57
λ	0.0011		0.0054
BreadWheat_Landraces		Durum_Biparent	
Item	LASSO coefficients	Item	LASSO coefficients
Intercept	2.893	Intercept	2.027
D_period1[a]	−0.102		−0.207
D_period2	0		−0.241
D_period3	−0.064		0
D_period4	0		0
H_period1	−0.143		0
H_period2	−0.725		−0.361
H_period3	0		−0.215
H_period4	−0.247		−0.135
ET_period1	0.161		0.224
ET_period2	0		0
ET_period3	0		0
ET_period4	0		0.278
RMSE	0.605		0.110
Adj. R^2	0.70		0.98
λ	0.0063		0.000

The explanatory variables consisted of water stress (*D*), heat stress (*H*; number of days with temperatures >30°C), ET (Priestley–Taylor method) and the response variable was wheat grain yield (t/ha) for each of the four growing periods (periods 1–4, see text for details).
[a]Explanatory variables (*D*, *H*, or ET) plus period number.

phenotypic and physiological changes (Abhinandan *et al*., 2018). Not only the degree of these environmental factors but also the timing of occurrence during the crop growing season is important. These environmental stressors might occur at different times or simultaneously.

This study investigated the combined effects of abiotic stresses: heat, water deficit (SDR) and ET, on six wheat genotype categories in Australia. The findings highlight the intricate and multifaceted nature of understanding how multiple stressors can impact crop performance.

https://doi.org/10.1017/S0021859624000479 Published online by Cambridge University Press

Delayed sowing resulted in longer crop emergence time, slower growth, less ground cover, lower biomass and higher non-productive (evaporation) component of water balance. Late sown crops were exposed to higher temperature and ET during the critical crop development stage. Previous studies have shown that an early sown crop has a deeper rooting system to access subsoil water during the reproductive growth stage (Zeleke and Nendel, 2019). For all the growth periods considered in this study (periods 1–4), the correlation between explanatory variables (TGW [1000-grain weight], ET, T [transpiration], SDR, $H > 30$ [number of days with temperatures >30°C]) and dependent variable (grain yield) is different for different genotype groups (results shown for only period 3). This can be due to the inherent difference of the genotypes or due to pooled data from two sites and two sowing times. Heat, ET, and transpiration are negatively correlated with yield. One would expect that the more a crop transpires, the higher the yield will be. However, in our data higher rainfall (or higher ET or T) years were affected by lodging, resulting in lower yield.

Our research confirms that climate change presents significant challenges for wheat production. The different growing periods exhibited variations in water stress, ET, and heat stress ($H > 30$ days), demonstrating the potential for diverse climatic pressures throughout the growing season (Nuttall *et al.*, 2018). These stresses were found to significantly impact grain yield and plant characteristics like 1000-grain weight.

Interestingly, this study emphasizes that the combined effect of these stressors is not simply additive. Interactions between factors like heat and water deficit can be complex and vary depending on the specific genotype category and growing period. For example, while heat stress generally reduced yield in most wheat genotype categories tested here, its impact was less pronounced in the 'BreadWheat_Landraces' group in period 3, while water stress in period 2 had the largest detrimental effect for this group. Conversely, the 'Durum_Biparent' group seemed less sensitive to stress overall, even showing a positive response to increased ET in some periods. Similar results were found by other studies including Sinha *et al.* (2021) and Ru *et al.* (2023).

These findings underline the need for nuanced approaches to managing wheat crops under increasing climate variability (FAO, 2016). Selecting stress-tolerant varieties and implementing targeted strategies based on specific environmental conditions and genotype characteristics will be crucial for ensuring food security in a changing climate. This study highlights the need for genotype specific management strategies to minimize the impact of these stresses. Further research exploring additional stress factors and their interactions will also be vital for optimizing wheat production and resilience. One such additional stress factor is frost, especially when it occurs during the flowering period.

While the LASSO model effectively captured the main stress effects (Shafiee *et al.*, 2021), it is important to acknowledge the limitations. The observed stress–yield relationships likely involve intricate interactions that the model might not fully capture. For instance, the contrasting response of 'BreadWheat_Landraces' to heat stress across different periods suggests potential moderating factors or complex physiological mechanisms may have been at play. Further research delving deeper into these interactions and incorporating additional stress factors like salinity or nutrient deficiency could provide a more comprehensive understanding of how multiple stresses collectively can impact wheat performance (Teixeira *et al.*, 2013; Ru *et al.*, 2023).

Despite these known limitations, the LASSO model demonstrated promising results in predicting yield based on weather and soil-based indices, particularly for the 'Durum_Biparent' group. This highlights its potential as a tool for:

(1) *Identifying stress-tolerant genotypes*: by analysing the LASSO coefficients and stress responses across diverse genotype, researchers can prioritize genotypes with inherent resistance or resilience to specific stress combinations. This can be achieved by identifying genotype categories that exhibit consistently lower yield reductions under various stress combinations.
(2) *Targeted stress mitigation strategies*: understanding which stress factors are most critical for specific genotypes and growth periods allows for tailored interventions. The relative importance of these stress factors at different periods around the flowering time helps to implement appropriate stress management strategies. For example, if water stress is the primary limiting factor for a particular genotype category during a specific growth period, implementing irrigation scheduling strategies can be crucial. Conversely, for genotype categories sensitive to heat stress, exploring heat stress management techniques such as by earlier sowing or by growing earlier maturing varieties or breeding for heat tolerance can be prioritized.

To summarize, the LASSO analysis provided valuable insights into the diverse and complex ways that abiotic stresses impact wheat yield across different genotype categories. While further research is needed to fully understand the intricate interactions between stresses, this study demonstrates the potential of LASSO as a tool for predicting and managing stress impacts, ultimately contributing to improved wheat production and food security in a changing climate. This study highlights the significant impact of heat stress and drought as potential causes of yield losses. Compound stressors (heat and drought) can have more severe impacts on crops than individual impacts. The findings are significant for breeders, farmers and policymakers. The genotypes screened using this technique can be used in breeding for yield stability under dry, normal and wet seasons and different heat stress scenarios. This can help in screening, genetic development and improvements in phenotyping. Depending on the suitability of soil moisture, seed availability and farming operations, farmers can decide which genotype to sow in the sowing window. This information is useful for farming system planners and policymakers in making resource allocation decisions and in the delivery of incentives to mitigate the impacts of climate change.

Conclusion

In this study, we demonstrated how LASSO can be used to identify bread wheat and durum wheat genotypes with stress-tolerance ability within genotype groupings using data from multi-site and multi-year field experiments grown in NSW, Australia. Grain yield, soil characteristics and daily weather data were recorded to predict grain yield using stress indices. LASSO predicted grain yield well but adding other variables like lodging score, disease incidence, weed incidence and insect damage could improve prediction accuracy. Not all growing periods were predicted well. We found that the growth period 30 days pre-flowering up to flowering was sensitive for yield loss from heat and water stress as compared to other three periods of similar duration. The study confirms the usefulness of statistical modelling in identifying genotypes worthy of investigation by breeders.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0021859624000479.

Acknowledgements. We acknowledge the contributions of Dr Nicholas Collins (School of Agriculture Food and Wine, The University of Adelaide, Adelaide, SA, Australia) who supervised these projects and thank all the technical staff involved in the field experiments. We are grateful for the insights offered in the comments from anonymous reviewers and editors.

Author contributions. M. R. A., L. E., R. H. L. I., D. J. L., Y. S. C. and K. T. Z.: investigation; methodology; data curation; formal analysis; writing – review & editing. M. R. A., D. J. L. and R. H. L. I.: conceptualized the model analysis and wrote the first draft of the manuscript, which all authors further revised. L. E. supervised the field experimentations and data collection.

Funding statement. The data used for this paper were derived from research projects funded by the Grains Research and Development Corporation (GRDC) under project UA00123 and UA00147.

Competing interests. None.

References

Abbate PE, Andrade FH and Culot JP (1995) The effects of radiation and nitrogen on number of grains in wheat. *The Journal of Agricultural Science* **124**, 351–360.

Abhinandan K, Skori L, Stanic M, Hickerson NMN, Jamshed M and Samuel MA (2018) Abiotic stress signaling in wheat – an inclusive overview of hormonal interactions during abiotic stress responses in wheat. *Frontiers in Plant Science* **9**, 734.

ABS (2024) *Australian Bureau of Statistics 2021–22, Agricultural Commodities*. Canberra: ABS. Available at https://www.abs.gov.au/ (accessed 11 January 2024).

Akumaga U and Alderman PD (2019) Comparison of Penman–Monteith and Priestley–Taylor evapotranspiration methods for crop modelling in Oklahoma. *Agronomy Journal* **111**, 1171–1180.

Anwar MR, Liu DL, Farquharson R, Macadam I, Abadi A, Finlayson J, Wang B and Ramilan T (2015) Climate change impacts on phenology and yields of five broadacre crops at four climatologically distinct locations in Australia. *Agricultural Systems* **132**, 133–144.

Anwar MR, Luckett DJ, Chauhan YS, Ip RHL, Maphosa L, Simpson M, Warren A, Raman R, Richards MF, Pengilley G, Hobson K and Graham N (2022) Modelling the effects of cold temperature during the reproductive stage on the yield of chickpea (*Cicer arietinum* L). *International Journal of Biometeorology* **66**, 111–125.

APSIM (2023) SoilWat. Available at https://www.apsim.info/documentation/model-documentation/soil-modules-documentation/soilwat/ (accessed 13 October 2023).

Asseng S, Ewert F, Martre P, Rötter RP, Lobell DB, Cammarano D, Kimball BA, Ottman MJ, Wall GW, White JW, Reynolds MP, Alderman PD, Prasad PVV, Aggarwal PK, Anothai J, Basso B, Biernath C, Challinor AJ, De Sanctis G, Doltra J, Fereres E, Garcia-Vila M, Gayler S, Hoogenboom G, Hunt LA, Izaurralde RC, Jabloun M, Jones CD, Kersebaum KC, Koehler A-K, Müller C, Kumar SN, Nendel C, O'Leary G, Olesen JE, Palosuo T, Priesack E, Rezaei EE, Ruane AC, Semenov MA, Shcherbak I, Stöckle C, Stratonovitch P, Streck T, Supit I, Tao F, Thorburn PJ, Waha K, Wang E, Wallach D, Wolf J, Zhao Z and Zhu Y (2015) Rising temperatures reduce global wheat production. *Nature Climate Change* **5**, 143–147.

Chapman SC, Hammer GL and Meinke H (1993) A sunflower simulation model. I. Model development. *Agronomy Journal* **85**, 725–735.

Chenu K, Cooper M, Hammer GL, Mathews KL, Dreccer MF and Chapman SC (2011) Environment characterization as an aid to wheat improvement: interpreting genotype–environment interactions by modelling water-deficit patterns in North-Eastern Australia. *Journal of Experimental Botany* **62**, 1743–1755.

Collins B and Chenu K (2021) Improving productivity of Australian wheat by adapting sowing date and genotype phenology to future climate. *Climate Risk Management* **32**, 100300.

Collins N, Hildebrand S, Taylor K, Taylor H, Plemin D, Lohraseb I, Shirdelmoghanloo H, Erena M, Rahman M, Taylor J, Munoz-Santa S, Mather D, Heuer S, Sissons M and Emebiri L (2017) *Understanding Heat Impacts on Wheat to Breed Future Tolerance*. Canberra: GRDC Update. Available at https://grdc.com.au/resources-and-publications/grdc-update-papers/tab-content/grdc-update-papers/2017/02/understanding-heat-impacts-on-wheat-to-breed-future-tolerance (accessed 22 October 2023).

Didari S, Talebnejad R, Bahrami M and Mahmoudi MR (2023) Dryland farming wheat yield prediction using the LASSO regression model and meteorological variables in dry and semi-dry region. *Stochastic Environmental Research and Risk Assessment* **37**, 3967–3985.

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Marquéz JRG, Gruber B, Lafourcade B, Leitão PJ, Münkemüller T, McClean C, Osborne PE, Reineking B, Schröder B, Skidmore AK, Zurell D and Lautenbach S (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**, 27–46.

Emebiri L, Erena MF, Taylor K, Hildebrand S, Maccaferri M and Collins NC (2024) Field screening for heat-stress tolerance of floret fertility in wheat (*Triticum aestivum* and *T. durum*). *Crop & Pasture Science* **75**, CP23214.

Erena MF, Lohraseb I, Munoz-Santa I, Taylor JD, Emebiri LC and Collins NC (2021) The WtmsDW locus on wheat chromosome 2B controls major natural variation for floret sterility responses to heat stress at booting stage. *Frontiers in Plant Science* **12**, 635397.

FAO (2016) Food and Agriculture Organization of the United Nations. Climate change and food security: Risks and responses. Available at https://www.fao.org/3/i5188e/I5188E.pdf (accessed 5 January 2024).

Fischer RA (1985) Number of kernels in wheat crops and the influence of solar radiation and temperature. *The Journal of Agricultural Science* **105**, 447–461.

Friedman JH, Hastie T and Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* **33**, 1–22.

GRDC GrowNote (2017) Grains Research and Development Corporation (GRDC), Durum Southern Region. Available at https://grdc.com.au/resources-and-publications/grownotes/crop-agronomy/durum-southern-region-grownotes and https://grdc.com.au/__data/assets/pdf_file/0016/301651/GRDC-GrowNotes-Durum-SOUTHERN.pdf (accessed 11 January 2024).

He D and Wang E (2019) On the relation between soil water holding capacity and dryland crop productivity. *Geoderma* **353**, 11–24.

Hochman Z, Gobbett DL and Horan H (2017) Climate trends account for stalled wheat yields in Australia since 1990. *Global Change Biology* **23**, 2071–2081.

Holzworth DP, Huth NI, deVoil PG, Zurcher EJ, Herrmann NI, McLean G, Chenu K, van Oosterom EJ, Snow V, Murphy C, Moore AD, Brown H, Whish JPM, Verrall S, Fainges J, Bell LW, Peake AS, Poulton PL, Hochman Z, Thorburn PJ, Gaydon DS, Dalgliesh NP, Rodriguez D, Cox H, Chapman S, Doherty A, Teixeira E, Sharp J, Cichota R, Vogeler I, Li FY, Wang E, Hammer GL, Robertson MJ, Dimes JP, Whitbread AM, Hunt J, van Rees H, McClelland T, Carberry PS, Hargreaves JNG, MacLeod N, McDonald C, Harsdorf J, Wedgwood S and Keating BA (2014) APSIM – evolution towards a new generation of agricultural systems simulation. *Environmental Modelling & Software* **62**, 327–350.

Isbell RF and National Committee on Soil and Terrain (2021) *The Australian Soil Classification*, 3rd Edn. Melbourne: CSIRO Publishing.

Kerr RB, Hasegawa T, Lasco R, Bhatt I, Deryng D, Farrell A, Gurney-Smith H, Ju H, Lluch-Cota S, Meza F, Nelson G, Neufeldt H and Thornton P (2022) Food, fibre, and other ecosystem products. In Pörtner HO, Roberts DC, Tignor M, Poloczanska ES, Mintenbeck K, Alegría A, Craig M, Langsdorf S, Löschke S, Möller V, Okem A and Rama B (eds), *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK and New York, NY, USA: Cambridge University Press, pp. 713–906. https://doi.org/10.1017/9781009325844.007.

Kutner MH, Nachtsheim CJ, Neter J and Li W (2005) *Applied Linear Statistical Models*, Vol. **5**. Boston: McGraw-Hill Irwin.

Langridge P and Reynolds M (2021) Breeding for drought and heat tolerance in wheat. *Theoretical and Applied Genetics* **134**, 1753–1769.

Li S, Wang B, Feng P, Liu DL, Li C, Shi L and Yu Q (2022) Assessing climate vulnerability of historical wheat yield in south-eastern Australia's wheat belt. *Agricultural Systems* **196**, 103340.

Li Q, Ye A, Wada Y, Zhang Y and Zhou J (2024) Climate change leads to an expansion of global drought-sensitive area. *Journal of Hydrology* **632**, 130874. https://doi.org/10.1016/j.jhydrol.2024.130874.

Liu DL, Anwar MR, O'Leary G and Conyers MK (2014) Managing wheat stubble as an effective approach to sequester soil carbon in a semi-arid environment: spatial modelling. *Geoderma* **214–215**, 50–61.

Lobell DB, Schlenker W and Costa-Roberts J (2011) Climate trends and global crop production since 1980. *Science* **333**, 616–620.

Lohithaswa HC, Shreekanth SM, Banakara SK, Sripathy KV and Mallikarjuna MG (2022) Genomic selection for enhanced stress tolerance in maize. In Mallikarjuna MG, Nayaka SC and Kaul T (eds), *Next-Generation Plant Breeding Approaches for Stress Resilience in Cereal Crops*. Singapore: Springer, pp. 121–160. https://doi.org/10.1007/978-981-19-1445-4_4.

Mittler R (2006) Abiotic stress, the field environment and stress combination. *Trends in Plant Science* **11**, 15–19.

Monahan JF (2011) *A Primer on Linear Models*. New York: Chapman and Hall/CRC. https://doi.org/10.1201/b11551.

NSW Government Water (2020) The 2017–20 drought. Drought stages and measures implemented. Available at https://water.dpie.nsw.gov.au/our-work/allocations-availability/drought-and-floods/drought-recovery/2017-20-drought#:~:text=From%20January%202017%20to%20December,Millennium%20drought%20in%20many%20areas (accessed 4 June 2024).

Nuttall J, Brady S, Brand J, O'Leary G and Fitzgerald GJ (2012) Heat waves and wheat growth under a future climate. In Yunusa I (ed.), Capturing Opportunities and Overcoming Obstacles in Australian Agronomy. *Proceedings of 16th Australian Agronomy Conference 2012, 14–18 October 2012, Armidale, NSW*. Available at http://www.regional.org.au/au/asa/2012/climate-change/8085_nuttalljg.htm.

Nuttall JG, Barlow KM, Delahunty AJ, Christy BP and O'Leary GJ (2018) Acute high temperature response in wheat. *Agronomy Journal* **110**, 1296–1308.

OEC (2023) The Observatory of Economic Complexity. Australia (AUS) Exports, Imports, and Trade Partners. Available at https://oec.world/en/profile/country/aus?yearlyTradeFlowSelector=flow1 (accessed 11 January 2024).

OECD/FAO (2023) *OECD-FAO Agricultural Outlook*. Paris: OECD Agriculture Statistics. Available at https://www.oecd.org/publications/oecd-fao-agricultural-outlook-19991142.htm (accessed 12 October 2023).

Ortiz-Monasterio JI, Dhillon SS and Fischer RA (1994) Date of sowing effects on grain-yield and yield components of irrigated spring wheat cultivars and relationships with radiation and temperature in Ludhiana, India. *Field Crops Research* **37**, 169–184.

Pandey P, Irulappan V, Bagavathiannan MV and Senthil-Kumar M (2017) Impact of combined abiotic and biotic stresses on plant growth and avenues for crop improvement by exploiting physio-morphological traits. *Frontiers in Plant Science* **8**, 537.

Paredes P, Rodrigues GC, Alves I and Pereira LS (2014) Partitioning evapotranspiration, yield prediction, and economic returns of maize under various irrigation management strategies. *Agricultural Water Management* **135**, 27–39.

PDI (2022) NSW Government Department of Primary Industries, Cropping Wheat. Available at https://www.dpi.nsw.gov.au/about-us/publications/pdi/2022/wheat#:~:text=As%20a%20result%2C%20North%20American,drought%20in%202017%20to%202019 (accessed 4 June 2024).

Prasad PVV, Pisipati SR, Momcilovic I and Ristic Z (2011) Independent and combined effects of high temperature and drought stress during grain filling on plant yield and chloroplast EF-Tu expression in spring wheat. *Journal of Agronomy and Crop Science* **197**, 430–441.

Priestly CHB and Taylor RJ (1972) On the assessment of surface heat and evaporation using large-scale parameters. *Monthly Weather Review* **100**, 81.

R Core Team (2023) R: A Language and Environment for Statistical Computing, Vienna, Austria. Available at http://www.R-project.org.

RStudio Team (2023) RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. Available at https://posit.co/downloads/.

Ru C, Hu X, Chen D, Wang W, Zhen J and Song T (2023) Individual and combined effects of heat and drought and subsequent recovery on winter wheat (*Triticum aestivum* L.) photosynthesis, nitrogen metabolism, cell osmoregulation, and yield formation. *Plant Physiology and Biochemistry* **196**, 222–235.

Sadras VO and McDonald G (2012) *Water Use Efficiency of Grain Crops in Australia: Principles, Benchmarks and Management*. Canberra: GRDC Project Code: DAS00089. Available at https://grdc.com.au/__data/assets/pdf_file/0030/159186/grdcpublicationwateruseefficiencyofgraincropsinaustraliapdf.pdf.pdf.

Shafiee S, Lied LM, Burud I, Dieseth JA, Alsheikh M and Lillemo M (2021) Sequential forward selection and support vector regression in comparison to LASSO regression for spring wheat yield prediction based on UAV imagery. *Computers and Electronics in Agriculture* **183**, 106036.

Shirdelmoghanloo H, Taylor JD, Lohraseb I, Rabie H, Brien C, Timmins A, Martin P, Mather DE, Emebiri L and Collins NC (2016) A QTL on the short arm of wheat (*Triticum aestivum* L.) chromosome 3B affects the stability of grain weight in plants exposed to a brief heat shock early in grain filling. *BMC Plant Biology* **16**, 100.

Sinha R, Fritschi FB, Zandalinas SI and Mittler R (2021) The impact of stress combination on reproductive processes in crops. *Plant Science* **311**, 111007.

Sissons M, Pleming D, Taylor JD, Emebiri L and Collins NC (2018) Effects of heat exposure from late sowing on the agronomic and technological quality of tetraploid wheat. *Cereal Chemistry* **95**, 274–287.

Sissons M, Pleming D, Taylor JD, Emebiri L, Eckermann P and Collins NC (2024) Effects of heat exposure from late sowing on agronomic traits and the technological quality of hexaploid wheat. *Journal of Cereal Science* **118**, 103950.

Slafer GA, Savin R and Sadras VO (2023) Wheat yield is not causally related to the duration of the growing season. *European Journal of Agronomy* **148**, 126885.

Stone PJ and Nicolas ME (1994) Wheat cultivars vary widely in their responses of grain yield and quality to short periods of post-anthesis heat stress. *Australian Journal of Plant Physiology* **21**, 887–900.

Street K, Bari A, Mackay M and Amri A (2016) How the focused identification of germplasm strategy (FIGS) is used to mine plant genetic resources collections for adaptive traits. In Maxted N, Dulloo ME and Ford-Lloyd BV (eds), *Enhancing Crop Genepool Use: Capturing Wild Relative and Landrace Diversity for Crop Improvement*. Wallingford, UK: CAB International, pp. 54–63.

Talukder ASMHM, McDonald GK and Gill GS (2013) Effect of short-term heat stress prior to flowering and at early grain set on the utilization of water-soluble carbohydrate by wheat genotypes. *Field Crops Research* **147**, 1–11.

Talukder SK, Babar MA, Vijayalakshmi K, Poland J, Prasad VV, Bowden R and Fritz A (2014) Mapping QTL for the traits associated with heat tolerance in wheat (*Triticum aestivum* L.). *BMC Genetics* **15**, 97.

Teixeira EI, Fischer G, van Velthuizen H, Walter C and Ewert F (2013) Global hot-spots of heat stress on agricultural crops due to climate change. *Agricultural and Forest Meteorology* **170**, 206–215.

Trethowan RM (2022) Abiotic stresses. In Reynolds MP and Braun HJ (eds), *Wheat Improvement*. Cham: Springer, pp. 159–175. https://doi.org/10.1007/978-3-030-90673-3_10.

Trout TJ and DeJonge KC (2017) Water productivity of maize in the US high plains. *Irrigation Science* **35**, 251–266.

Unkovich M, Baldock J and Farquharson R (2018) Field measurements of bare soil evaporation and crop transpiration, and transpiration efficiency, for rainfed grain crops in Australia – a review. *Agricultural Water Management* **205**, 72–80.

Unkovich M, McBeath T, Moodie M and Macdonald LM (2023) High soil strength and cereal crop responses to deeper tillage on sandy soils in a semi-arid environment. *Field Crops Research* **291**, 108792.

Wang B, Liu DL, Asseng S, Macadam I and Yu Q (2017) Modelling wheat yield change under CO_2 increase, heat and water stress in relation to plant available water capacity in eastern Australia. *European Journal of Agronomy* **90**, 152–161.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K and Yutani H (2019) Welcome to the tidyverse. *Journal of Open Source Software* **43**, 1686.

Xing H, Liu DL, Li G, Wang B, Anwar MR, Crean J, Lines-Kelly R and Yu G (2017) Incorporating grain legumes in cereal-based cropping systems to improve profitability in southern New South Wales, Australia. *Agricultural Systems* **154**, 112–123.

Yang Y, Liu DL, Anwar MR, O'Leary G, Macadam I and Yang Y (2016) Water use efficiency and crop water balance of rainfed wheat in a semi-arid environment: sensitivity of future changes to projected climate changes and soil type. *Theoretical and Applied Climatology* **123**, 565–579.

Zeleke K and Nendel C (2019) Growth and yield response of faba bean to soil moisture regimes and sowing dates: field experiment and modelling study. *Agricultural Water Management* **213**, 1063–1077.

Zeleke KT, Anwar MR, Emebiri L and Luckett D (2023) Weather indices during reproductive phase explain wheat yield variability. *The Journal of Agricultural Science* **161**, 617–632.

Zhao C, Liu B, Piao S, Wang X, Lobell DB, Huang Y, Huang M, Yao Y, Bassu S, Ciais P, Durand J-L, Elliott J, Ewert F, Janssens IA, Li T, Lin E, Liu Q, Martre P, Müller C, Peng S, Penuelas J, Ruane AC, Wallach D, Wang T, Wu D, Liu Z, Zhu Y, Zhu Z and Asseng S (2017) Temperature increase reduces global yields of major crops in four independent estimates. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 35. https://doi.org/10.1073/pnas.1701762114.