Historic wheat robustness

Historic Trends and Sources of Year-over-Year Stability in Montana Winter Wheat Yields

Jennifer Lachowiec*, James E. Berg, Meng Liang

J. Lachowiec, M. Liang, and J. E. Berg, Montana State University, 119 Plant Bioscience Building, Bozeman, MT 59717, USA. Corresponding author (jennifer.lachowiec@montana.edu).

ORCID information

Jennifer Lachowiec 0000-0003-2962-6448,

Meng Liang 0000-0002-1467-946X

ABSTRACT

Producers desire varieties that consistently perform with high yields and end-use qualities. Unlike easily recognized average yield improvements, yield stability over time is less examined, especially when considering the role of breeding relative to other factors like management and changing climatic conditions. Our study system was a 70-year historical dataset from which we estimated the year-over-year stability of *Triticum aestivum*, winter wheat varieties released by Montana's Agricultural Experimental Station. We examined yield stability within six locations representing diverse growing conditions across Montana and found evidence that breeding has improved stability at specific locations and not at others. Newer varieties showed improved year-over-year stability at locations that tended to have the lowest yields and more extreme weather conditions, reflecting that year-over-year stability has a genotype-by-environment component.

We examined the role of climatic conditions, including temperature and rainfall to understand if reduced climatic variability was driving patterns of improved stability at these sites. However, the impact of breeding remained, or became evident when accounting for climatic variables. Together, these findings suggest that breeding's strong selective pressures improve second order traits.

INTRODUCTION

Volatile crop yields caused by environmental extremes threaten our global food supply (Lobell et al., 2008). Further, producers lose economic value when crops grow variably across fields, due to increased difficulties in management and harvest and meeting end-use quality requirements. To counteract these obstacles, multiple approaches are taken to optimize yield stability and quality, including breeding and agronomic practices (Reckling et al., 2021). Crops grown at a specific geographic location experience not only varied weather and disease pressure year-over-year, but also microenvironmental differences within a field and stochastic errors during development (Lachowiec et al., 2015). Genotype greatly influences how plants withstand collective perturbations (Hill & Mulder, 2010), herein referred to as robustness. Because genotypes associated with differing degrees of robustness have been identified in plants (Hall et al., 2007; Jimenez-Gomez et al., 2011; Sangster et al., 2008), including crops (Fisher & Zamir, 2021; Makumburage & Stapleton, 2011; Ordas et al., 2008; Tollenaar & Wu, 1999), we hypothesize that breeding has increased crop robustness to such perturbations.

Defining and measuring robustness is a complicated task. Not to be confused with the stability (or conversely, the plasticity) of a trait across geographic location, robustness is a way to describe the uniformity, repeatability, intra-genotypic variability (Bruijning et al., 2020), and predictability of a genotype within a location. In many studies examining agronomic traits, the coefficient of variation (CV) of a trait is calculated. CV traditionally is used to identify undesired high levels of variability to determine experimental validity (Bowman, 2001) but also potentially holds information about robustness (Fasoula & Fasoula, 2002). In agriculture, robustness also has been captured using many statistics like Ve (Schou et al., 2020) and heterogeneity of environmental variance (Hill & Mulder, 2010), among many others (Reckling et al., 2021), revealing how different genotypes can influence robustness. Using these and similar robustness statistics, the underlying genetic architecture and genetic loci controlling robustness have been identified (Fisher & Zamir, 2021; Hall et al., 2007; Sangster et al., 2008).

Like most quantitative traits, robustness is affected not only by genotype, but also environment and the interaction of genotype and environment (Falconer & Mackay, 1996). Most commonly, how the environment affects robustness is measured within a single variety to assess the impact of different agronomic practices. Varied planting density (Lu et al., 2020), application of fertilizers, and irrigation (Kristensen et al., 2008; Kukal & Irmak, 2018) are examples of agronomic practices examined for impacts on robustness. Long-term studies have produced mixed findings regarding how robustness has changed over time in crops. An economic study of global wheat and maize yields produced since the Green Revolution show an increase in robustness over time (Gollin, 2006). In contrast, an agronomic study of a single, highly controlled, long-term site found both barley and wheat yields decreased in robustness, dependent on fertility management (Macholdt et al., 2021). A breeding study considering the impact of a small number of specific maize hybrids released over five decades, found increased robustness is associated with greater yields, due to improved stress tolerance (Tollenaar & Wu, 1999).

Additional information is needed to understand how breeding, location, and management influence yield robustness.

To explore how breeding affects robustness, we examined historical data collected on winter wheat grown in Montana. In Montana, winter wheat is one of the main cereal crops with a planting area of over 627,000 hectares in 2020, accounting for nearly $400 M in production value (National Agricultural Statistics Service, 2021). Since 1949, Montana's winter wheat breeding program grew winter wheat systematically at multiple research centers across the state, collecting yield data. With these data, we examined yield year-to-year variability. We inspected how yield robustness varies with geographic location. Given the important role of weather variability from year-to-year at a specific location, we examined the impacts of temperature and precipitation on yields and its robustness. We propose the role of breeding in yield robustness based on examining the most grown Montana Agricultural Experiment Station varieties developed over the last 100 years.

MATERIALS AND METHODS

Data sets

Yield data

Winter wheat yields were obtained from the Montana Winter Wheat Breeding Program, reported as bushel per acre and converted to kg ha^{-1}. The 48 varieties included those that have been planted across the state (Table S1, National Agricultural Statistics Service 1958-2019). The data include the yields of released winter wheat varieties tested in Montana from 1949 to 2019 at six research centers with the following soil types: Northern Agricultural Research Center with silt loam (48° 30′, 109° 48′; Havre, MT), Northwestern Agricultural Research Center with silt loam (48° 10′, 114° 15′; Kalispell/Creston, MT), Central Agricultural Research Center clay loam (47°

03′, 109° 57′; Moccasin, MT), Southern Agricultural Research Center with silt loam (45° 55′, 108° 15′; Huntley, MT), Eastern Agricultural Research Center with clay loam (47° 40′, 104° 08′; Sidney, MT), and the Post Research Farm silt loam (45° 41′, 111° 00′; Bozeman, MT). Extreme winterkill resulting in zero yields were excluded from all analyses as outliers (Reckling et al., 2021). Only dryland results were included with no additional irrigation.

Weather data

Weather data was obtained using the web interface and search capabilities provided by the National Oceanic and Atmospheric Administration's National Centers for Environmental Information (https://www.ncdc.noaa.gov/cdo-web/search), accessed April 1, 2021. The datasets chosen were "Global Summary of the Month" and "Global Summary of the Year" ranging from January 1949 to December 2019. Weather stations were chosen for their proximity to the research centers and data availability. Data were obtained from Kalispell Glacier Airport (Station IDUSC00244558), Sidney (Station ID USC00247560), Huntley Experimental Station (Station ID USC00244345), and Moccasin Experiment Station (Station ID USC00245761). In Bozeman, two stations were needed to obtain desired weather data. From January 1949 to October 1966, data were obtained from Bozeman Montana State University (Station ID USC00241044), and from November 1966 to December 2019, data were obtained from Bozeman 6 W Experimental Farm, (Station ID USC00241047). In Havre, two stations were used to obtain data spanning the focus period: January 1949 to February 1961 Havre Weather Bureau (Station ID USW00024035) and March 1961 to December 2019 Havre Airport (Station ID USW00094). From each weather station, four metrics of weather were obtained: monthly average (TAVG), monthly minimum (TMIN), and monthly maximum (TMAX) air temperatures and monthly cumulative precipitation

(PRCP). The percentage data coverage across the weather metrics follows: Bozeman-98.5%, Havre-100%, Huntley-96.4%, Kalispell-99.4%, Moccasin-95.8%, and Sidney-97.8%.

Statistical analyses

All statistical analyses were performed using R version 4.1.0 (R Development Core Team, 2011).

Yield analyses

To examine the change in yields over time, we calculated the mean yield each year across varieties at the six research centers. We used a linear model to evaluate the rates of change when calculating the mean across locations per year and at each location separately.

To examine the change in yield robustness over time, we utilized multiple approaches. First, we visually examined the residuals of the linear model capturing the relationship between yield and year. Second, we calculated the coefficient of variation ($\mathrm{CV} = \frac{s}{\bar{Y}}$) of yield of a variety across the years it was grown (Ray et al., 2015). Because the number of years a variety was tested was positively correlated with the CV (Fig. S2), we employed two alternative methods to account for this relationship. First, we observed a threshold in the impact of number of years grown on the CV and used a quadratic plateau model to estimate the transition point, implemented with the nlstools package in R (Baty et al., 2015). The 95% confidence interval for the transition point was 6.57-12.14, with an estimate of 9.36. Therefore, we used a threshold of greater than 9 years grown to estimate CV values for varieties. In a second approach, data were sampled with replacement to eliminate effect of number of years sampled on the CV. Varieties were filtered to those that included measurements from at least 5 years. For each bootstrap sample, five years of yield data were selected for each variety. The linear relationship between the CV of the bootstrapped samples and release year was computed. This process was repeated 1000 times, and the distributions of b, R^2, and p-value for b for each bootstrapped sample was examined. Bootstrap estimates were calculated pooling across locations as well as for each individual location.

Weather analyses

To understand trends in weather metrics between 1949-2019, we examined annual mean values obtained for each location. The relationship between metrics and year were modeled with linear regression for each location. The change in each weather metric was calculated per decade using the slope estimate. Annual weather variability was examined by calculating a "moving CV" with a moving window of 30 years using the annual mean values for each weather metric.

Seasonality in the weather metrics were also examined at each location to understand if certain months were highly variability in temperatures relative to others. The monthly means were calculated between the years of 1949-2019. Weather variability over time was examined by using the statistic CV after transforming the data to only include positive values by adding 30 to all values. The CV was then calculated for the monthly weather metrics obtained for each month.

Combined weather and yield analyses

Because weather data were available each year at each location, we examined the relationship between the weather metrics and yields. Using principal components analysis, we examined how the four weather metrics, release year (capturing the impact of breeding), and year grown (capturing, at least, changes in agronomic practices) were acting in concert. We then calculated the residual CV (CV_{res}) (Kukal & Irmak, 2018; Schou et al., 2020) by fitting a linear model with yields as the response variable and using the mean annual temperature and cumulative precipitation, which contributed most strongly to PC1 and PC2, and their interaction as the explanatory variables. Using the residuals from this model after transformation to positive values by adding the minimum residual, the CV_{res} was calculated for each variety. Next, the relationship between CV_{res} and release year was examined using linear regression.

RESULTS

Montana wheat yields have increased with varied impact on yield robustness

To assess how winter wheat yields have changed historically, we focused on the most tested varieties in Montana's winter wheat breeding program since the 1920s. We examined forty-eight varieties ranked on acreage planted in Montana after their release (Table S1, Montana Agriculture Statistics Service 1958-2019), assessing the results of yield trials across six research centers representing different production environments in Montana. We first examined winter wheat's interannual improvement between 1949-2019 by calculating the mean yield each year. On average across the six locations, yields increased 43.55 kg ha^{-1} yr^{-1}based on fitting a linear model (Fig. 1a). Depending on location, yields showed over a five-fold average improvement over the last 70 years (Fig. 1b). Huntley and Kalispell had the largest year-over-years gains (Fig. 1c). These yield changes are not only dependent on breeding improvement but also on improvement in agronomic practices (Lanning et al., 2010) and changing climatic conditions and atmospheric CO_2 (McGrath & Lobell, 2013).

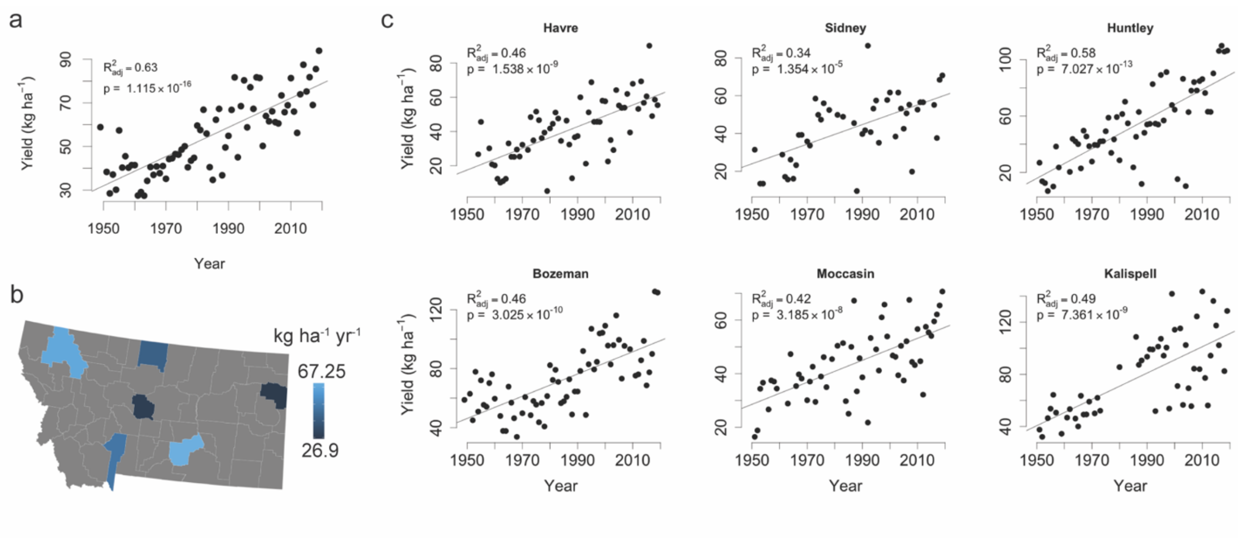

Fig. 1 Winter wheat yield increases across Montana from 1949-2019

a) The average yields from Montana's released varieties from breeding nurseries across six research centers is shown from 1949-2019. The linear regression fit is indicated with the solid black line, with adjusted R^2 and p-values for the slope shown. b) Yields from 1949-2019 is shown for each research station. The linear regression fit is indicated with the solid black line, with adjusted R^2 and p-values for the slope displayed. c) The linear regression coefficient for yield and year for each research center from panel b) is indicated with the color scale for the corresponding county.

In addition to observing increased yields over time, we hypothesized increased year-over-year yield stability. To distinguish stability year-over-year within a location from the more common definition of stability across geographic locations (Becker & Leon, 1988; Finlay & Wilkinson, 1963), we will use the term robustness to refer to year-over-year stability. In other words, within a single geographic location we can describe how robust yields are to yearly fluctuations in weather, plot location at a research center, biotic stresses, and more. We hypothesized that the residuals of the linear regression models estimating the change in yield over time at each location would decrease with time. However, the residuals showed little change over time, except at Kalispell, where they instead appeared to increase, contrary to our prediction (Fig. S1). These data suggest that the collective impacts of changing agronomic practices, climate, breeding, etc. are not resulting in greater yield robustness, with Kalispell decreasing in robustness, and other locations unchanging.

The association of variety release year and yield robustness depends on location

Varieties were pooled together in the previous analyses, potentially obscuring the impact of breeding on robustness. Given that wheat is primarily released as inbred varieties, the selection pressure to increase yields during breeding may also tend to select genotypes with more narrower trait distributions (Gavrilets & Hastings, 1994; Wagner et al., 1997)—or more predictable genotypes. We first tested whether breeding could be a component affecting yield robustness by comparing the release year of a variety to its robustness using the statistic CV. We calculated the yield robustness for each variety across years planted at a single location; however, we observed that robustness was sensitive to the number of years it was grown (Fig. S2).

To reduce or eliminate the impact of number of years grown on robustness, we used two approaches. First, based upon an estimated plateau in decreased robustness due to number of years grown (Fig. S2a), we set a conservative threshold of greater than 9 years planted for inclusion. This approach reduced the number of varieties included and eliminated those most recently publicly released from further analysis, because they have not existed long enough to be tested for 10 years. With these filtered data, we detected a very weak negative linear relationship between CV and release year among locations in this filtered dataset (Fig. 2a, Table 1).

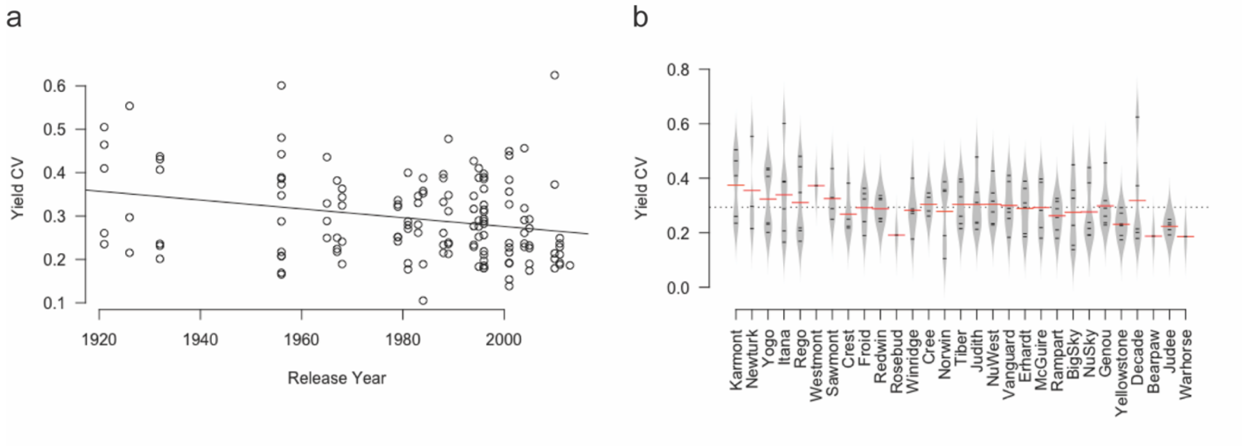

Fig. 2 Yield CV for varieties decreases with release year

a) The yield CV was calculated for each variety at each research station for those grown at least 10 years. The solid line indicates the fit of yield CV and release year of corresponding variety from linear regression. b) The yield CV for each variety is shown. Black lines indicating the yield CV from each research center for which the variety was grown at least 10 years. The red lines indicate the mean yield CV for each variety, and the dotted black line shows the grand mean yield CV. The varieties are arranged in order of release.

TABLE 1 Linear regression of variety yield CV on variety release year

Location	***n***	***b***	**SE_b**	**R^2**
All	138	-0.001071***	0.0003359	+0.0651
Havre	24	-0.002589***	0.0004373	+0.5969
Sidney	17	-0.003069***	0.0009891	+0.4878
Huntley	25	-0.001776***	0.0006205	+0.2306
Moccasin	26	-0.000475	0.0003297	+0.04118
Bozeman	27	-0.000340	0.0004239	-0.01395
Kalispell	19	+0.003141*	0.001137	+0.2693

** Significant at the .01 probability level, *** Significant at the .001 probability level.

Given the yield of a variety is sensitive to geographic location in which it was grown, we hypothesized that robustness of a variety also would depend on location. We examined the location-specific yield CV values for varieties grown over nine years and observed that the CV was not consistent across locations and could vary widely (Fig. 2b). Only two varieties, Yellowstone and Judee, showed CV values across all locations that were less than the overall mean. The impact of location on the relationship between release year and robustness suggests multiple potential explanations, including that the environmental conditions experienced at each locations drove the specific yield CV patterns observed over time. Indeed, when examining robustness *within* specific locations, we found different tendencies due to location between the relationship of release year and yield CV. Havre and Sidney showed the strongest relationship between release year and robustness (Fig. 3a, Table 1). Kalispell exhibited a pattern opposite to the other locations with robustness decreasing in more recently released varieties (Fig. 3a).

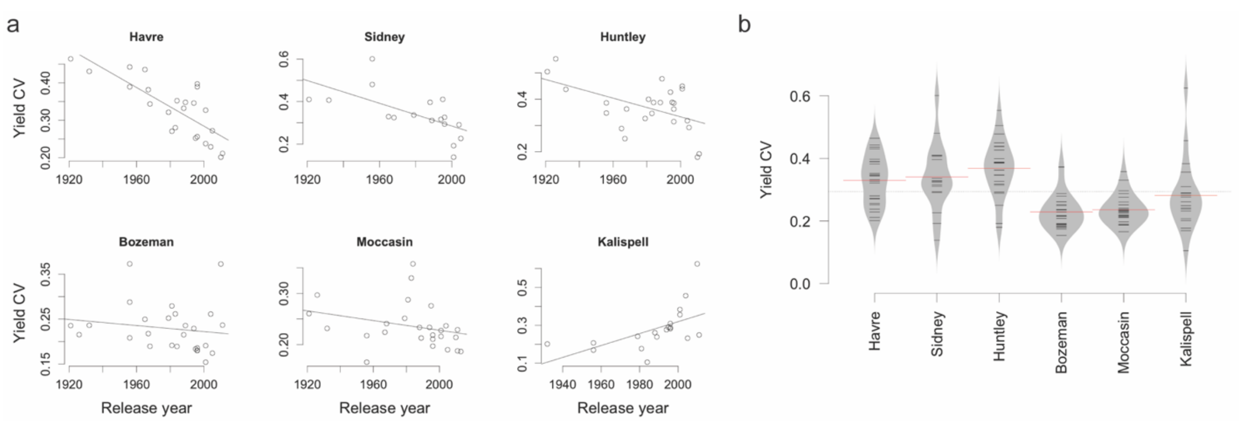

Fig. 3 The relationship between yield CV and release year of a variety depends on location

a) The Yield CV was calculated for each variety at each research station for those grown at least 10 years. The solid line indicates the fit of yield CV and release year of corresponding variety from linear regression. b) The yield CV is shown for each variety grown at least ten years at each research station (black lines). The mean yield CV for each research station is indicated with the red lines, and the grand mean is indicated with the dotted black line.

The threshold-based approach used above to determine the minimum number of years from which to calculate robustness may be biased. It is possible that lines that were planted for many years may simply be more robust due to breeder selection. Therefore, we also performed a second, similar analysis using bootstrapping to understand the relationship between robustness and release year. In this approach, we calculated robustness using five randomly chosen yield values for each released line measured at least five years. Then we modelled the relationship between release year and robustness. We performed 1000 bootstraps and examined the summary statistics' distributions for evidence of a relationship between release year and robustness. Using the data across locations, we observed a negative relationship between CV and release year (mean of 1000 bootstraps: b = -0.0125, P = 0.0197, R^2 = 0.179, Fig. S3a-c), a very similar outcome as the threshold approach. Next, we performed this same bootstrap procedure for each location (Fig. S3d-i). Similar to the previous approach using a 10-year filter, we again found support for a negative relationship between release year and robustness in Havre and Huntley. At Kalispell and Sidney, the indication of relationships between robustness and release year was not as evident; the data do not support a relationship.

To begin to understand why locations showed differences in the relationship between robustness and release year, we first examined the range of CV values. We noted that the CV varied with geographic location, with higher CV values at Huntley, Sidney, and Havre (Fig. 3b). We suspect that to observe a relationship between CV and release year, locations must have environments that have weather and field perturbations to be overcome, and the absence of this variability would prohibit observing a pattern. Thus, this analysis suggests that breeding has contributed to improved robustness in a manner that is dependent on location or can only be observed in certain locations or environmental conditions.

We also found evidence that biotic perturbations influenced levels of robustness in a location-specific manner. At Kalispell, Decade exhibited the lowest robustness in yield of the entire dataset (yield CV > 0.6, Fig. 3b). This is due to very large range of yields produced over 12 years of testing, ranging from 605.25 kg ha^{-1} to 9280.5 kg ha^{-1}. The low yields for Decade were location specific and not observed at other sites. Examining the breeder's notes, we found that second lowest value (1076 kg ha^{-1}) measured was associated with stripe rust at Kalispell. The lowest value (605.25 kg ha^{-1}) was the following year. In addition to the abiotic, agronomic practices, and breeding impacts considered, biotic stresses also can greatly impact the robustness measured. In this case, however, we observe that the robustness is associated with genotype, and conclude that Decade represents a genotype particularly susceptible relative to others tested at the same location and year (Kertho, 2014; Riveland et al., 2011) .

Assessment of weather variability across locations

We hypothesized that the differences in yield robustness between locations in part reflects the weather conditions experienced by a variety the years it was grown. Therefore, we first assessed mean weather patterns at the six locations at multiple scales, focusing on temperature and precipitation, both of which are critical input to crop yield (Kukal & Irmak, 2018). Between 1949-2019, the six locations varied in average annual temperature and precipitation, with Huntley having the highest average annual temperatures and Havre having the lowest precipitation (Fig. S4a-d). At the same time, all six locations increased in maximum annual temperatures between 1949-2019 (Table 2). Sidney had the largest average annual temperature change between 1949 and 2019, increasing 0.41°C per decade. Sidney, Bozeman, and Moccasin also had increasing minimum average annual temperatures during this time. Precipitation trends also varied per location. Bozeman and Moccasin, which tended to have higher precipitation among these locations, were unchanged or tending towards decreasing during this time period, but Sidney had an annual average precipitation increasing 1.33 cm per decade between 1949-2019. For a better resolved perspective of seasonal temperatures at each location, we calculated the monthly means from 1949-2019 (Fig. S4e-g). Havre and Sidney experienced the most extreme temperatures, not only facing high mean, minimum and maximum temperatures in the summer months, but also the lowest temperatures in the winter months relative to the other four locations. Monthly total precipitation means were also calculated for each month from 1949-2019 (Fig. S4g). Kalispell has higher rainfall, especially during the winter and early spring, relative to the other locations. Total monthly precipitation varied much more from month to month and across locations relative to temperature (Fig. S4d).

TABLE 2 Linear regression coefficients from regression of weather metrics on year from 1949-2019.

	$b \times 10^{a}$			
Location	**Temperature annual mean**	**Temperature annual maximum**	**Temperature annual minimum**	**Precipitation annual cumulative**
Havre	0.09212	0.17410*	0.01086	0.00453
Sidney	0.41412***	0.39452***	0.43258***	1.33099*
Huntley	0.15281*	0.19018*	0.11719	0.80292
Bozeman	0.16543**	0.19713**	0.13363**	-0.15007
Moccasin	0.22628**	0.19209*	0.24095***	-0.03251
Kalispell	0.11965**	0.16146**	0.07672	0.19277

* Significant at the .05 probability level, ** Significant at the .01 probability level, *** Significant at the .001 probability level. [a]To indicate trends per decade, yearly coefficients are shown multiplied by 10.

Using these monthly and annual weather metrics, we examined weather variability over time in the dataset. We calculated monthly CV statistics for the mean, minimum, and maximum temperatures, and the precipitation at each location from 1949-2019 (Fig. 4a-d). For all three summaries of temperature, we observed high CV in the fall and winter months, with a peak in January. Across locations, Havre and Sidney ranked as having the highest CV values during these winter months. In contrast, during the spring and summer months of winter wheat growth, the temperature CV were indistinguishable among locations, indicating little year-over-year variation in monthly temperatures. These patterns suggested that if year-over-year variability is impacting yield robustness, the source may be variation in the fall and winter months. Because winter wheat is planted in the fall and vernalizes over winter, temperature throughout winter also has direct impacts on the plants, including winter kill. Year-over-year precipitation showed a different pattern from temperature, with more CV values ranging from 0.40 – 1.15, again with Sidney and Havre exhibiting a tendency for more variability.

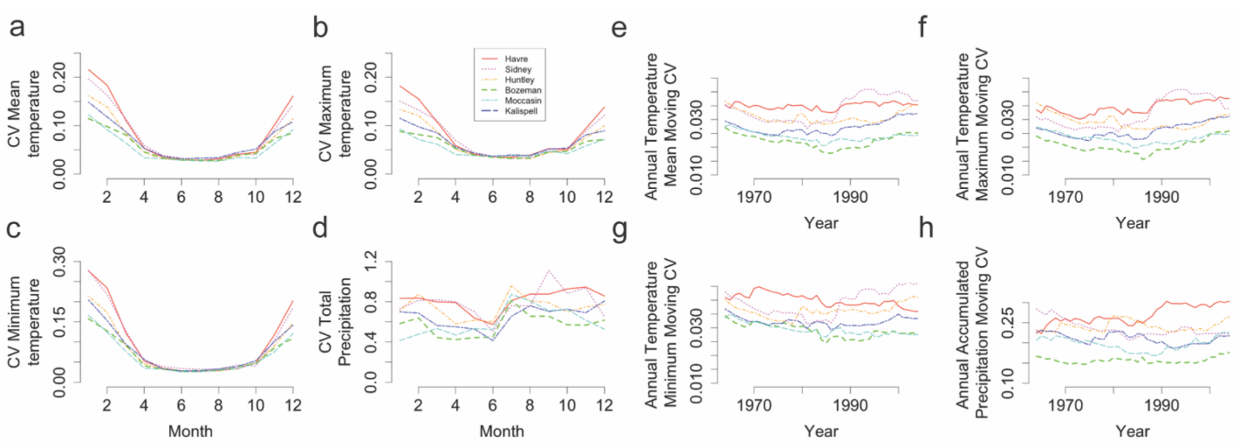

Fig. 4 Climate metrics show location-specific patterns in variability

a-d) The CV of monthly values from 1949-2019 for a) mean, b) maximum, c) minimum temperatures, and d) total precipitation are shown for each research center. e-h) CV values calculated from a 30-year moving window are displayed for each location for the average e) mean, f) maximum, g) minimum temperatures, and h) total precipitation.

Overall Montana's weather variability is projected to increase based on climate models (Whitlock et al., 2017). We used two approaches to examine historical trends in weather variability at the six locations examined. First, we calculated a "moving CV", based upon the annual temperature and precipitation values, using window sizes of 30 years, a standard in calculating climate normal (*U.S. Climate Normals*, 2021). Contrary to our expectation that weather variability is generally increasing, we observed that historic weather variability depended on location (Fig. 4e-h). Fitting a linear model to the moving CV values, we found that Bozeman weather in fact was converging for minimum temperature variability but showed no change for other weather metrics (Table 3). At Kalispell, all measures of temperature were converging, with precipitation variability remaining steady, consistent with previous patterns observed in Montana (Zhang et al., 2021). At Havre, Sidney, Huntley and Moccasin, the different weather metrics showed mixed patterns of variability (Table 3). Sidney, for example was diverging in all temperature metrics but converging for precipitation. Moreover, an abrupt upward variability shift in all three temperature metrics occurred in Sidney in approximately 1986 (Fig. 4e-g). We confirmed this shift was not due to changes in equipment at the Sidney weather station or missing weather data (only found in 2010s), but we have not identified sources of this shift. Out of the 24 weather-location combinations, eight showed convergence in weather variability, eight with divergence, and eight with no change.

TABLE 3 Linear regression coefficients from 30-year moving window CV of multiple weather metrics on year from 1954-2004.

	b			
Location	**Temperature annual mean CV**	**Temperature annual maximum CV**	**Temperature annual minimum CV**	**Precipitation annual cumulative CV**
Havre	3.625×10^{-5}	1.947×10^{-4}**	-2.460×10^{-4}**	1.6190×10^{-3}**
Sidney	3.411×10^{-4}**	4.423×10^{-4}**	2.817×10^{-4}**	-1.2177×10^{-3}**
Huntley	-2.430×10^{-5}	-1.332×10^{-4}**	1.390×10^{-4}**	1.729×10^{-4}
Bozeman	-4.244×10^{-5}	7.501×10^{-5}	-2.238×10^{-4}**	1.507×10^{-4}
Moccasin	1.766×10^{-4}**	2.102×10^{-4}**	-5.304×10^{-5}	-5.766×10^{-4}**
Kalispell	-1.069×10^{-4}**	-7.245×10^{-5}*	-2.398×10^{-4}**	-1.416×10^{-4}

* Significant at the .05 probability level, ** Significant at the .01 probability level.

Accounting for weather demonstrates remaining geographic influences on robustness

Weather, breeding, and agronomic practice are expected to impact yield CV at all locations but are not independent of one another. For example, while year grown reflects improvement in agronomic practices, it also captures increasing temperatures and is strongly correlated to release year. Therefore, we first examined the relationships among release year (to reflect breeding), year grown (to reflect agronomic practice at least), and the four weather metrics using principal component analysis. The first dimension described 47.8% of the variation (Fig. 5a), primarily capturing the three temperature metrics, along with a substantial contribution of year and release year. The second dimension was driven by year and release year, representing an addition 27.4% of the variation, which showed a close visual association with yield increase, as expected (Fig. 5b). The third dimension describing 17.4% of the variation primarily represented precipitation differences (Fig. 5a). Because of the inter-relatedness of weather, agronomic practices, and breeding due to their correlations with time, we isolated their potential impacts on yields. A partial correlation analysis showed that after accounting for the correlations between release year, weather metrics, and year, a remaining correlation of $r = 0.0583$ was remaining between yield and release year of varieties examined (Table 4) though the strongest correlation with yield was with precipitation ($r = 0.29$).

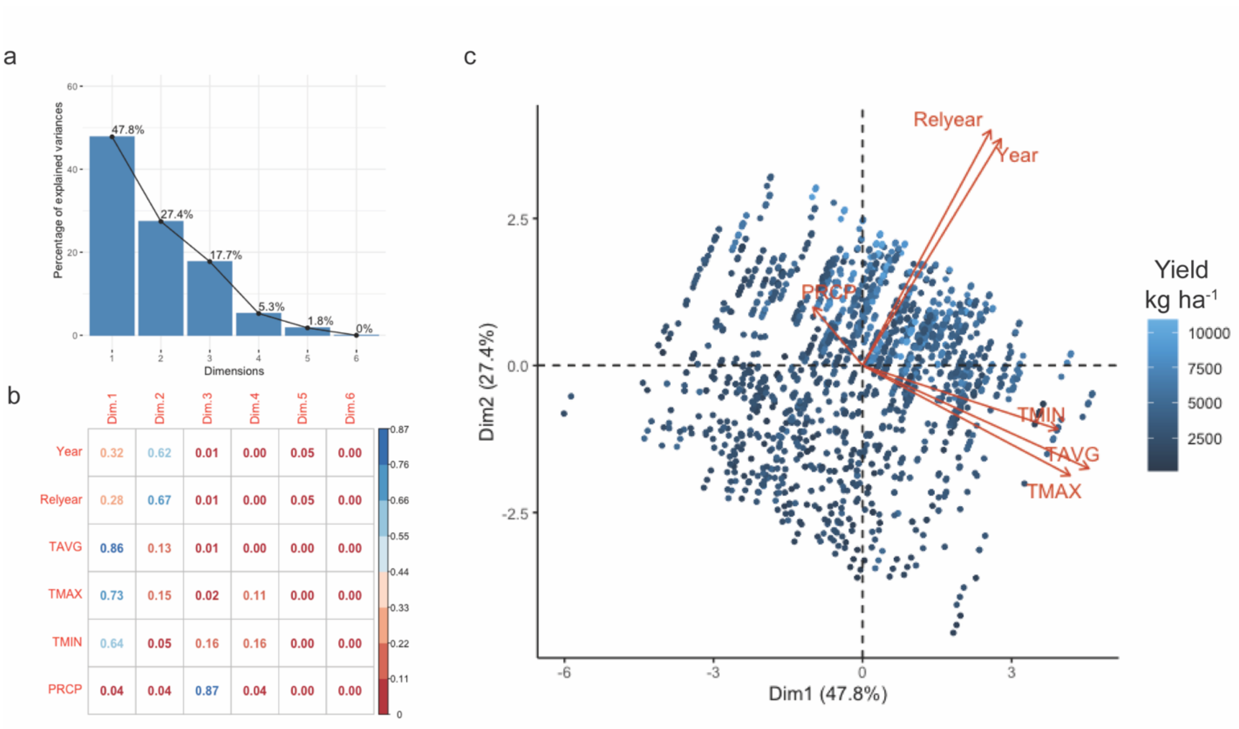

Fig. 5 Non-independence of multiple metrics affecting robustness

a) Bar chart of percent of variance explained per principal component. b) Loadings of metrics to each component (labeled Dim1, Dim2). c) Biplot for each variety tested, colored by yield.

TABLE 4 Partial correlations[a] analysis among yield and potential explanatory factors, all locations

		(ºC)				
	Yield	**Temperature annual mean**	**Temperature annual maximum**	**Temperature annual minimum**	**Precipitation annual cumulative (cm)**	**Year**
Temperature annual mean	-0.0013					
Temperature annual maximum	-0.0024	0.9994***				
Temperature annual minimum	0.0112	0.9984***	-0.9965***			
Precipitation annual cumulative	0.2929***	0.0548*	-0.0682**	-0.0400		
Year	0.2026***	0.0257	-0.0238	-0.0225	-0.1124***	
Release year	0.0583*	-0.0220	0.0218	0.0202	0.0293	0.8551***

* Significant at the .05 probability level, ** Significant at the .01 probability level, *** Significant at the .001 probability level. [a]Pearson correlation coefficients given.

With this firmer basis of how yields are collectively affected by weather, we next aimed to parse out the impact of weather for influencing yield robustness in these data. We revisited the filtered dataset with years grown greater than nine, to avoid the impact of years grown. We obtained the corresponding average annual temperature and annual average precipitation for each variety during the specific years it was grown. Due to missing weather station data, we were unable to include data for some years. We modeled how yields at each location were affected by average temperature and precipitation, the largest contributors to principal components 1 and 3, as well as their interaction. Finally, we recalculated each varieties' yield robustness as the CV_{res} based on the residuals of the models following the approach by Schou et al (2020). Overall, we observed similar patterns as before. Havre continued to show a strong negative relationship between release year and CV_{res} even after removing the impact of average temperature and precipitation (Fig. 6). Sidney and Huntley showed less strong relationships once removing the impact of weather. We detected no relationship between release year and CV_{res} in Bozeman, and in Kalispell, the opposite pattern was again found. In contrast, we noted a non-linear relationship in Moccasin, with a decrease in CV_{res} after the year 2000. Overall, even after some correction for weather differences, we continued to observe location specific impacts on robustness, pointing to many other factors involved in influencing patterns of robustness over time.

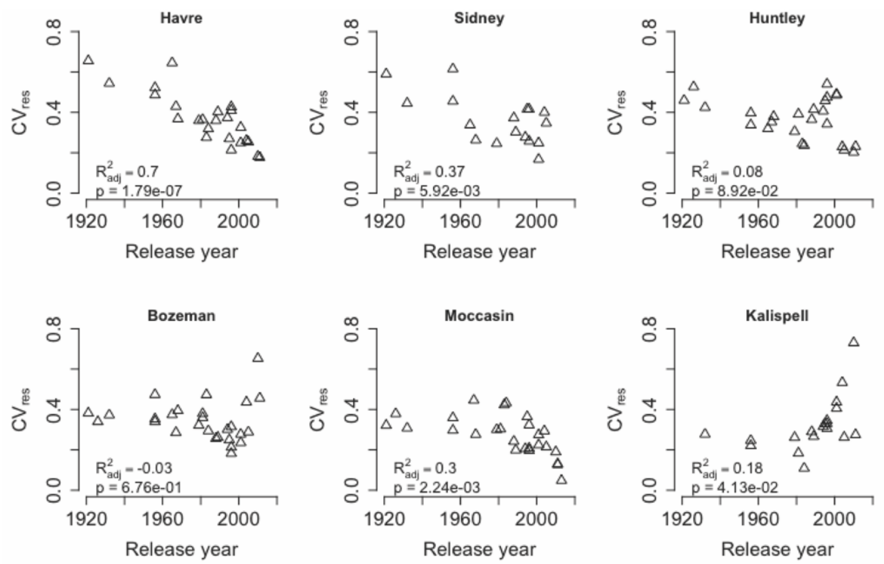

Fig. 6 The relationship between yield CV_{res} and release year of a variety is dependent on location

After modeling the impact of average temperature and precipitation on yield, yield CV_{res} was calculated at each research station for each variety grown at least 10 years and plotted against the variety release year. Adjusted R^2 and p-values for the regression coefficient are displayed.

DISCUSSION

Historical data collected from specific varieties of winter wheat grown at defined locations in Montana enable studies of the change in year-over-year stability, or robustness, of yields over time. It is important to consider to which perturbations the trait is robust (Wagner, 2007). Examining the robustness of repeated organs formed on an individual, such as tillers in wheat, reveals robustness to stochastic errors of development. Examining individual plants within a single plot reveals robustness to microenvironmental variation as well as developmental differences. On a larger scale, measuring robustness across multiple plots captures robustness to field variation, and finally, robustness has also been considered across years, revealing robustness to weather or variation in agronomic practices at the very least. In this experiment, we measured aggregate robustness across these levels (Falconer & Mackay, 1996) and detected tendencies for robustness to increase over time, dependent on breeding, but this was dependent on geographic location.

A simple examination of the residuals of a linear fit of yield over time as a proxy for robustness did not point to improved robustness over time. This approach cannot distinguish varieties, agronomic practices, weather changes or other influences on yield. For example, improved management and other non-genetic inputs at these research sites are known to affect yields. A study of spring wheat at these six research centers from 1950-2007 found that the variety "Thatcher" showed increased grain yields throughout this period, ranging from 14.8 kg ha^{-1} yr^{-1} in Moccasin to 50.8 kg ha^{-1} yr^{-1} in Kalispell (Lanning et al., 2010). However, examining robustness of each variety separately by calculating a standardized variance using the statistic CV showed that more recently released varieties tend to be more robust (Fig. 2, 3). Importantly, estimating robustness requires a sufficiently large dataset. We found that estimates of robustness using the statistic CV was dependent on the number of years a variety was grown. If a variety was only grown for a few years, the CV was not representative of robustness. We detected a plateau of approximately 10 years of data as sufficient to estimate robustness using CV (Fig. S2a). We speculate that this plateau represents the number of years needed for a variety to experience an extreme weather event on average.

Geographic location was a strong indicator of whether robustness increased with breeding. The relationship between robustness and release year is observed at Havre, Sidney, and Huntley (Fig. 3a) and not at the other locations. We noted several potential shared characteristics at Havre, Sidney and Huntley that may contribute to these locations demonstrating improved robustness. First, Havre, Sidney, and Huntley had the lowest starting yields (Fig. 1c) in the 1950s. It is possible that these locations had the most to gain in both yield mean and robustness relative to the other locations. Accordingly, the CV values across varieties tended to be higher and cover a broad range at these locations (Fig. 3b), relative to Moccasin and Bozeman. Notably, Havre, Sidney, and Huntley represent where most of the wheat is grown in the state by producers (National Agricultural Statistics Service, 2021), possibly reflecting a focus of breeders to optimize wheat’s growth to those locations and/or that robustness improvement had the most impact in locations with more extreme temperatures. Indeed, Havre, Sidney, and Huntley have had the highest summer temperatures, with Sidney and Havre also having the lowest winter temperatures (Fig. S4e-g). The greater tolerance of perturbations at these highly variable locations is consistent with observations in other crops, such as maize yields increasing due to reduced plant-to-plant variability (Tollenaar & Wu, 1999). Kalispell stands out from the other locations that comprise the primary focus of the breeding program and represents a region of the state more like eastern Washington. The selected lines are not well-adapted to this region, and the decrease in robustness in Kalispell may reflect this. Pathogen and herbivore pressure also vary over the years and were not considered in this study. Wheat stem sawfly and stripe rust have devastating impacts on yields that can be location specific and breeding has focused on producing resistant varieties that likely contribute to improved robustness. Finally, we also note that management also did vary across research centers: during this period, Havre and Huntley transitioned from tillage to chemical fallow, increasing soil moisture (Lenssen et al., 2007), while other sites maintained tillage, including Sidney.

Weather conditions greatly impact yields of crops, evidenced by both the great impacts that droughts and floods have on yields (Ray et al., 2015) within a region and how wheat yields are sensitive to one-degree changes in temperatures in controlled conditions (Lanning et al., 2010). We surmised that variability in weather conditions from year to year should be reflected in the robustness of crops. Upon examining climate patterns of temperature and precipitation, we found that temperatures are increasing at all research centers, and precipitation is unchanging or increasing (Table 2). Second, the variabilities of weather metrics showed patterns of convergence, divergence, and no change (Table 3). No general statements could be made to relate change in robustness in yields as dependent on weather variability. After removing the effect of weather metrics assessed, we continued to detect relationships between robustness and release year. Havre continued to show a strong negative relationship between release year and CV_{res}. Sidney and Huntley showed less strong relationships once removing the impact of weather. In the case of Sidney, this possibly reflects a tendency towards less variable precipitation (Fig. 4h, pink line) and at Huntley, this could reflect convergence in maximum temperatures (Fig. 4f, orange line).

Overall, in this historical wheat dataset spanning 70 years, we can detect a signal that yield robustness is dependent on breeding and location across many levels of perturbations, from within plant to across years. Because of the extensive confounding of management and environmental conditions, such as global CO_2 levels, the impact of breeding could best be tested compared by growing historical varieties at the same site and time for multiple seasons to estimate robustness. This work reveals how breeding influences robustness and how environment influences its evolution—an inadvertent evolutionary selection experiment.

ACKNOWLEDGMENTS

We thank Nate Silverman, Jason Cook, and Justin Vetch for comments on the manuscript. This work was supported by the National Science Foundation project OIA-1929113, USDA National Institute of Food and Agriculture Hatch project MONB00393, and the Montana Wheat and Barley Committee.

AUTHOR CONTRIBUTIONS

JL: conceptualization, methodology, formal analysis, writing-original draft preparation, writing-reviewing and editing, funding acquisition. JEB: data curation, writing-reviewing and editing. ML: data curation, formal analysis, writing-reviewing and editing.

SUPPLEMENTAL MATERIAL

Supplemental figures and tables are available as Supplemental Material.

CONFLICT OF INTEREST

The authors have no relevant financial or non-financial interests to disclose.

DATA AVAILABILITY

Code generated and data analyzed in this study are available at https://github.com/Lachowiec-Lab/historical-stability.

REFERENCES

Baty, F., Ritz, C., Charles, S., Brutsche, M., Flandrois, J.-P., & Delignette-Muller, M.-L. (2015). A Toolbox for Nonlinear Regression in R: The Package nlstools. *Journal of Statistical Software*, *66*, 1–21. https://doi.org/10.18637/jss.v066.i05

Becker, H. C., & Leon, J. (1988). Stability Analysis in Plant Breeding. *Plant Breeding*, *101*(1), 1–23. https://doi.org/10.1111/j.1439-0523.1988.tb00261.x

Bowman, D. T. (2001). Common use of the CV: a statistical aberration in crop performance trials (Contemporary Issue). *Journal of Cotton Science*, *5*(2), 137–141.

Bruijning, M., Metcalf, C. J. E., Jongejans, E., & Ayroles, J. F. (2020). The Evolution of Variance Control. *Trends in Ecology & Evolution*, *35*(1), 22–33. https://doi.org/10.1016/j.tree.2019.08.005

Falconer, D. S., & Mackay, T. F. C. (1996). Introduction to quantitative genetics. Essex. *UK: Longman Group*.

Fasoula, V. A., & Fasoula, D. A. (2002). Principles underlying genetic improvement for high and stable crop yield potential. *Field Crops Research*, *75*(2–3), 191–209. http://dx.doi.org/10.1016/S0378-4290(02)00026-6

Finlay, K. W., & Wilkinson, G. N. (1963). The analysis of adaptation in a plant-breeding programme. *Australian Journal of Agricultural Research*, *14*(6), 742–754. https://doi.org/10.1071/ar9630742

Fisher, J., & Zamir, D. (2021). Genes for Yield Stability in Tomatoes. *Advanced Genetics*, *2*(4), 2100049. https://doi.org/10.1002/ggn2.202100049

Gavrilets, S., & Hastings, A. (1994). A Quantitative-Genetic Model for Selection on Developmental Noise. *Evolution*, *48*(5), 1478–1486. https://doi.org/10.1111/j.1558-5646.1994.tb02190.x

Gollin, D. (2006). *Impacts of International Research on Intertemporal Yield Stability in Wheat and Maize: An Economic Assessment*. CIMMYT.

Hall, M. C., Dworkin, I., Ungerer, M. C., & Purugganan, M. (2007). Genetics of microenvironmental canalization in Arabidopsis thaliana. *Proc Natl Acad Sci U S A*, *104*(34), 13717–13722. https://doi.org/0701936104 [pii] 10.1073/pnas.0701936104

Hill, W. G., & Mulder, H. A. (2010). Genetic analysis of environmental variation. *Genetics Research*, *92*(5–6), 381–395. https://doi.org/10.1017/S0016672310000546

Jimenez-Gomez, J. M., Corwin, J. A., Joseph, B., Maloof, J. N., & Kliebenstein, D. J. (2011). Genomic analysis of QTLs and genes altering natural variation in stochastic noise. *PLoS Genet*, *7*(9), e1002295. https://doi.org/10.1371/journal.pgen.1002295 PGENETICS-D-11-00547 [pii]

Kertho, A. O. (2014). *Evaluation of winter wheat germplasm for resistance to stripe rust and leaf rust*. North Dakota State University.

Kristensen, L., Olsen, J., & Weiner, J. (2008). Crop Density, Sowing Pattern, and Nitrogen Fertilization Effects on Weed Suppression and Yield In Spring Wheat. *Weed Science*, *56*(1), 97–102. https://doi.org/10.1614/WS-07-065.1

Kukal, M. S., & Irmak, S. (2018). Climate-Driven Crop Yield and Yield Variability and Climate Change Impacts on the U.S. Great Plains Agricultural Production. *Scientific Reports*, *8*(1), 3450. https://doi.org/10.1038/s41598-018-21848-2

Lachowiec, J., Queitsch, C., & Kliebenstein, D. J. (2015). Molecular mechanisms governing differential robustness of development and environmental responses in plants. *Annals of Botany*, *117*(5), 795–809. https://doi.org/10.1093/aob/mcv151

Lanning, S. P., Kephart, K., Carlson, G. R., Eckhoff, J. E., Stougaard, R. N., Wichman, D. M., Martin, J. M., & Talbert, L. E. (2010). Climatic Change and Agronomic Performance of Hard Red Spring Wheat from 1950 to 2007. *Crop Science*, *50*(3), 835–841. https://doi.org/10.2135/cropsci2009.06.0314

Lenssen, A. W., Johnson, G. D., & Carlson, G. R. (2007). Cropping sequence and tillage system influences annual crop production and water use in semiarid Montana, USA. *Field Crops Research*, *100*(1), 32–43. https://doi.org/10.1016/j.fcr.2006.05.004

Lobell, D. B., Burke, M. B., Tebaldi, C., Mastrandrea, M. D., Falcon, W. P., & Naylor, R. L. (2008). Prioritizing climate change adaptation needs for food security in 2030. *Science 319*(5863), 607–610. https://doi.org/10.1126/science.1152339

Lu, P., Jiang, B., & Weiner, J. (2020). Chapter Three—Crop spatial uniformity, yield and weed suppression. In D. L. Sparks (Ed.), *Advances in Agronomy* (Vol. 161, pp. 117–178). Academic Press. https://doi.org/10.1016/bs.agron.2019.12.003

Macholdt, J., Hadasch, S., Piepho, H.-P., Reckling, M., Taghizadeh-Toosi, A., & Christensen, B. T. (2021). Yield variability trends of winter wheat and spring barley grown during 1932–2019 in the Askov Long-term Experiment. *Field Crops Research*, *264*, 108083. https://doi.org/10.1016/j.fcr.2021.108083

Makumburage, G. B., & Stapleton, A. E. (2011). Phenotype uniformity in combined-stress environments has a different genetic architecture than in single-stress treatments. *Frontiers in Plant Science*, *2*, 12. https://doi.org/10.3389/fpls.2011.00012

McGrath, J. M., & Lobell, D. B. (2013). Regional disparities in the CO 2 fertilization effect and implications for crop yields. *Environmental Research Letters*, *8*(1), 014054. https://doi.org/10.1088/1748-9326/8/1/014054

National Agricultural Statistics Service. (2021). *U.S. National Agricultural Statistics Service NASS.* [Web Archive]. https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=MONTANA

Ordas, B., Malvar, R. A., & Hill, W. G. (2008). Genetic variation and quantitative trait loci associated with developmental stability and the environmental correlation between traits in maize. *Genet Res (Camb)*, *90*(5), 385–395. https://doi.org/10.1017/S0016672308009762 S0016672308009762 [pii]

R Development Core Team, R. (2011). R: A Language and Environment for Statistical Computing. In R. D. C. Team (Ed.), *R Foundation for Statistical Computing* (Vol. 1, Issue 2.11.1, p. 409). R Foundation for Statistical Computing. https://doi.org/10.1007/978-3-540-74686-7

Ray, D. K., Gerber, J. S., MacDonald, G. K., & West, P. C. (2015). Climate variation explains a third of global crop yield variability. *Nature Communications*, *6*(1), 5989. https://doi.org/10.1038/ncomms6989

Reckling, M., Ahrends, H., Chen, T.-W., Eugster, W., Hadasch, S., Knapp, S., Laidig, F., Linstädter, A., Macholdt, J., Piepho, H.-P., Schiffers, K., & Döring, T. F. (2021). Methods of yield stability analysis in long-term field experiments. A review. *Agronomy for Sustainable Development*, *41*(2), 27. https://doi.org/10.1007/s13593-021-00681-4

Riveland, N. R., Berg, J. E., Kephart, K. D., Wichman, D. M., Carlson, G. R., Kushnak, G. D., Stougaard, R. N., Eckhoff, J. L., Nash, D. L., Johnston, M., Grey, W. E., Jin, Y., Chen, X., & Bruckner, P. L. (2011). Registration of 'Decade' Wheat. *Journal of Plant Registrations*, *5*(3), 345–348. https://doi.org/10.3198/jpr2011.04.0191crc

Sangster, T., Salathia, N., Undurraga, Milo, R., Schellenberg, K., Lindquist, S. L., & Queitsch, C. (2008). HSP90 affects the expression of genetic variation and developmental stability in quantitative traits. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(8), 2963–2968. https://doi.org/10.1073/pnas.0712200105

Schou, M. F., Kristensen, T. N., & Hoffmann, A. A. (2020). Patterns of environmental variance across environments and traits in domestic cattle. *Evolutionary Applications*, *13*(5), 1090–1102. https://doi.org/10.1111/eva.12924

Tollenaar, M., & Wu, J. (1999). Yield Improvement in Temperate Maize is Attributable to Greater Stress Tolerance. *Crop Science*, *39*(6), 1597–1604. https://doi.org/10.2135/cropsci1999.3961597x

U.S. Climate Normals. (2021, February 25). National Centers for Environmental Information (NCEI). http://www.ncei.noaa.gov/products/land-based-station/us-climate-normals

Wagner, A. (2007). *Robustness and evolvability in living systems*. Princeton University Press. https://press.princeton.edu/titles/8002.html

Wagner, G. P., Booth, G., & Bagheri-Chaichian, H. (1997). A population genetic theory of canalization. *Evolution*, 329–347.

Whitlock, C., Cross, W., Maxwell, B., Silverman, N., & Wade, A. (2017). *2017 Montana Climate Assessment* (p. 318 p.). https://doi.org/doi:10.15788/m2ww8w

Zhang, F., Biederman, J. A., Dannenberg, M. P., Yan, D., Reed, S. C., & Smith, W. K. (2021). Five Decades of Observed Daily Precipitation Reveal Longer and More Variable Drought Events Across Much of the Western United States. *Geophysical Research Letters*, *48*(7), e2020GL092293. https://doi.org/10.1029/2020GL092293