

## Appendix A

Monthly average of 2m temperature metrics for October 1979

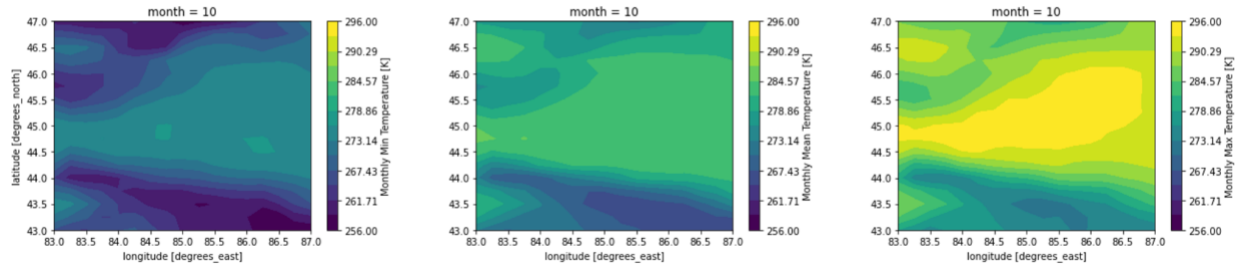


Figure 1: sample plot of 2m temperature data from ERA5

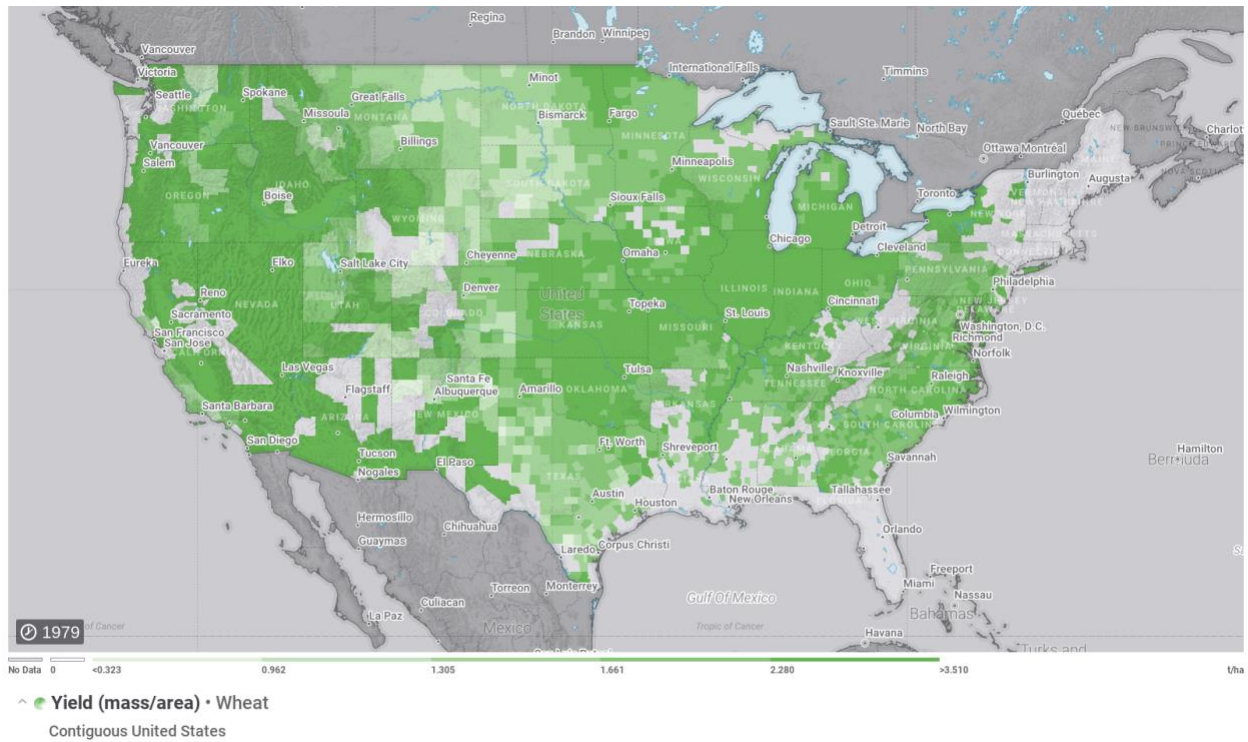


Figure 2: sample of wheat yield data by district from gro-intelligence for the year 1979.

Parameter	Source	Units	Time Range	Format
Annual Wheat Yield by district, contiguous US	Gro-Intelligence (USDA NASS Crops)	tonne per hectare	1955-2007	Excel file (start date, end date, district, value)
Monthly mean 2m relative humidity	ERA5, accessed on tigercpu (/tigress/wenchang/data/era5)	percent	1979-2018	NetCDF4 (latitude, longitude, time, value)
Daily mean 2m temperature	ERA5, accessed on tigercpu (/tigress/wenchang/data/era5)	K	1979-2020	NetCDF4 (latitude, longitude, time, value)
Daily maximum 2m temperature	ERA5, accessed on tigercpu (/tigress/wenchang/data/era5)	K	1979-2020	NetCDF4 (latitude, longitude, time, value)
Daily minimum 2m temperature	ERA5, accessed on tigercpu (/tigress/wenchang/data/era5)	K	1979-2020	NetCDF4 (latitude, longitude, time, value)
Monthly mean total precipitation rate	ERA5, accessed on tigercpu (/tigress/wenchang/data/era5)	mm/day	1979-2020	NetCDF4 (latitude, longitude, time, value)
Daily mean surface downward UV radiation flux	ERA5, accessed on tigercpu (/tigress/wenchang/data/era5)	W/m <sup>2</sup>	1979-2021	NetCDF4 (latitude, longitude, time, value)

Table 1: description of all data sets.

Feature	Description	Dependencies
Mean 2m temperature	Daily 2m temperature averaged over the month	Year, latitude, longitude
Maximum 2m temperature	Maximum daily 2m temperature value during the month	Year, latitude, longitude
Minimum 2m temperature	Minimum daily 2m temperature value during the month	Year, latitude, longitude
Mean total precipitation rate	Mean total precipitation rate averaged over the month	Year, latitude, longitude
Maximum mean surface downward UV radiation flux (msdwuvrf)	Maximum daily msdwuvrf value during the month	Year, latitude, longitude
Mean mean surface downward UV radiation flux	Daily msdwuvrf value averaged over the month	Year, latitude, longitude
Minimum mean surface downward UV radiation flux	Minimum daily msdwuvrf value during the month	Year, latitude, longitude
Mean relative humidity	Daily relative humidity value averaged over the month	Year, latitude, longitude
Mean 2m temperature anomaly	Mean daily 2m temperature for the month difference from expected mean 2m temperature for that month	Year, latitude, longitude
Maximum 2m temperature anomaly	Maximum daily 2m temperature for the month difference from expected maximum 2m temp for that month	Year, latitude, longitude
Minimum 2m temperature anomaly	Minimum daily 2m temperature for the month difference from expected minimum 2m temp for that month	Year, latitude, longitude
Mean total precipitation rate anomaly	Mean total precipitation rate for the month difference from expected mean total precipitation rate for the month	Year, latitude, longitude
Maximum mean surface downward UV radiation flux anomaly	Maximum daily msdwuvrf for the month difference from expected maximum daily msdwuvrf for that month	Year, latitude, longitude
Mean mean surface downward UV radiation flux anomaly	Mean daily msdwuvrf for the month difference from expected mean daily msdwuvrf for that month	Year, latitude, longitude
Minimum mean surface downward UV radiation flux anomaly	Minimum daily msdwuvrf for the month difference from expected minimum daily msdwuvrf for that month	Year, latitude, longitude
Mean relative humidity anomaly	Mean daily relative humidity for the month difference from expected mean daily relative humidity for that month	Year, latitude, longitude

Table 2: description of all features extracted from ERA5 data. It should be noted that within each row, each listed feature actually contains 12 total features, because I have considered all features on a monthly basis as being potential influences on annual wheat yields. Anomalies are calculated for each month of each year as the difference between the average value of that feature in the corresponding month across all years in the dataset.

## Appendix B

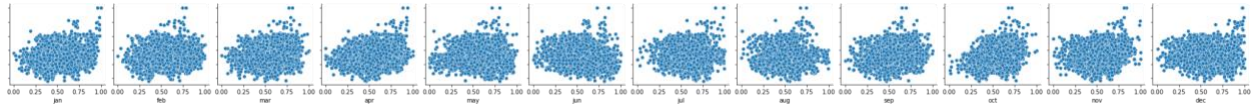


Figure 1: pairplot of annual yield versus mean 2m temperature for each month.

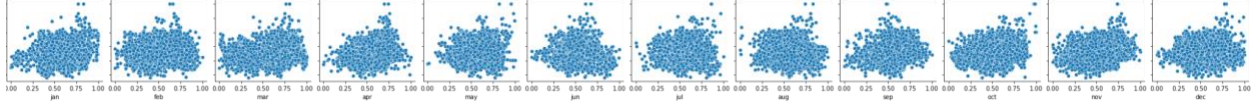


Figure 2: pairplot of annual yield versus maximum 2m temperature for each month.

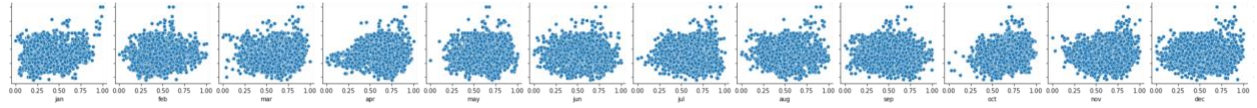


Figure 3: pairplot of annual yield versus minimum 2m temperature for each month.

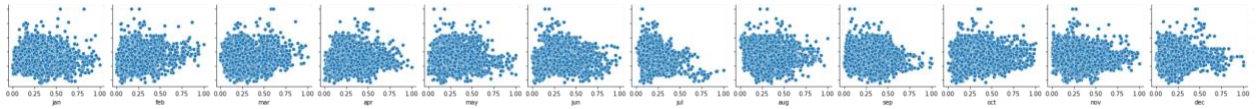


Figure 4: pairplot of annual yield versus mean total precipitation rate for each month.

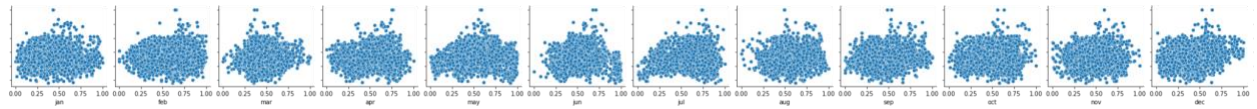


Figure 5: pairplot of annual yield versus mean msdwuvrf for each month.

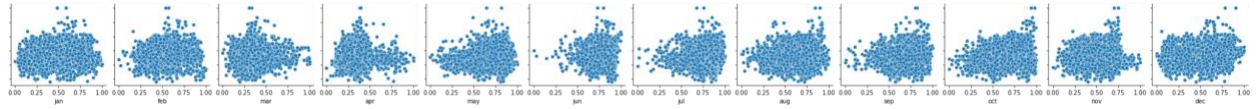


Figure 6: pairplot of annual yield versus maximum msdwuvrf for each month.

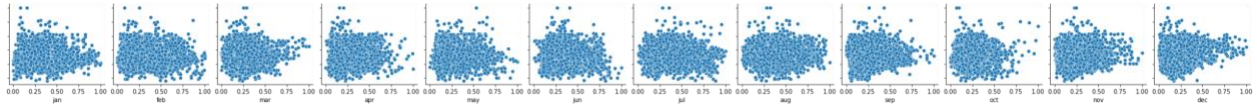


Figure 7: pairplot of annual yield versus minimum msdwuvrf for each month.

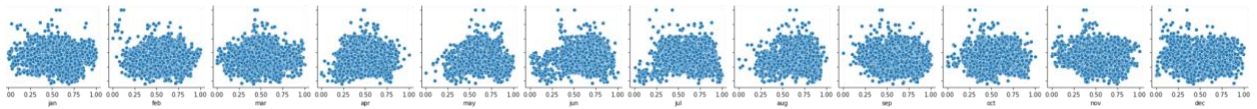


Figure 8: pairplot of annual yield versus mean relative humidity for each month.

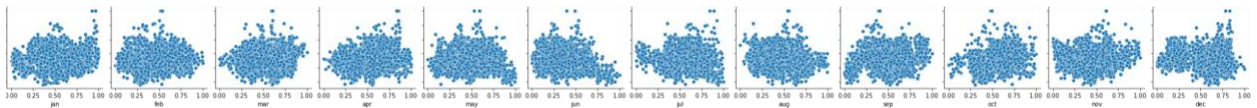


Figure 9: pairplot of annual yield versus mean 2m temperature anomaly for each month.

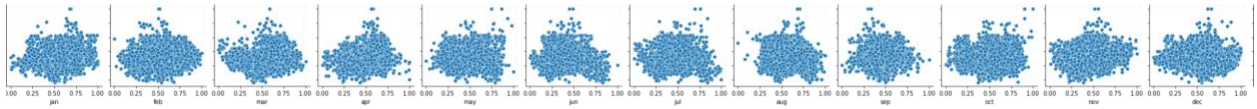


Figure 10: pairplot of annual yield versus maximum 2m temperature anomaly for each month.

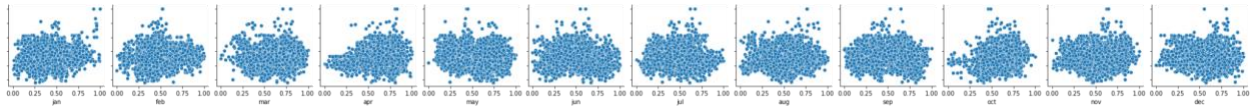




Figure 11: pairplot of annual yield versus minimum 2m temperature anomaly for each month.

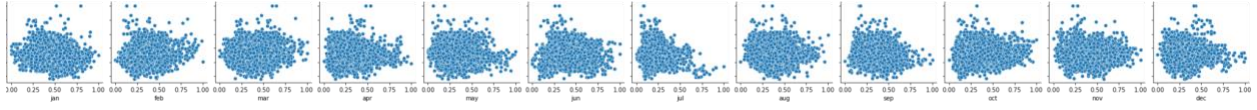


Figure 12: pairplot of annual yield versus mean total precipitation rate anomaly for each month.

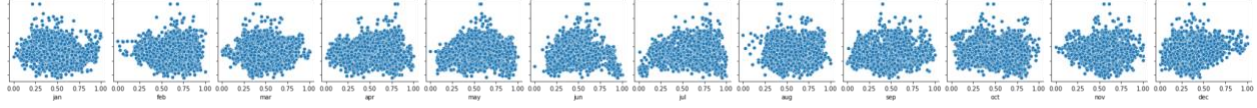


Figure 13: pairplot of annual yield versus mean msdwuvrf anomaly for each month.

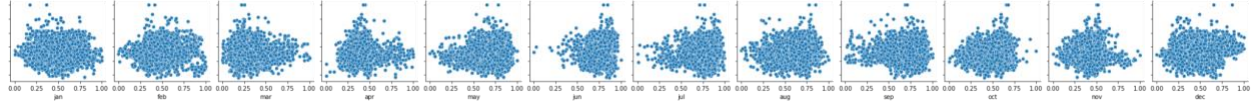


Figure 14: pairplot of annual yield versus maximum msdwuvrf anomaly for each month.

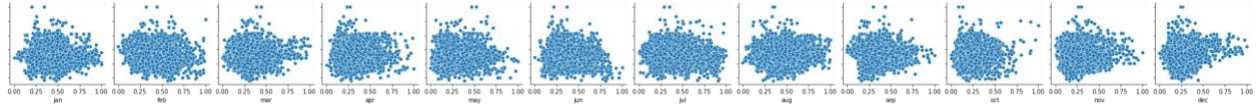


Figure 15: pairplot of annual yield versus minimum msdwuvrf anomaly for each month.

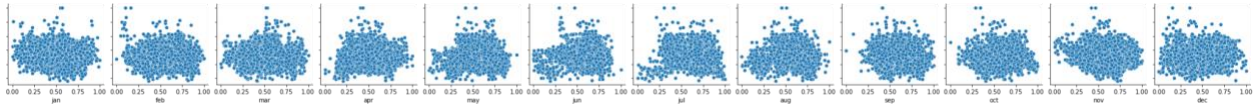


Figure 16: pairplot of annual yield versus mean relative humidity anomaly for each month.

FILE	FEATURE	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
monthly_avg_temp.xlsx	Mean 2m temperature	1	0	0	1	0	0	0	0	0	1	0	0
monthly_max_t2m.xlsx	Maximum 2m temperature	1	0	1	1	0	0	0	0	0	0	1	1
monthly_min_t2m.xlsx	Minimum 2m temperature	1	0	0	0	0	0	0	0	0	1	1	0
monthly_precip_rate.xlsx	Mean total precipitation rate	0	1	1	1	0	0	1	0	0	1	0	0
monthly_uv_mean.xlsx	Mean msdwuvrf	0	0	0	1	1	1	1	0	0	0	0	1
monthly_uv_max.xlsx	Maximum msdwuvrf	0	0	0	0	0	0	0	0	0	0	0	0
monthly_uv_min.xlsx	Minimum msdwuvrf	0	0	0	0	0	0	0	0	0	0	0	0
monthly_avgt2m_anom.xlsx	Mean 2m temperature anomaly	1	0	0	0	1	1	0	0	1	1	1	0
monthly_maxt2m_anom.xlsx	Max 2m temperatur anomaly	0	0	0	0	0	1	1	1	0	0	1	0
monthly_mint2m_anom.xlsx	Min 2m temperature anomaly	0	0	0	0	1	1	0	0	0	1	0	0
monthly_precipr_anom.xlsx	Precipitation rate anomaly	0	1	0	0	1	1	1	0	0	1	0	1
monthly_uvmean_anom.xlsx	Mean msdwuvrf anomaly	0	0	1	1	1	1	1	0	0	1	0	1
monthly_uvmax_anom.xlsx	Maximum msdwuvrf anomaly	0	0	0	0	0	0	0	0	0	0	0	0
monthly_uv_min_anom.xlsx	Minimum msdwuvrf anomaly	0	0	0	0	0	0	0	1	1	0	1	1
monthly_rh_mean.xlsx	Mean relative humidity	0	0	0	1	1	1	1	0	0	0	1	0
monthly_rh_mean_anom.xlsx	Mean relative humidity anomaly	1	0	0	0	1	1	1	1	0	0	1	0

Table 1: initial feature selection based on inspection of pairplots during exploratory data analysis. A 1 indicates that the feature is used in model training. A 0 indicates that the feature is removed prior to model training.

## Appendix C:

Model	Description
A	Multivariate linear regression with manual feature selection (by first recursively dropping features with p-values above a threshold, and next by recursively dropping features with Variance Inflation Factor above a threshold)
B	Multivariate linear regression with variance threshold feature selection
C	Multivariate linear regression with recursive feature elimination
D	Multivariate linear regression with sequential feature selection
E	Multivariate polynomial regression (degree 2). First, features are selected by sequential feature selection. Next, features are transformed to polynomial degree 2.
F	Multivariate polynomial regression (degree 2). First, features are transformed to polynomial degree 2. Next, choose features based on variance selection.

Table 1: description of all model types trained.

Model		Input parameters		Model Analysis				
A	Run	Max allowed P-value	Max allowed VIF	R2 score training	R2 score validation	MSE training	MSE validation	Number of features
	1	0.01	5	-0.53	-0.62	0.61	0.63	6
	2	0.01	12	0.08	0.12	0.37	0.35	13
	3	0.01	20	0.15	0.14	0.34	0.34	17
	4	0.03	12	0.13	0.1	0.35	0.35	15
	5	0.05	12	0.1	0.08	0.36	0.36	15
	6	0.01	50	0.33	0.26	0.27	0.29	26
s								
B	Run	Variance Threshold		R2 score training	R2 score validation	MSE training	MSE validation	Number of features
	1	0.02		0.48	0.41	0.21	0.23	40
	2	0.03		0.28	0.22	0.29	0.31	29
	3	0.04		-0.37	-0.54	0.55	0.6	5
C	Run	Number of features		R2 score training	R2 score validation	MSE training	MSE validation	
	1	20		0.3	0.31	0.28	0.27	
	2	30		0.42	0.42	0.23	0.23	
	3	40		0.47	0.39	0.21	0.24	
	4	50		0.48	0.48	0.21	0.21	
D	Run	Number of features		R2 score training	R2 score validation	MSE training	MSE validation	
	1	5		-0.04	-0.17	0.42	0.45	
	2	10		0.3	0.24	0.28	0.3	
	3	15		0.35	0.32	0.26	0.27	
	4	20		0.41	0.38	0.24	0.24	
	5	30		0.47	0.39	0.21	0.24	
	6	40		0.47	0.46	0.21	0.21	
E	Run	Number of features (for SFS)		R2 score training	R2 score validation	MSE training	MSE validation	Number of features (final)
	1	7		0.35	0.31	0.26	0.27	36
	2	10		0.44	0.35	0.22	0.26	66
	3	15		0.55	0.49	0.18	0.2	136
	4	20		0.63	0.55	0.15	0.18	231
	5	24		0.68	0.55	0.13	0.18	325
F	Run	Variance Threshold		R2 score training	R2 score validation	MSE training	MSE validation	Number of features
	1	0.03		0.28	0.22	0.29	0.31	29
	2	0.025		0.35	0.32	0.26	0.27	38
	3	0.02		0.46	0.41	0.21	0.23	57
	4	0.015		0.46	0.45	0.22	0.21	60

Table 2: description of all model training statistics from varying input parameters.



Figure 1: performance metrics of all model runs on the validation set plotted against the number of features.

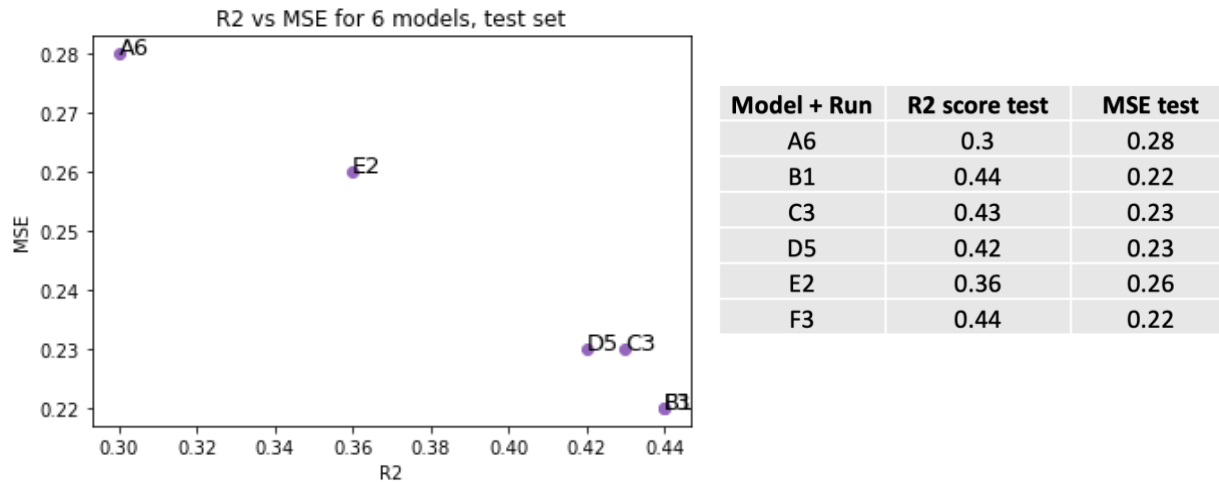


Figure 1: performance metrics of final models on test set.

## References:

1. Ansarifard, J., Wang, L. and Archontoulis, S.V. (2021) 'An interaction regression model for crop yield prediction', *Scientific Reports*, 11(1), p. 17754. doi:[10.1038/s41598-021-97221-7](https://doi.org/10.1038/s41598-021-97221-7).
2. *Farming 101: Planting Wheat* (2019) *Successful Farming*. Available at: <https://www.agriculture.com/crops/wheat/farming-101-planting-wheat> (Accessed: 8 December 2021).
3. Guillory, A. (2017) *ERA5, ECMWF*. Available at: <https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5> (Accessed: 10 December 2021).

4. Haqiqi, I. *et al.* (2021) 'Quantifying the impacts of compound extremes on agriculture', *Hydrology and Earth System Sciences*, 25(2), pp. 551–564. doi:[10.5194/hess-25-551-2021](https://doi.org/10.5194/hess-25-551-2021).
5. Searchinger, T. *et al.* (2019) *Creating a Sustainable Food Future*. Available at: <https://www.wri.org/research/creating-sustainable-food-future> (Accessed: 10 December 2021).
6. *Wheat Sector at a Glance* (2021) *USDA Economic Research Service*. Available at: <https://www.ers.usda.gov/topics/crops/wheat/wheat-sector-at-a-glance/> (Accessed: 9 December 2021).
7. *Wheat Yield, Contiguous US* (no date) *Gro-Intelligence*. Available at: <https://app.gro-intelligence.com/displays/wRL7rggqp> (Accessed: 10 December 2021).

I pledge my honor that this work represents my own in accordance with University guidelines.  
Demetra Yancopoulos