“Plugin”-based architecture of software to predict corn phenotypes

Francisco Munoz-Arriola[1,2]
Diego Jarquin[3]
Hallie Hohbein[4]
Parisa Sarzaeim[4]
Joseph Carter[4]
David Recic[4]
Zoe Trautman[4]
Anna Zhang[4]
Byrav Ramamurthy[4]

[1]Department of Biological Systems Engineering,
[2] School of Natural Resources
[3]Department of Agronomy and Horticulture
[4]Department of Computer Sciences and Engineering

2020 G2F Collaborator’s Meeting, Phenome Meeting

Acknowledgements

This project was supported by the USDA National Institute of Food and Agriculture, **Plant Health and Production and Plant Products: Plant Breeding for Agricultural Production, A1211). Accession No.1015252**

Some ideas are associated with the USDA National Institute of Food and Agriculture, Agriculture and Food Research Initiative HATCH project NEB-21-166 Accession No. No.1009760

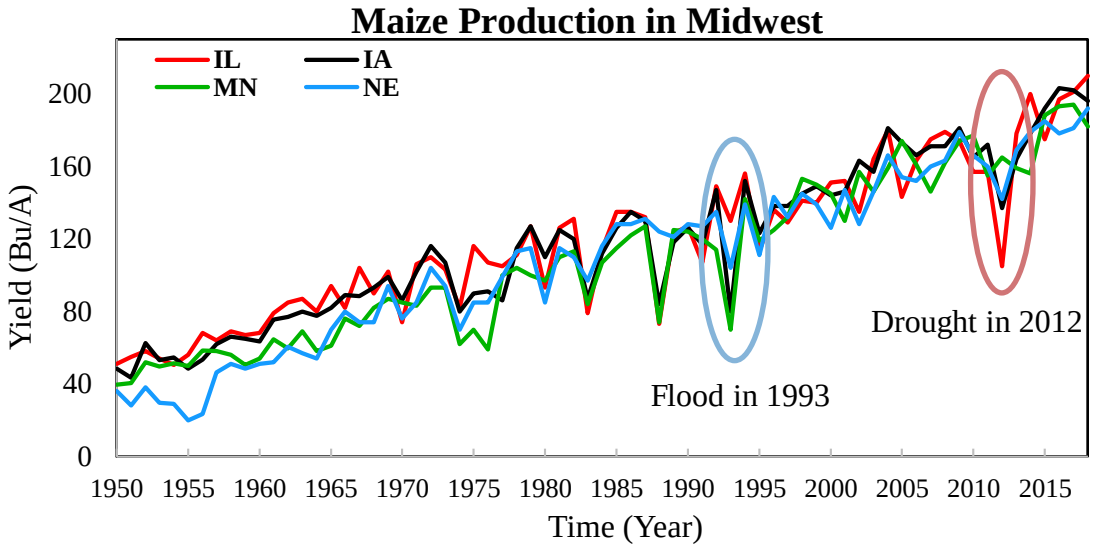
Genomes to Fields initiative

UNL's Department of Computer Sciences and Engineering Senior Design

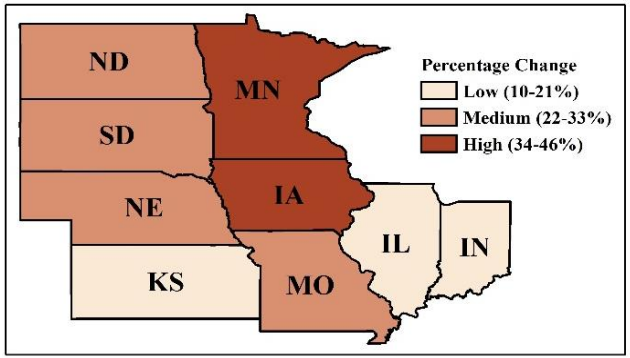
Motivation

- Consistent increase of water use efficiency, farmers revenues and yields

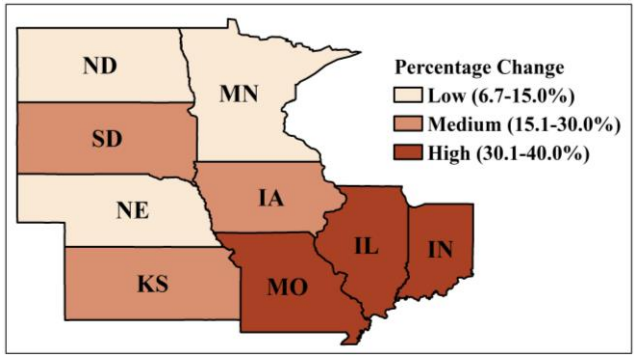
- Drops in water use efficiency, farmers revenues and yields after the occurrence of floods and droughts



Maize Yield Reduction (1992 -1993)



Maize Yield Reduction (2010 -2012)

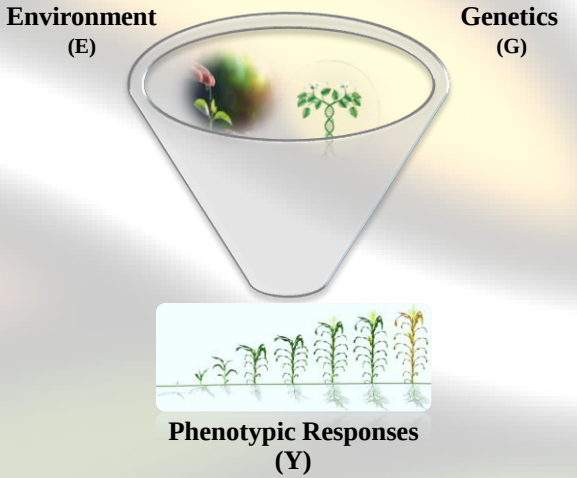


Ref: USDA NASS

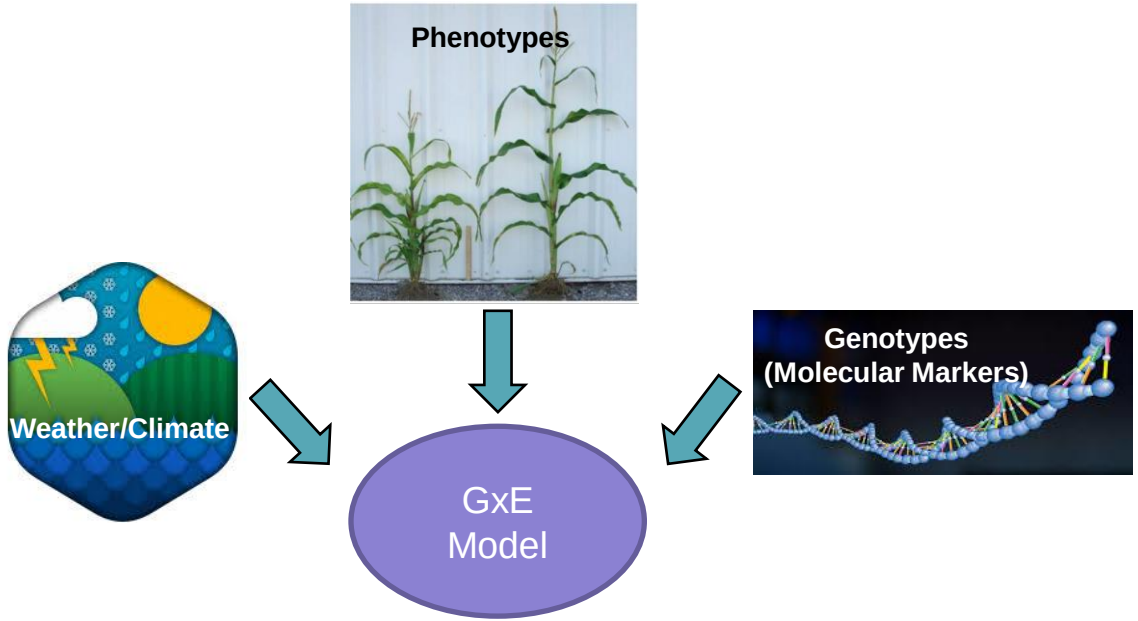
Outline

- Framework
- G2F
- Software Architecture
  - Preprocessing
  - Option Selection
  - Processing
  - Postprocessing
- Software Demo
- Complexities
- Conclusion
- Future Work

Framework



Develop a framework to collect, store, manage, and use weather/climate data to predict plant phenotypes using GxE model

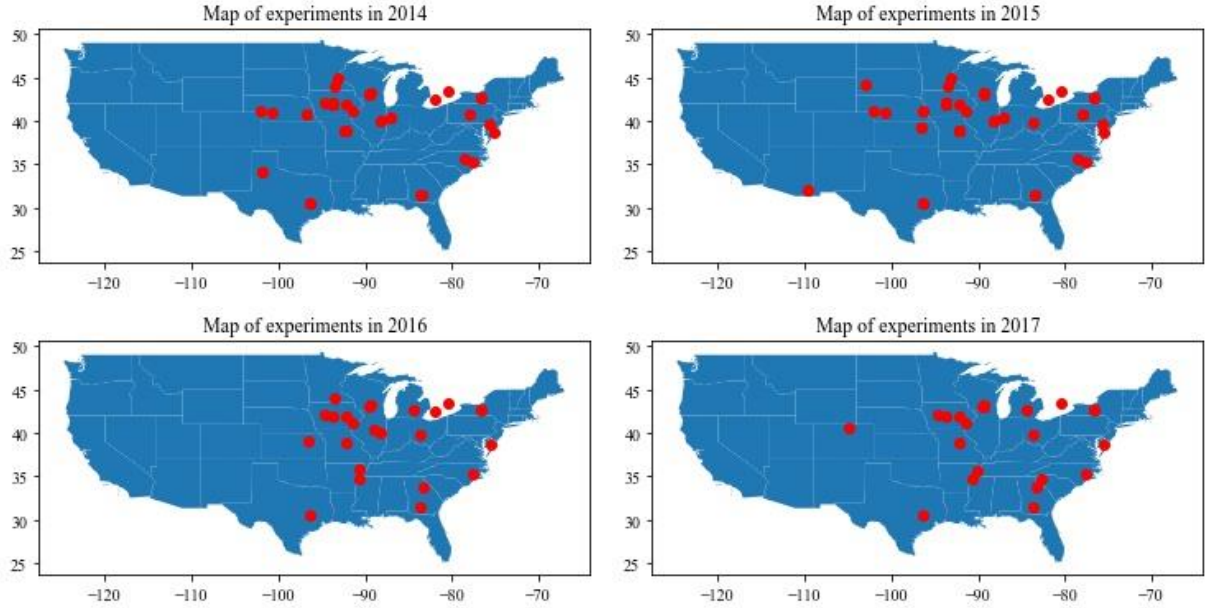


G2F

- Genetics
- Environments
- Phenotypes (traits)

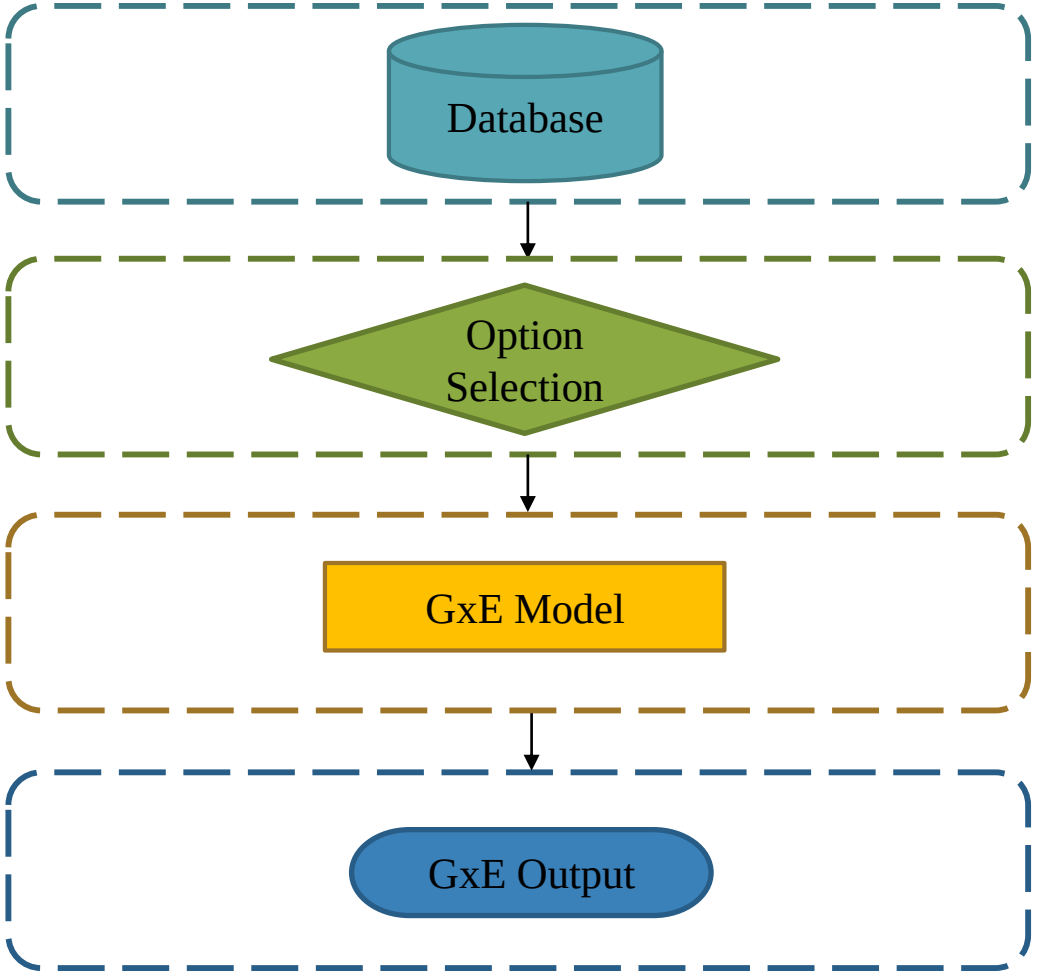
Incorporation of Environmental Information to Improve Phenotypic Predictability in Maize G2F-GxE Hybrid Project

G2F Experiments Distribution



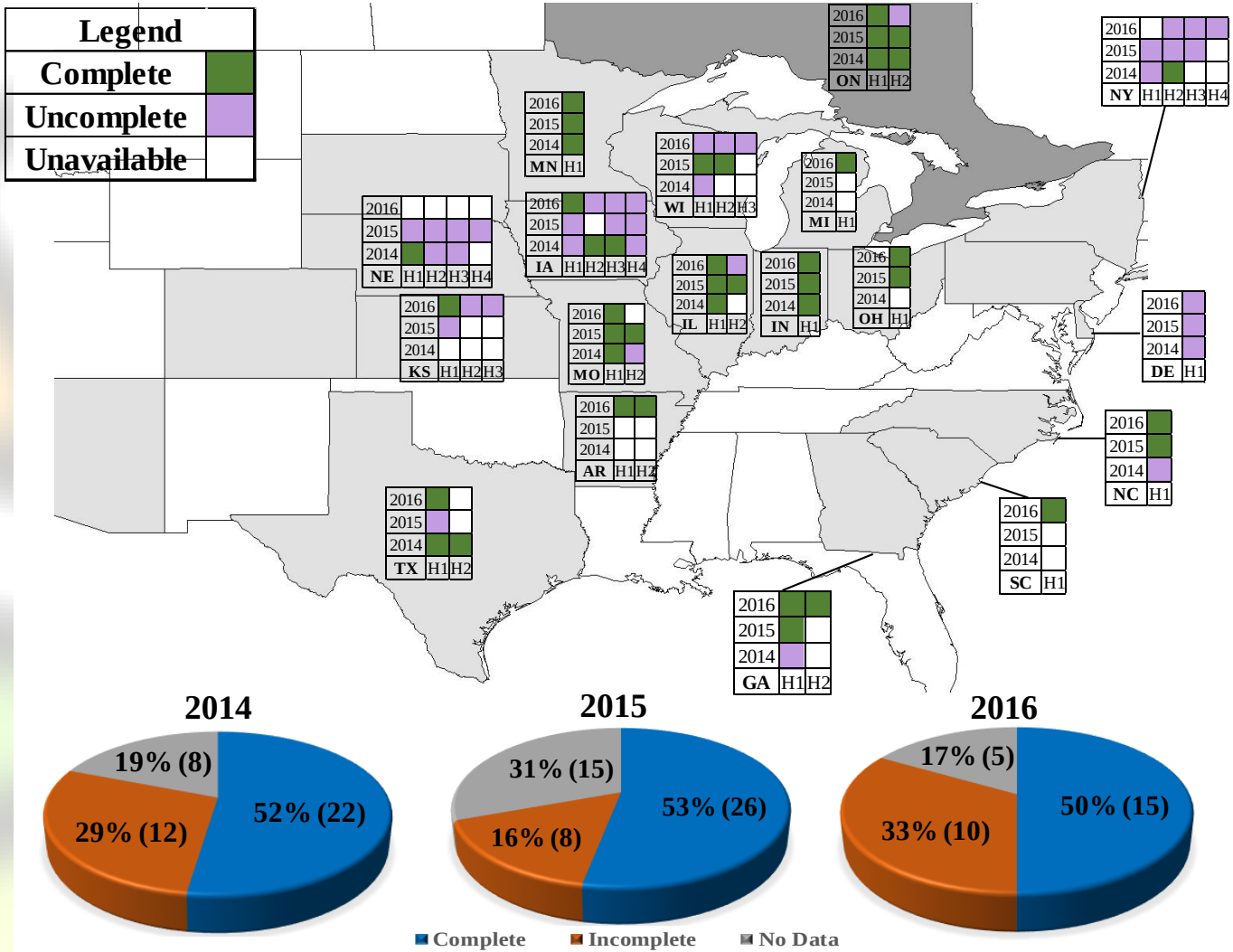
Software Architecture

- Preprocessing
- Selection
- Processing
- Postprocessing



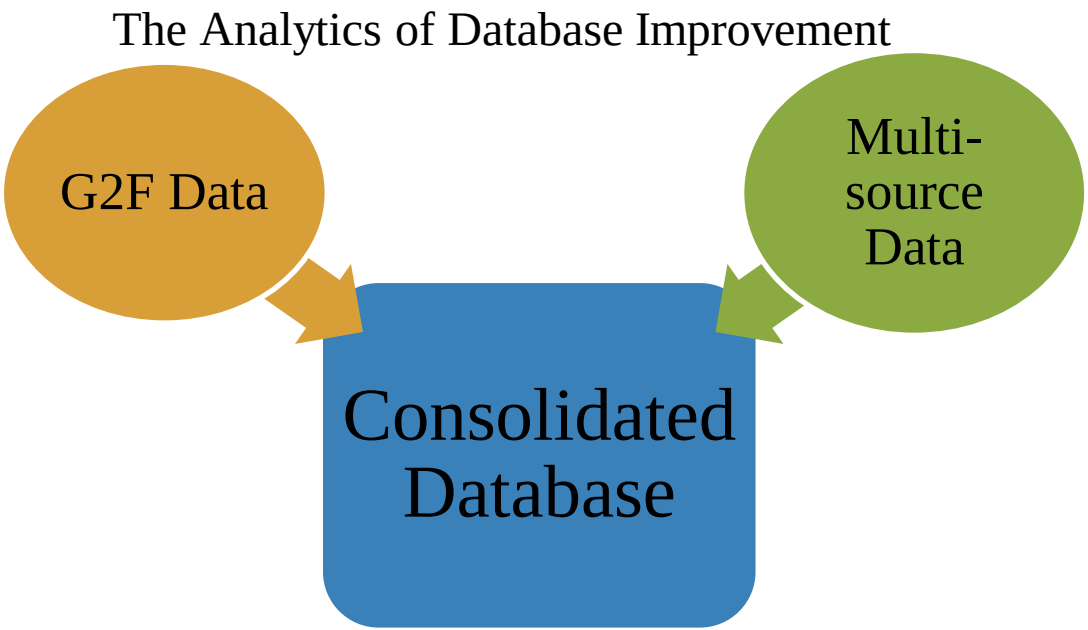
Database

- Temperature
- Dew Point
- Relative Humidity
- Solar Radiation
- Rainfall
- Wind Speed
- Wind Direction
- Wind Gust



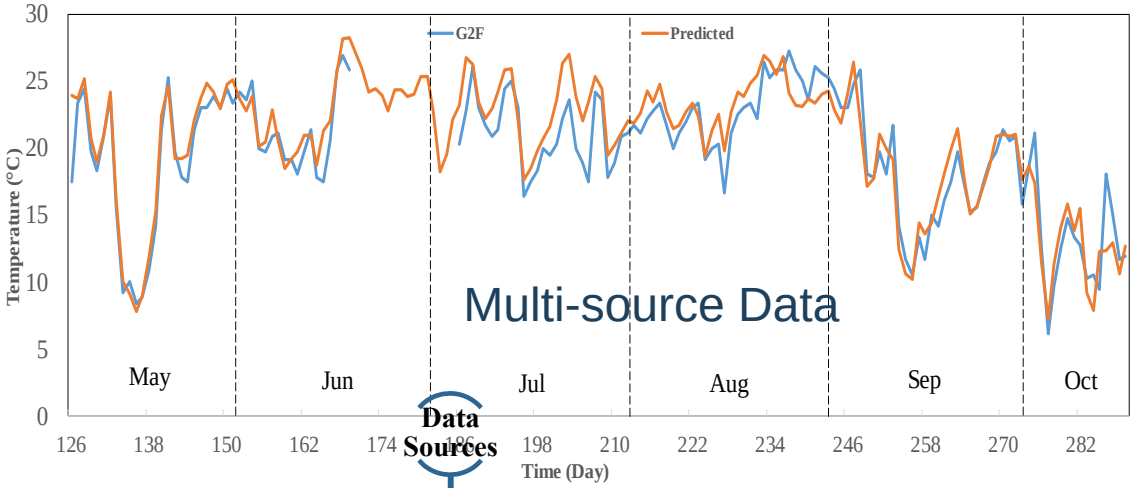
Pre-processing

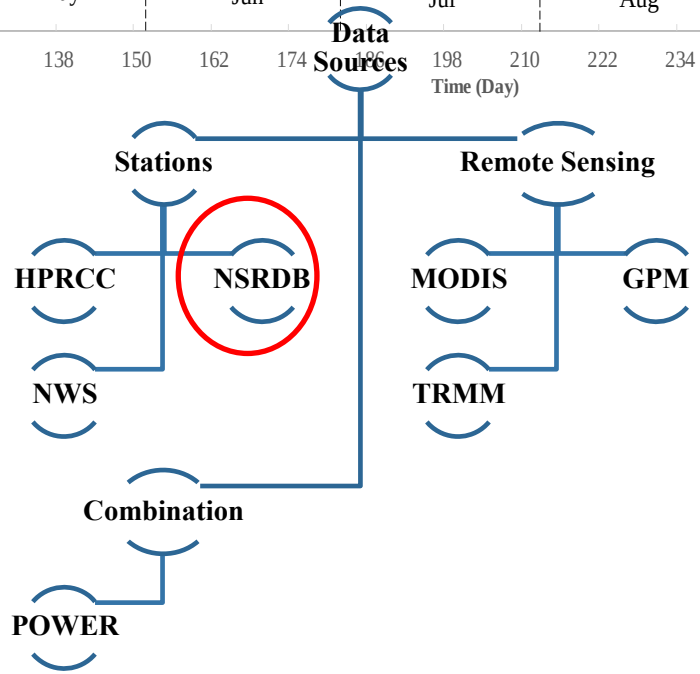
1. **Integration of various data sources**
2. **Correction the data**
3. **Synthesis the data**



Data consolidation

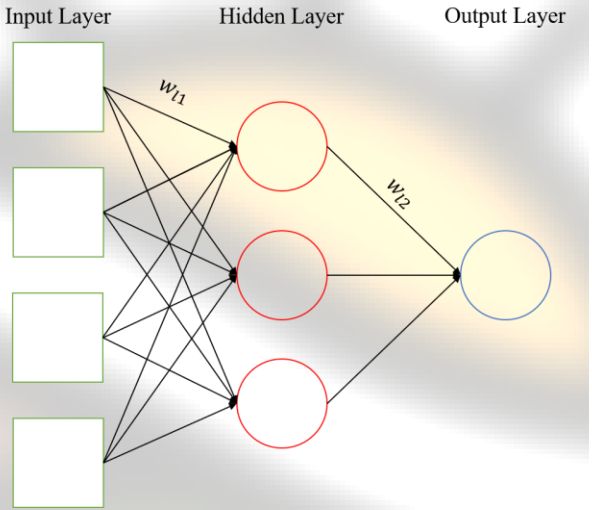
- Temperature
- Dew Point
- Relative Humidity
- Solar Radiation
- Rainfall
- Wind Speed
- Wind Direction

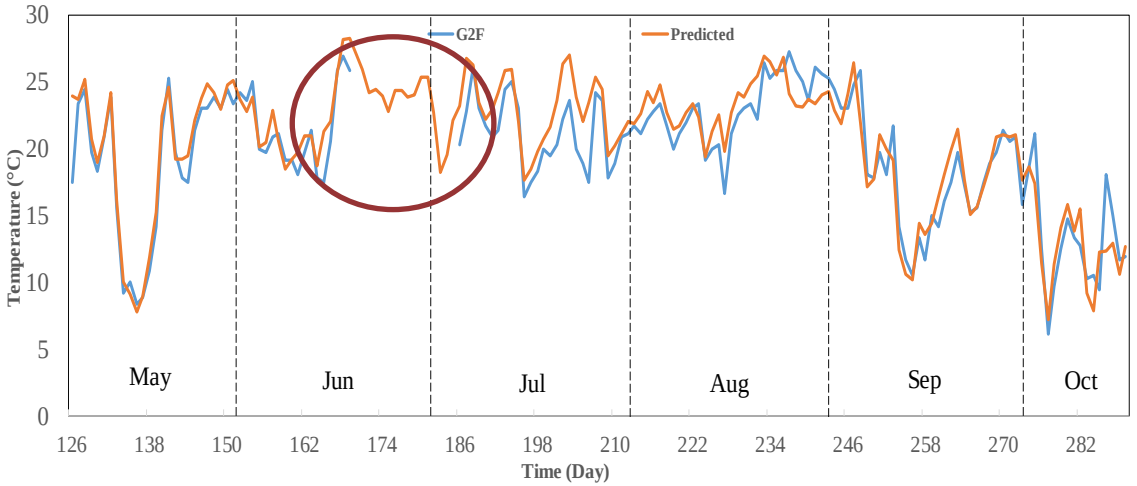




Ref: NSRDB-NREL

Data-driven analytics





Performance Metrics

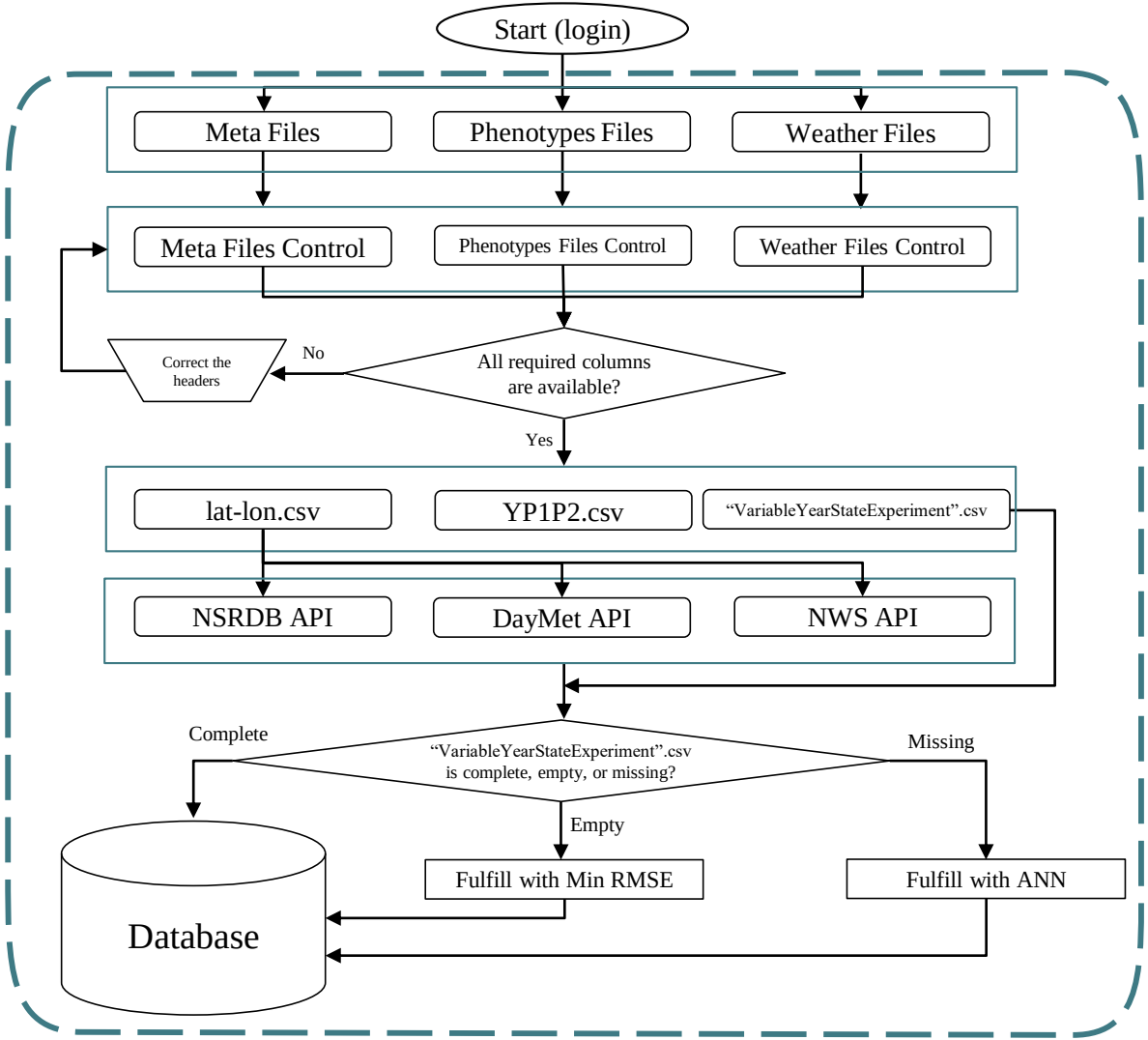
- R^2
- Bias
- RMSE
- NSE

Performance Metric	Mean	Min	Max	SD
R^2	0.88	0.61	0.96	0.10
Bias	-0.52	-1.15	0.13	0.37
RMSE	1.67	1.13	3.00	0.55
NSE	0.87	0.80	0.98	0.05

Pre-Processing

Preprocessing

- Input Files
- Files Control
- Data Processing
- APIs
- Database
- Option Selection
- GxE Model
- GxE Output

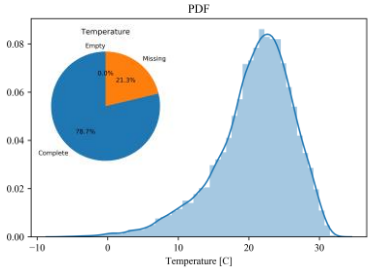


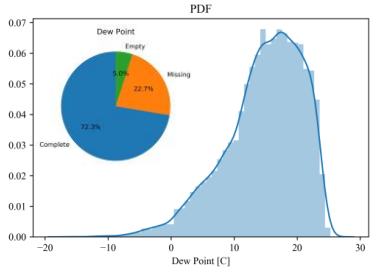
Pre-Processing

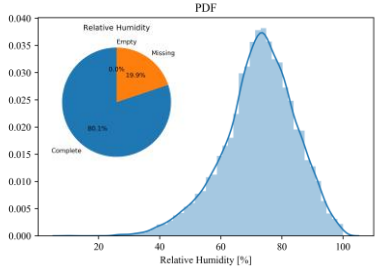
Preprocessing

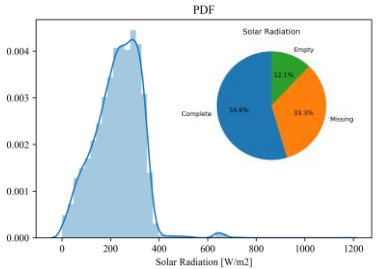
- Input Files
- Files Control
- Data Processing
- APIs
- Database

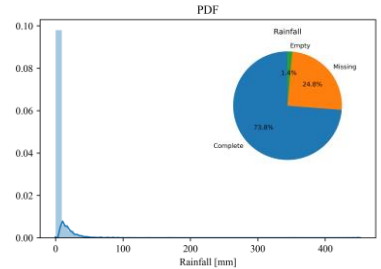
- Option Selection
- GxE Model
- GxE Output

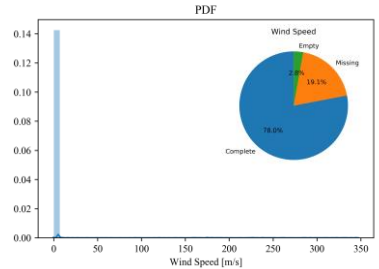


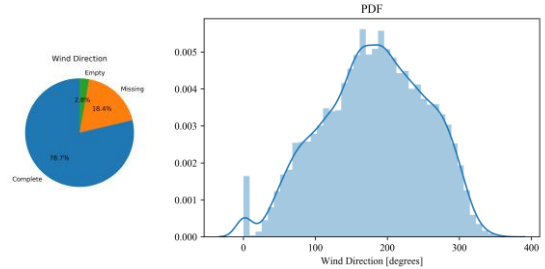


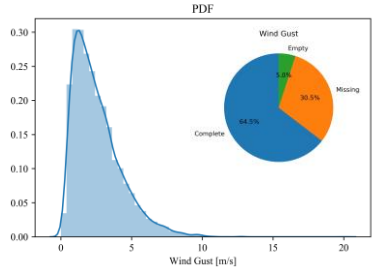












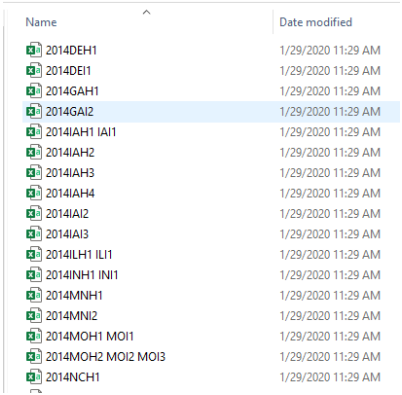
Complete

Empty

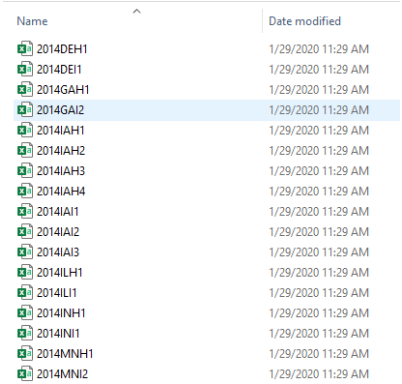
Missing

Pre-processing

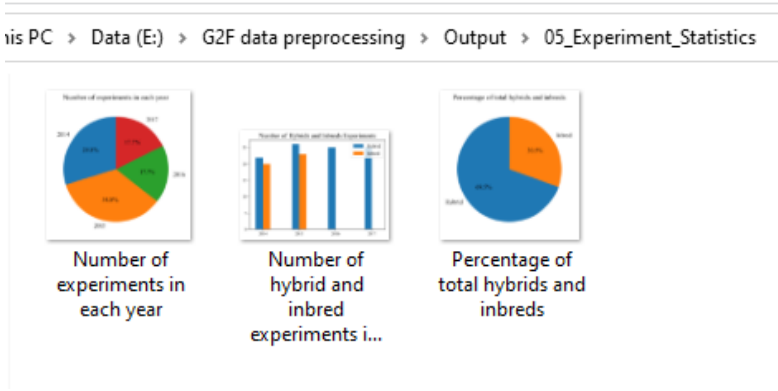
Data separation for each experiment



Correction of the Experiment names and Check the sequence of days



Charts for experiments analysis

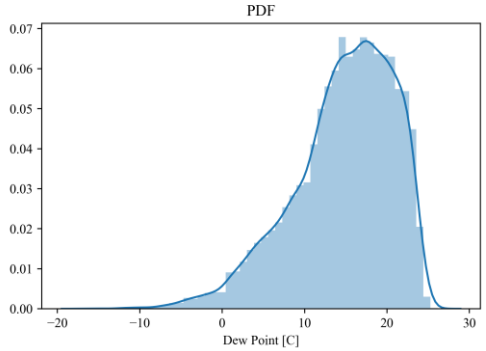


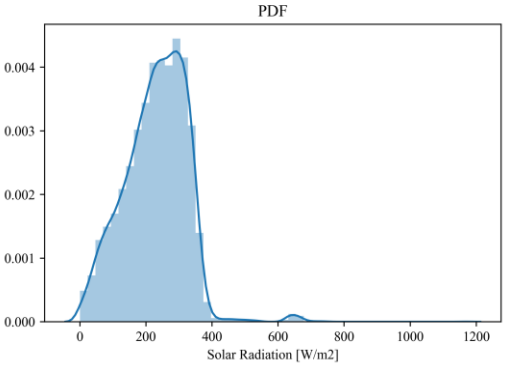
Pre-processing

Separating data for each variable

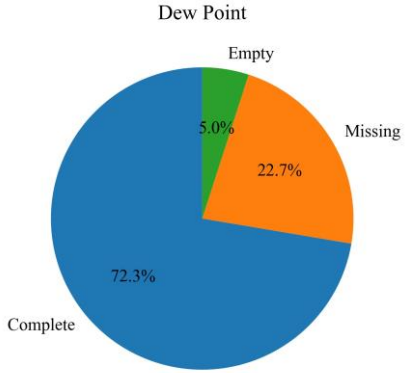
Name	Date modified
D2014DEH1	1/29/2020 11:31 AM
D2014DEI1	1/29/2020 11:31 AM
D2014GAH1	1/29/2020 11:31 AM
D2014GAI2	1/29/2020 11:31 AM
D2014IAH1	1/29/2020 11:31 AM
D2014IAH2	1/29/2020 11:31 AM
D2014IAH3	1/29/2020 11:31 AM
D2014IAH4	1/29/2020 11:31 AM
D2014IAI1	1/29/2020 11:31 AM
D2014IAI2	1/29/2020 11:31 AM
D2014IAI3	1/29/2020 11:31 AM
D2014ILH1	1/29/2020 11:31 AM
D2014ILI1	1/29/2020 11:31 AM
D2014INH1	1/29/2020 11:31 AM
D2014INI1	1/29/2020 11:31 AM
D2014MNH1	1/29/2020 11:31 AM
D2014MNI2	1/29/2020 11:31 AM
D2014MOH1	1/29/2020 11:31 AM
D2014MOH2	1/29/2020 11:31 AM
D2014MOI1	1/29/2020 11:31 AM
D2014MOI2	1/29/2020 11:31 AM
D2014MOI3	1/29/2020 11:31 AM
D2014NCH1	1/29/2020 11:31 AM

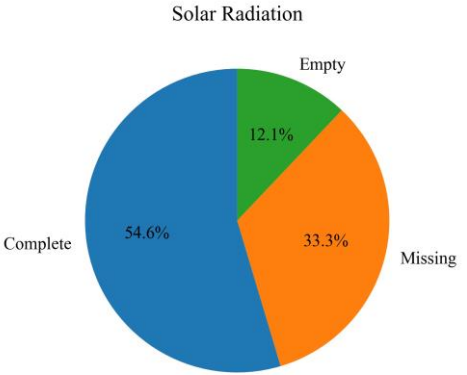
Providing PDFs for each variable





Providing charts to analyze data availability for each variable



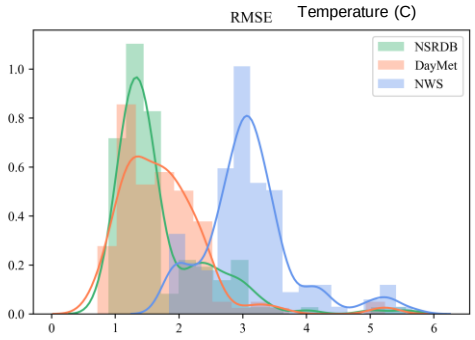


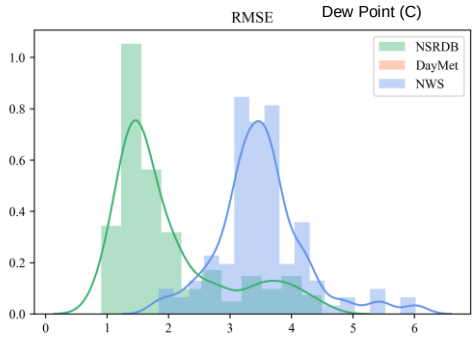
Pre-Processing

Preprocessing

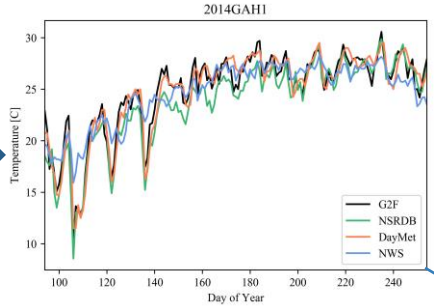
- Input Files
- Files Control
- Data Processing
- APIs
- Database

- Option Selection
- GxE Model
- GxE Output

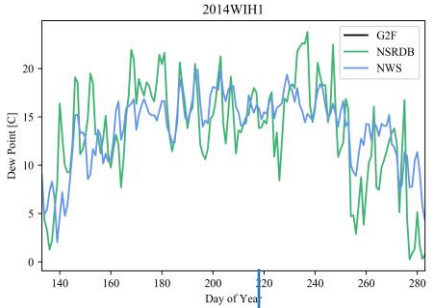




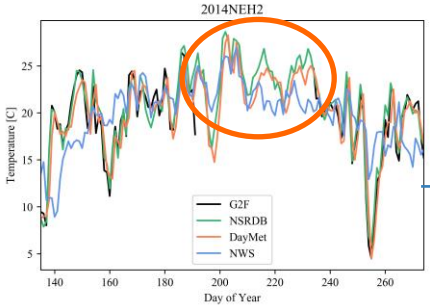
Complete



Empty



Missing

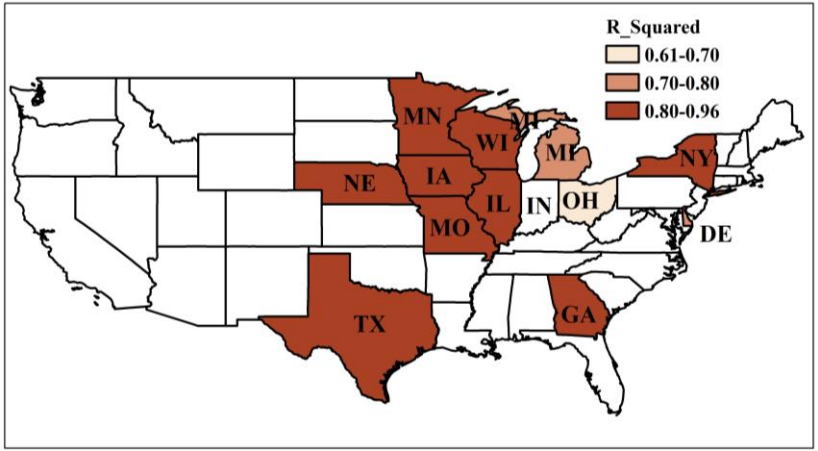


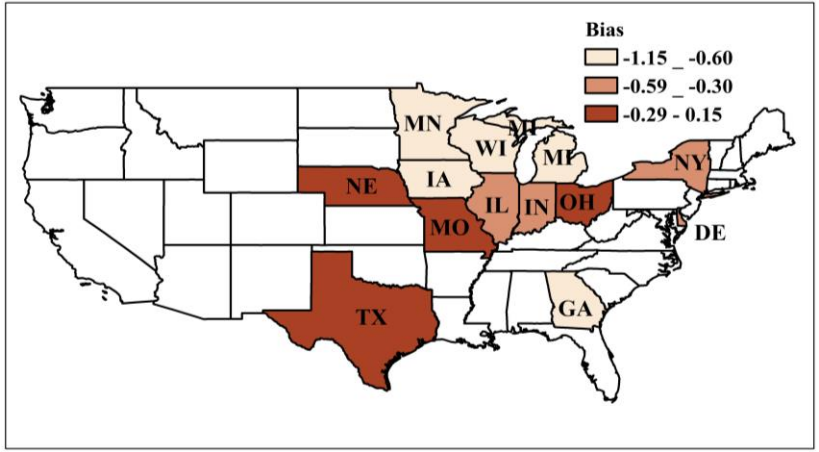
Filling

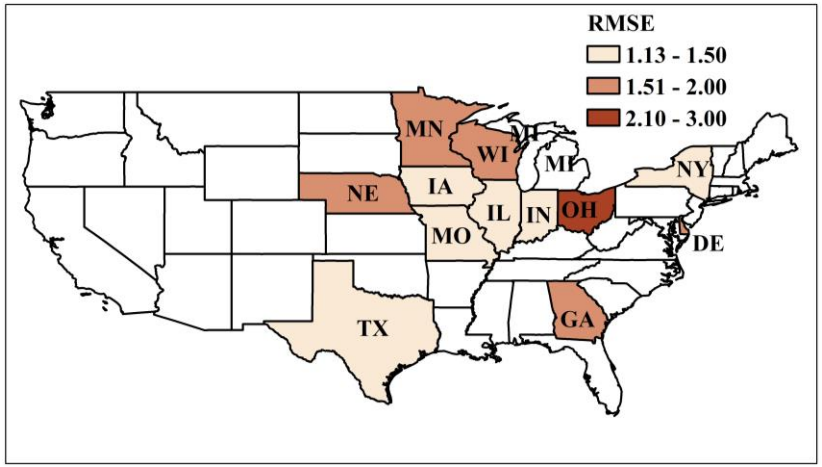
ANN

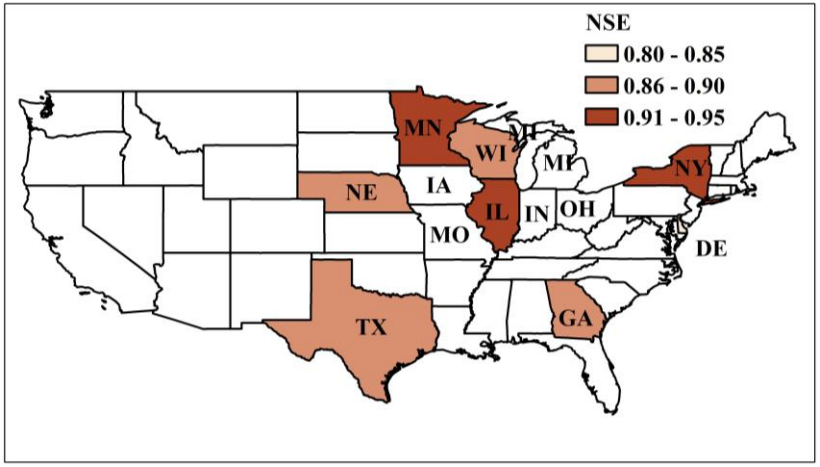
Database

Performance Metrics









Selection

Preprocessing

Option Selection

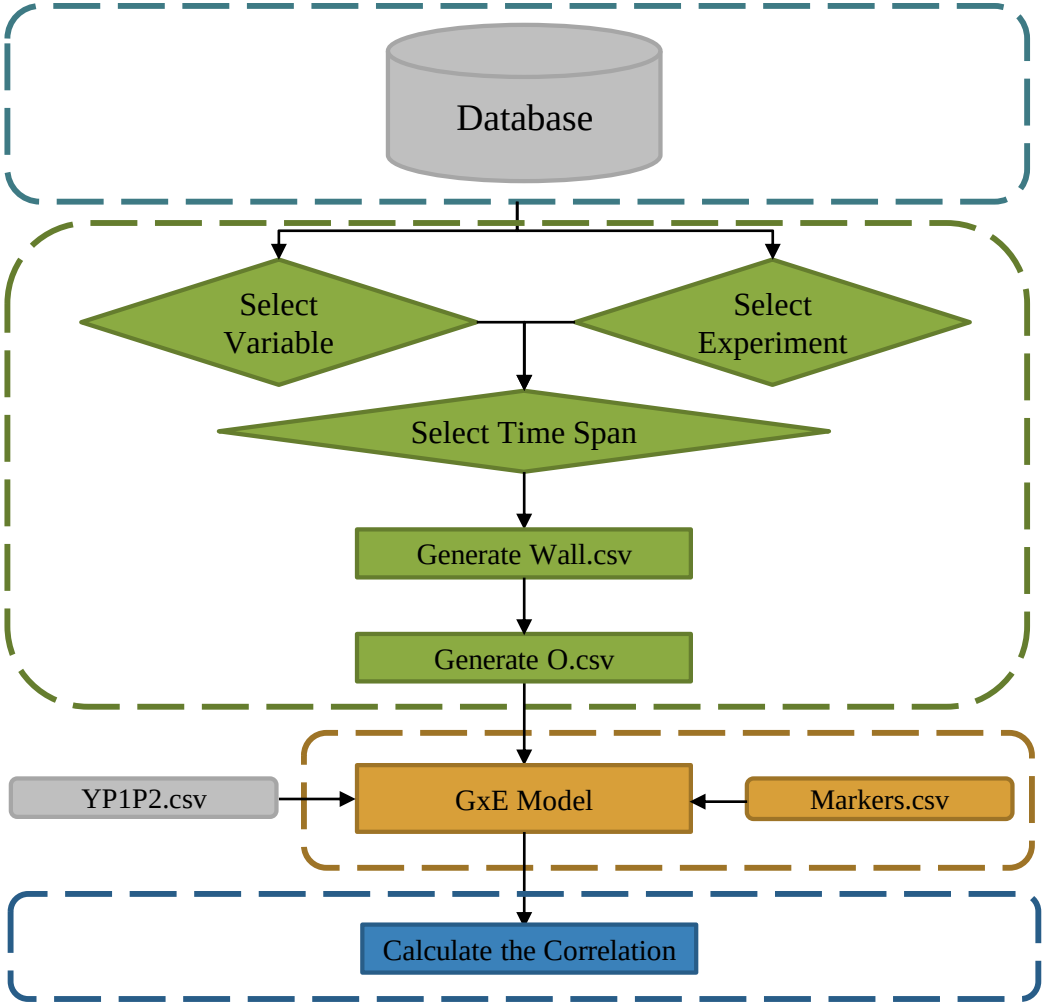
Select Variables

Select Experiment

Select Time Span

GxE Model

GxE Output



Selection

Option Selection

- Preprocessing
- Select Variables
- Select Experiment
- Select Time Span

GxE Execution

- GxE Model
- GxE Output

Select Variable(s)

- [ ] Temperature (C)
- [ ] Dew Point (C)
- [ ] Relative Humidity (%)
- [ ] Solar Radiation (W/m2)
- [ ] Rainfall (mm)
- [ ] Wind Speed (m/s)
- [ ] Wind Direction (degrees)
- [ ] Pressure (mb)
- [ ] Precipitable water (mm)

Select Experiment(s)

- [ ] 2014IAH3
- [ ] 2015NEH3
- [ ] 2017MOH1

.
.
.
.
.

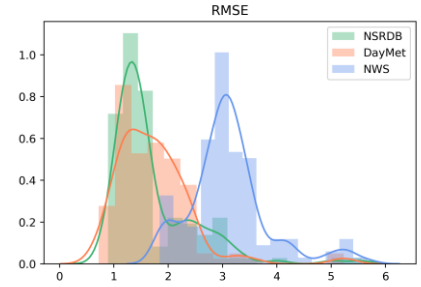
Start day =

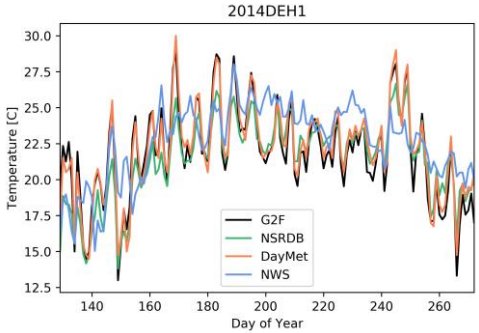
End day =

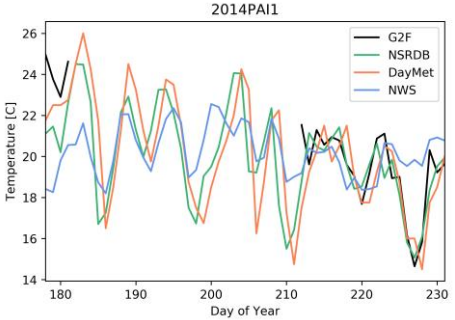
		Tested Genotypes: YES	Tested Genotypes: NO
Tested Environments	YES	CV2	CV1
Tested Environments	NO	CV0	CV00

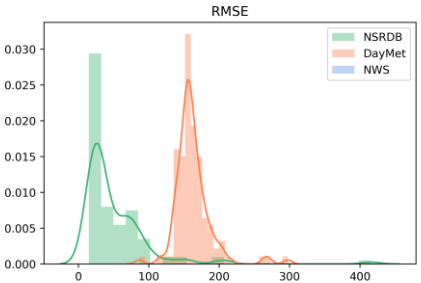
CV00: Predicting performance of unobserved lines in unobserved environments;
CV0: Predicting performance of unobserved environments;
CV1: Predicting performance of new developed lines through relationships with others;
CV2: Predicting Performance of Lines Captured in Other Environments

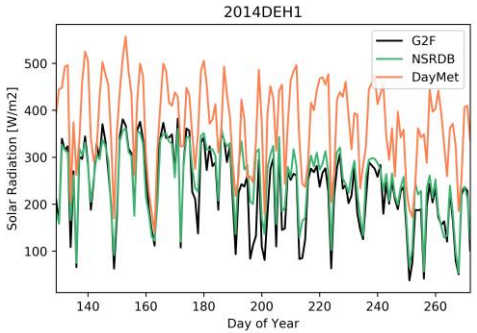
Selection

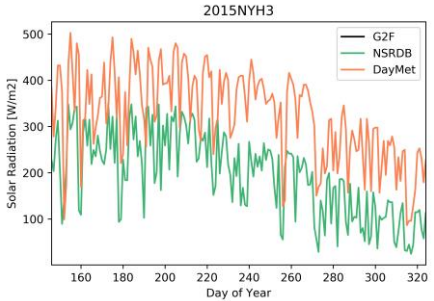

RMSE
NSRDB
DayMet
NWS


2014DEH1
Temperature [C]
Day of Year
G2F
NSRDB
DayMet
NWS


2014PAI1
Temperature [C]
Day of Year
G2F
NSRDB
DayMet
NWS


RMSE
NSRDB
DayMet
NWS

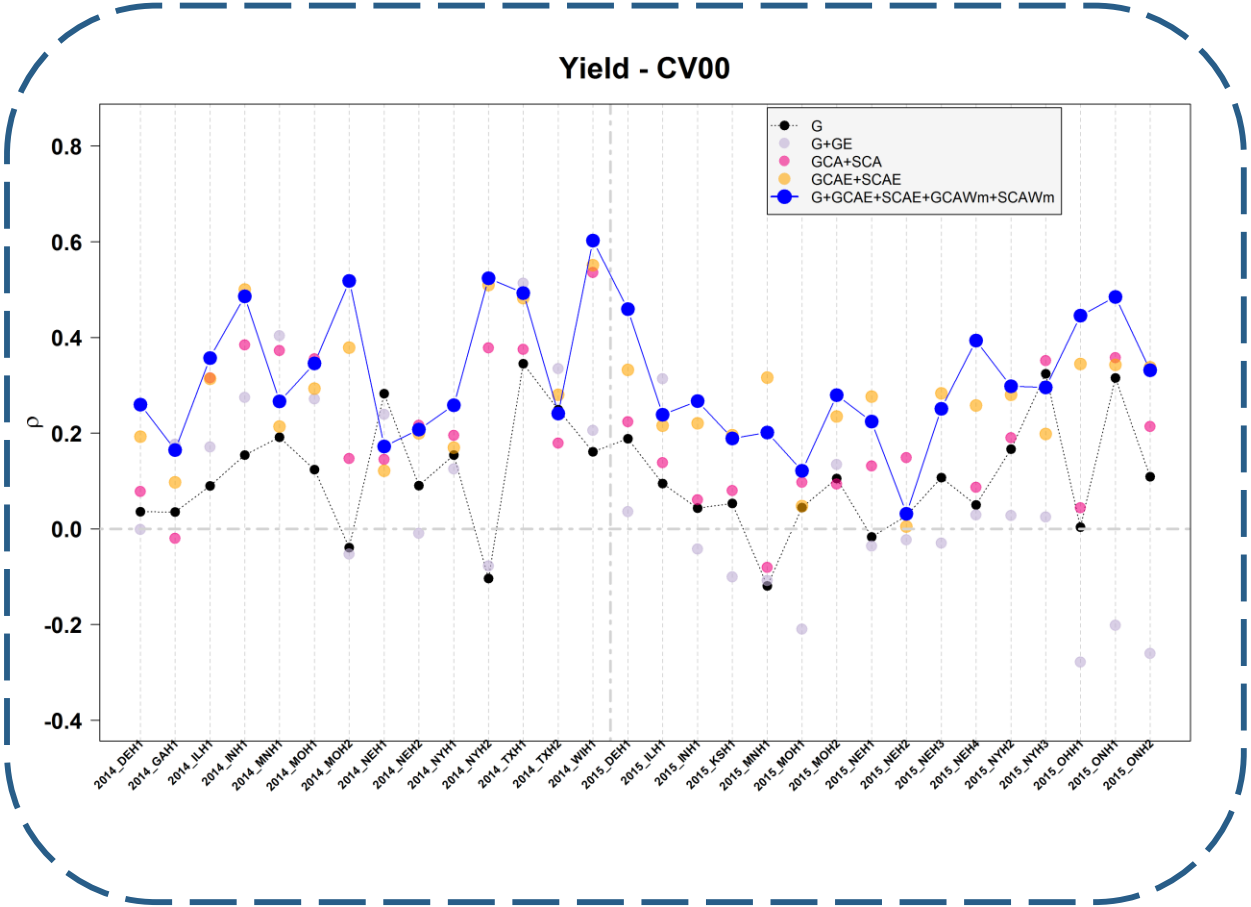

2014DEH1
Solar Radiation [W/m2]
Day of Year
G2F
NSRDB
DayMet


2015NYH3
Solar Radiation [W/m2]
Day of Year
G2F
NSRDB
DayMet

Post-Processing

- Preprocessing
- Select Variables
- Select Experiment
- Select Time Span
- GxE Model
- GxE Output

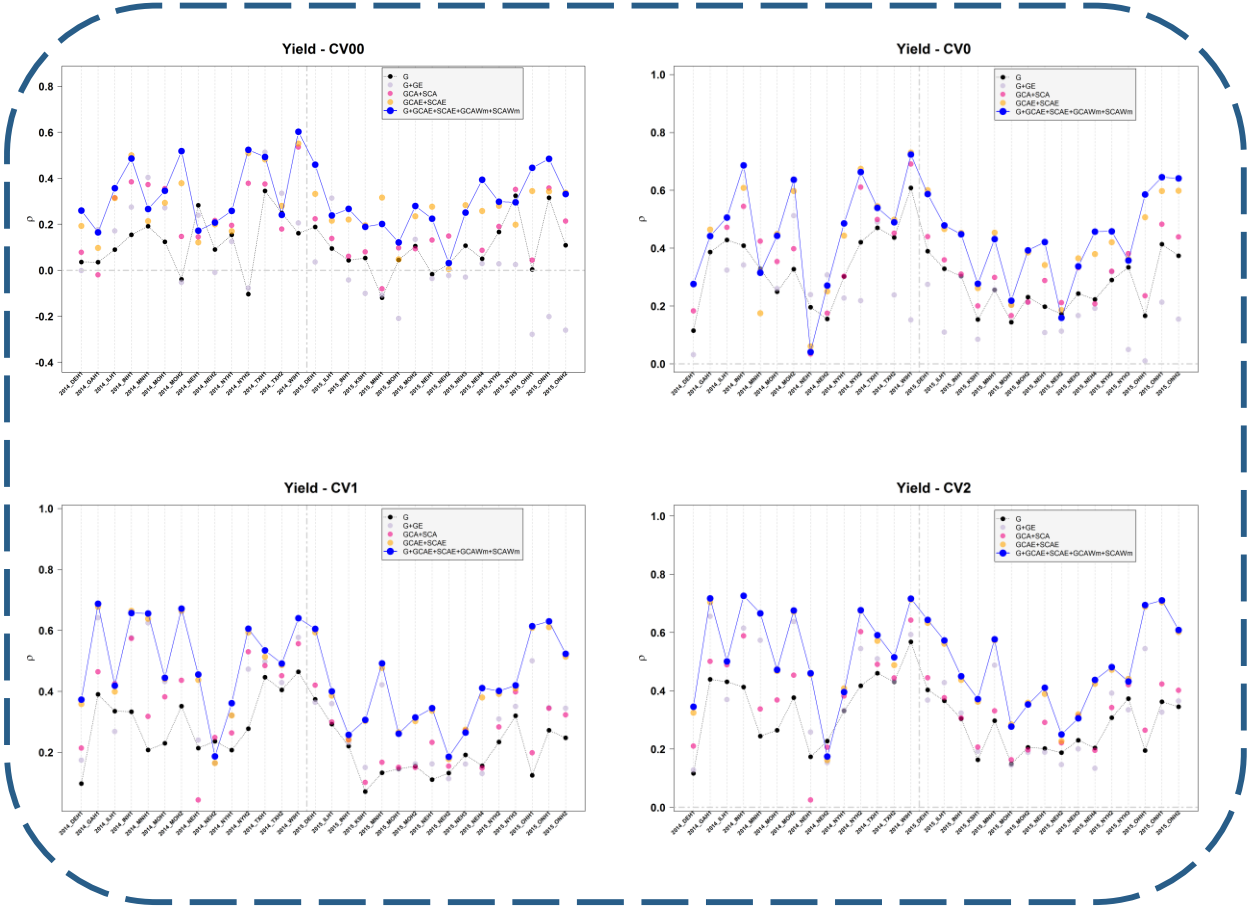
GxE Predictability

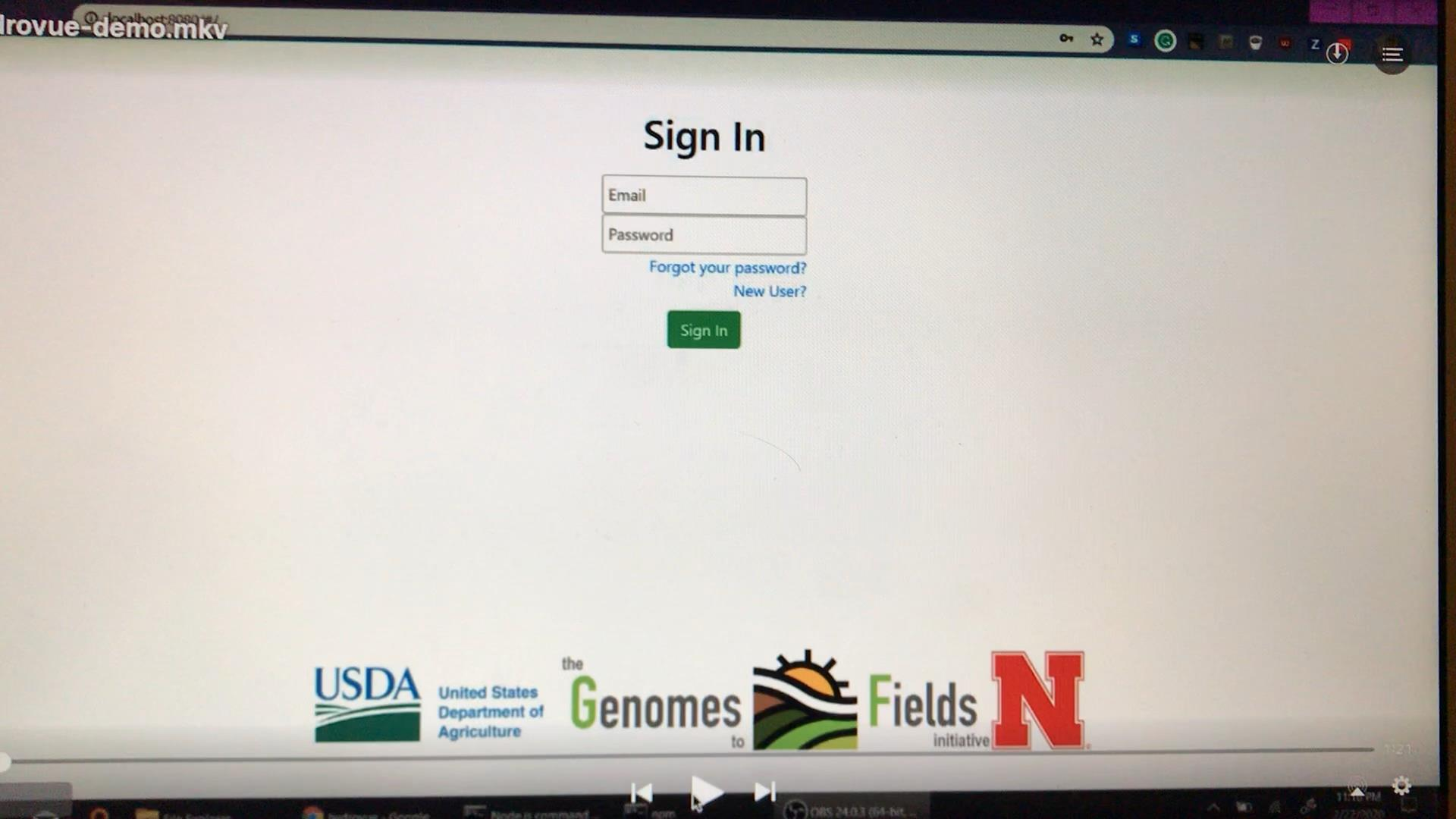


Post-Processing

- Preprocessing
- Select Variables
- Select Experiment
- Select Time Span
- GxE Model
- GxE Output

GxE Predictability




rovue-demo.mkv
Sign In
Email
Password
Forgot your password?
New User?
Sign In
USDA
United States
Department of
Agriculture
the
Genomes
to
Fields
initiative

Complexities

Providing AWS (Amazon Web Service) as platform for the phenotypic predictability application;

Authentication for different users;

Transferring all the data (G2F, NSRDB, DayMet, and NWS) and scripts (R and Python) to the platform;

Coupling R and Python scripts to develop an integrated software for phenotype-prediction in the G2F experiment.

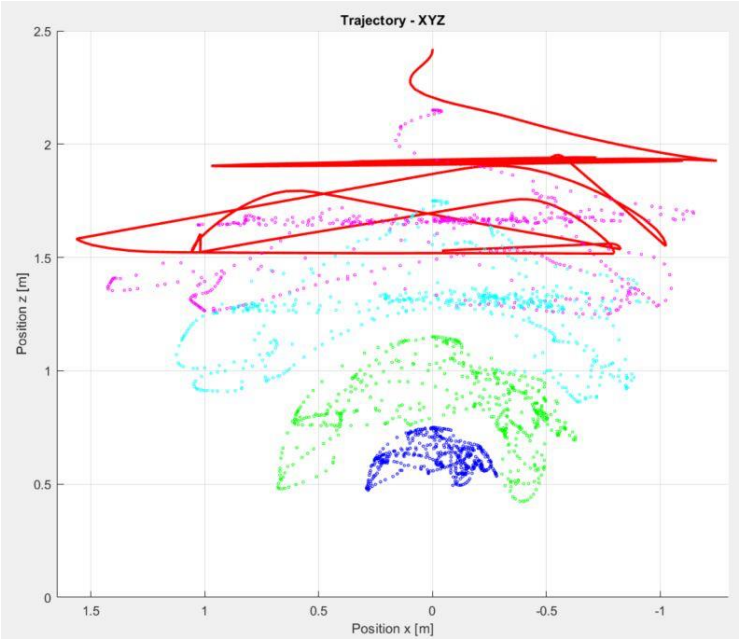
Conclusions

- The integration of other data sources to improve G2F database unclearly improved the predictability of phenotypes;
- Transferring and coupling the hydroclimate data analytics and GxE modeling scripts to the web service platform is feasible;
- Increasing the number of experiments may lead to a better accuracy of phenotype predictability.

Future work

- Add climatic spatial and temporal analytics of GxE predictability module;
- Add a global sensitivity of GxE accuracy module to estimate sources and propagation of uncertainty in response to various climatic (environmental) factors;
- Add the remote sensing data plugin module to increase the number of climatic variables and phenotypes in the database.

Some more future work



Team members and tasks:

- **Francisco Munoz-Arriola**; Team leader
- **Diego Jarquin: GxE model developer**; Develops R scripts for phenotypes predictions using GxE
- **Hallie Hohbein: Project Manager**; Takes care of project management tasks, documentation, and testing
- **Parisa Sarzaeim: Hydroclimate data scientist**; Develops Python scripts to manage hydroclimate database
- **Joseph Carter: Frontend/Backend Developer**; Works on user authentication, frontend development, and testing.
- **David Recic: Backed Developer**; Creates the database and works on user authentication.
- **Zoe Trautman: Frontend Developer**; Develops the frontend and writes documentation.
- **Anna Zhang: Development Manager**; In charge of AWS and helps with backend development.
- **Byrav Ramamurthy and Francisco Munoz-Arriola;** Computer science advisers

Thank You

This project was supported by the Agriculture and Food Research Initiative Grant number NEB-21-176 and NEB-21-166 from the **USDA National Institute of Food and Agriculture, Plant Health and Production and Plant Products: Plant Breeding for Agricultural Production, A1211). Accession Nos.1015252 and No.1009760**