

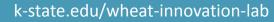


### LEVERAGING 'BIG DATA' FOR WHEAT IMPROVEMENT IN KANSAS AND AROUND THE WORLD

### JESSE POLAND

Kansas State University

Feed the Future Innovation Lab for Applied Wheat Genomics



Cornell University









## 60%

Increase in demand for wheat by 2050

- 20%

Potential average yield loss from climate change

>2%

Rate of gain needed to meet projections

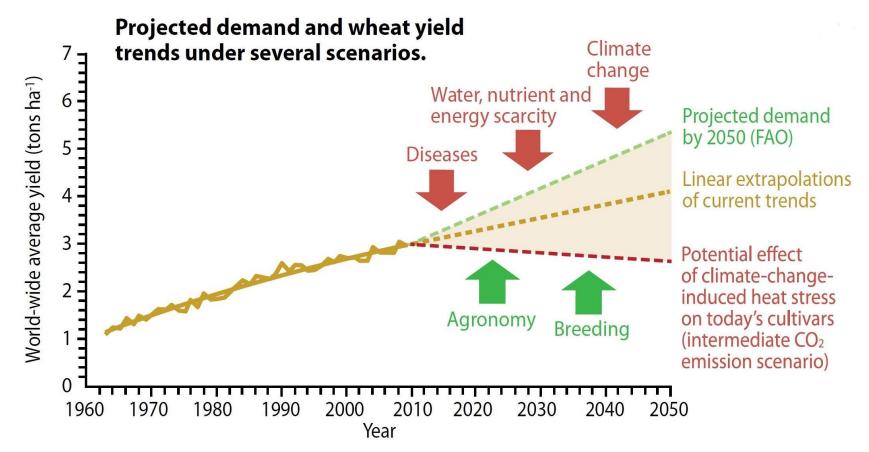
~ 1%

Current rate of gain











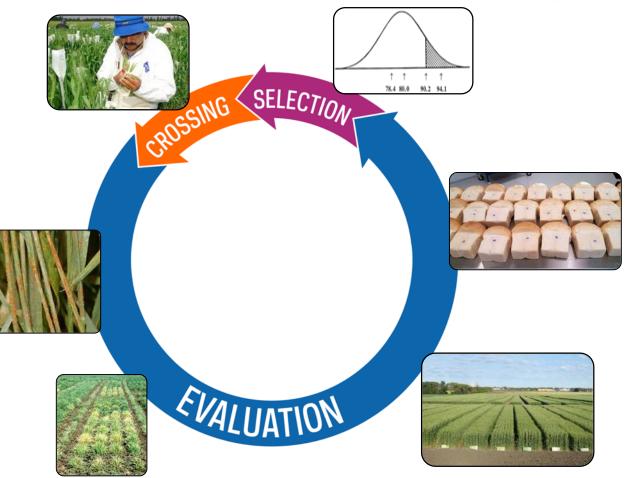




### THE BREEDING CYCLE

### genetic gain

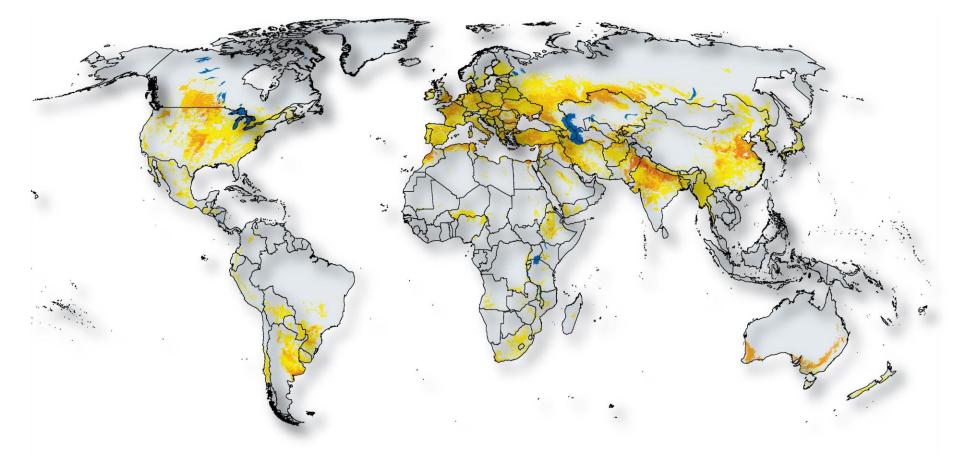
year on year progress in overall performance of breeding lines











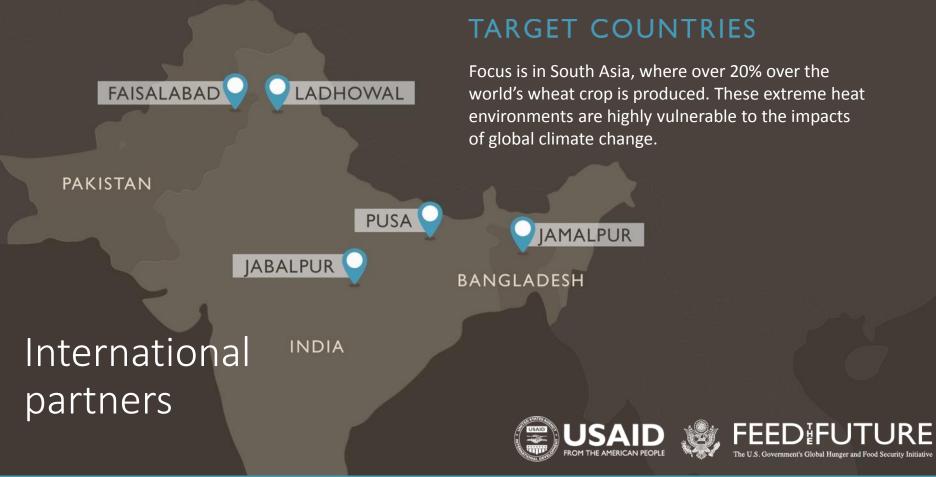






@wheatlab

y





La Habana

New York

República

d States

merica

udad **México** 



## **PREDICTION MODELS**

- billions of data points
  - wind speeds
- sea temperatures
- mathematical equations of physical process
- algorithms modeling the atmosphere at every point on the globe

### super computer to simulate possible outcomes









## PREDICTION MODELS

- billions of data points -
- yield and baking measurements -
- genomic information / sequence variants
  - mathematical equations of genetics -
- algorithms to calculate all of genetic effects -



processed on super computer to solve models







### ACCELERATING † † † † THE BREEDING SELECTION CROSSING 78.4 80.0 90.2 94.1 CYCLE GENOMIC PREDICTION EVALUATION







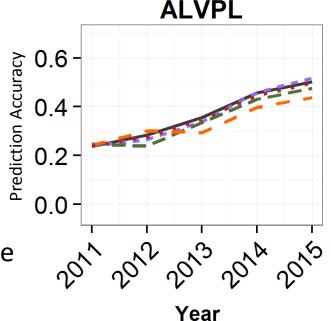
#### BIG DATA data sets that are so large or complex that traditional data processing application software is inadequate to deal with them

data driven modeling approaches .... 'the more data, the better'

The **GOOD**... we can predict about anything

The **BAD**... big challenges in modeling

The UGLY... big data is messy and hard to manage



ELNET --- GAUSS - - PLSR - - RF ··· RRBLUP

@wheatlab

Battenfield, S. et al. (2016) The Plant Genome. DOI: 10.3835/plantgenome2016.01.0005

www.en.wikipedia.org/wiki/Big\_data

Method





y





# FEEDIFUTURE

The U.S. Government's Global Hunger & Food Security Initiative

#### www.feedthefuture.gov













