A Case Study in Prince Edward Island, Canada: Comparing A Process Based Model and Machine Learning for Potato Yield Prediction

CGU 2023 - Banff, Alberta
May 8th, 2023

Mariaelisa Polsinelli
McGill University, Agriculture and Agri-Food Canada
Morteza Mesbah
Agriculture and Agri-Food Canada
Zhiming Qi
McGill University
Matt Ramsay
Kensington North Watersheds Association and Oyster Cove Farms

mariaelisa.polsinelli@mail.mcgill.ca

McGill UNIVERSITY

Agriculture and Agri-Food Canada   Agriculture et Agroalimentaire Canada

Image source: https://atlantic.ctvnews.ca/prince-edward-island-farmers-destroy-136-million-kilos-of-spuds-1.5799465

Potato Production in Prince Edward Island

- 5th most produced crop globally (FAO, 2021)
- 24.2% of land area (Maqsood, 2020)
- 84% of total exported agrifood products (Maqsood, 2020)
- 10.8% of the GDP (Maqsood, 2020)

Image source: https://www.mayrand.ca/en/our-products/fruits-et-legumes/legumes/pomme-de-terre/patate-russet-chef--26771/

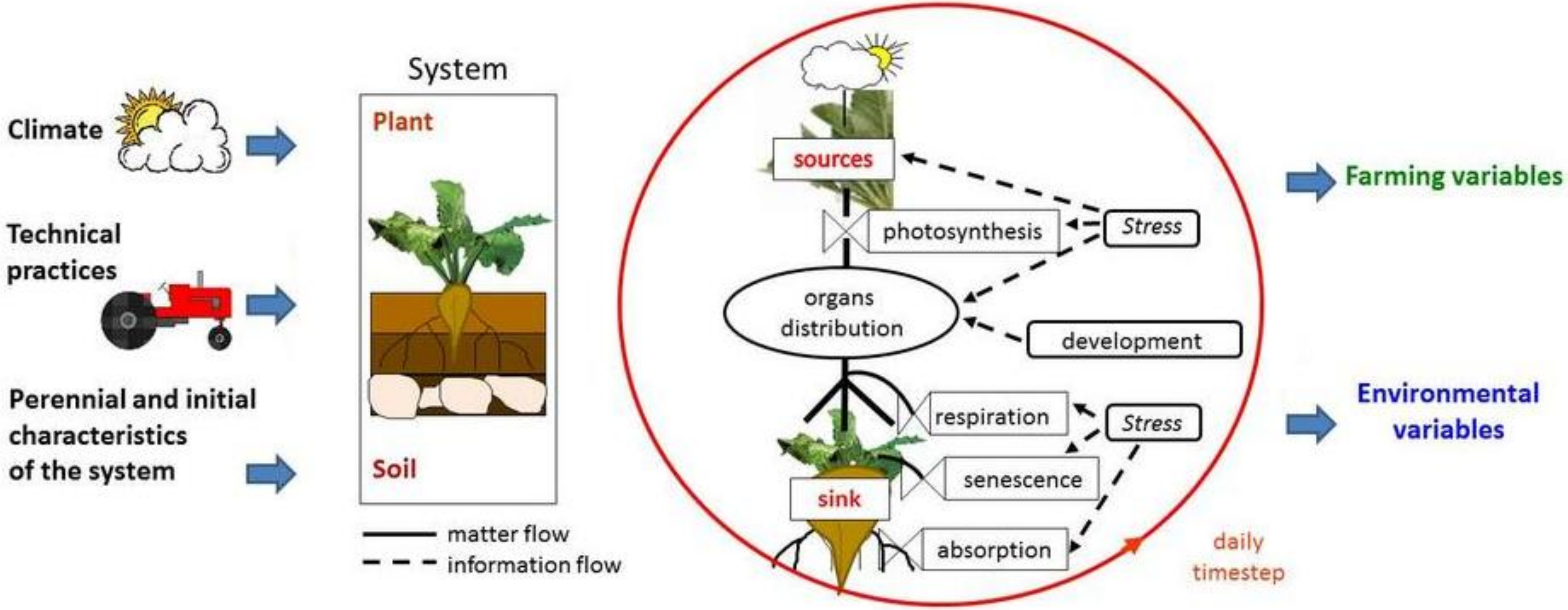
Yield Prediction

- Crop management decisions
  (Morissette et al., 2016, Shahhosseini et al., 2019)
- Economic trading
  (Jeong et al., 2016, Shahhosseini et al., 2019)
- Agricultural policy development
  (Jeong et al., 2016)
- Food Security
  (Shahhosseini et al., 2019)
- Adaptation strategies with climate change
  (Jeong et al., 2016)
  - Increase 1–1.4°C can reduce potato yields by 18–32% (Maqsood, 2020)

STICS

Simulateur mulTIdisciplinaire pour les Cultures Standard / Multidisciplinary Simulator for Standard Crops

Inputs **System - process** **Outputs**



Random Forest (RF)

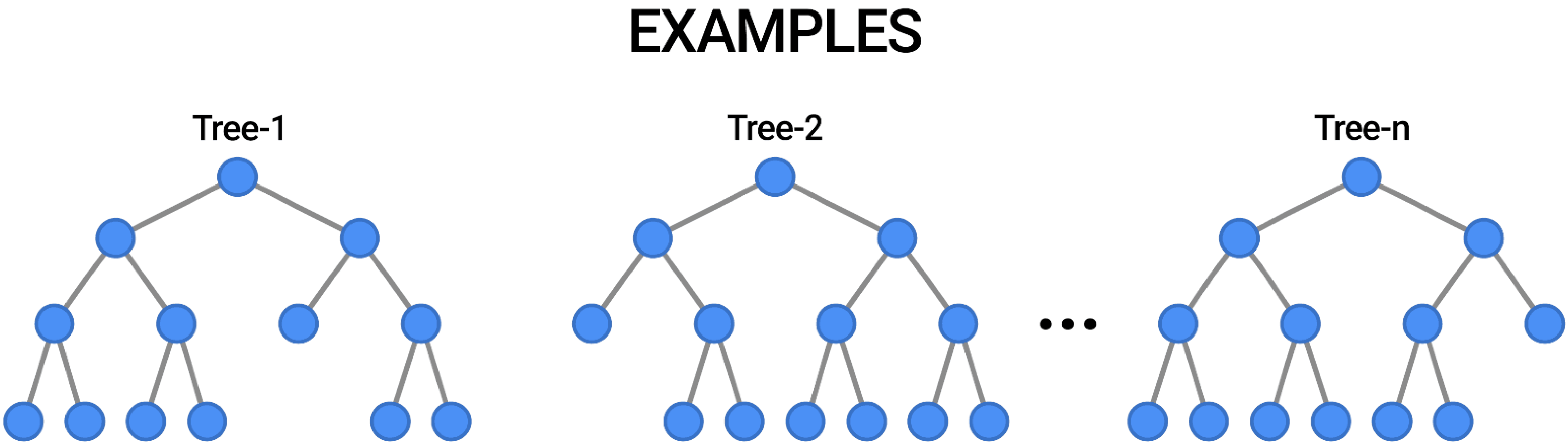


Image Source: https://blog.tensorflow.org/2021/05/introducing-tensorflow-decision-forests.html

Research Questions

- Can a crop model calibrated on research farm data be used for industrial fields?
- Can Random Forest predict the yield of unseen fields/years?

Industrial Farm Data

Provided by Matt Ramsay
Oyster Cove Farms, Prince County, PEI
46°29’N, 63°42’ W
2015-2021, excluding 2018

STICS Calibrated Model

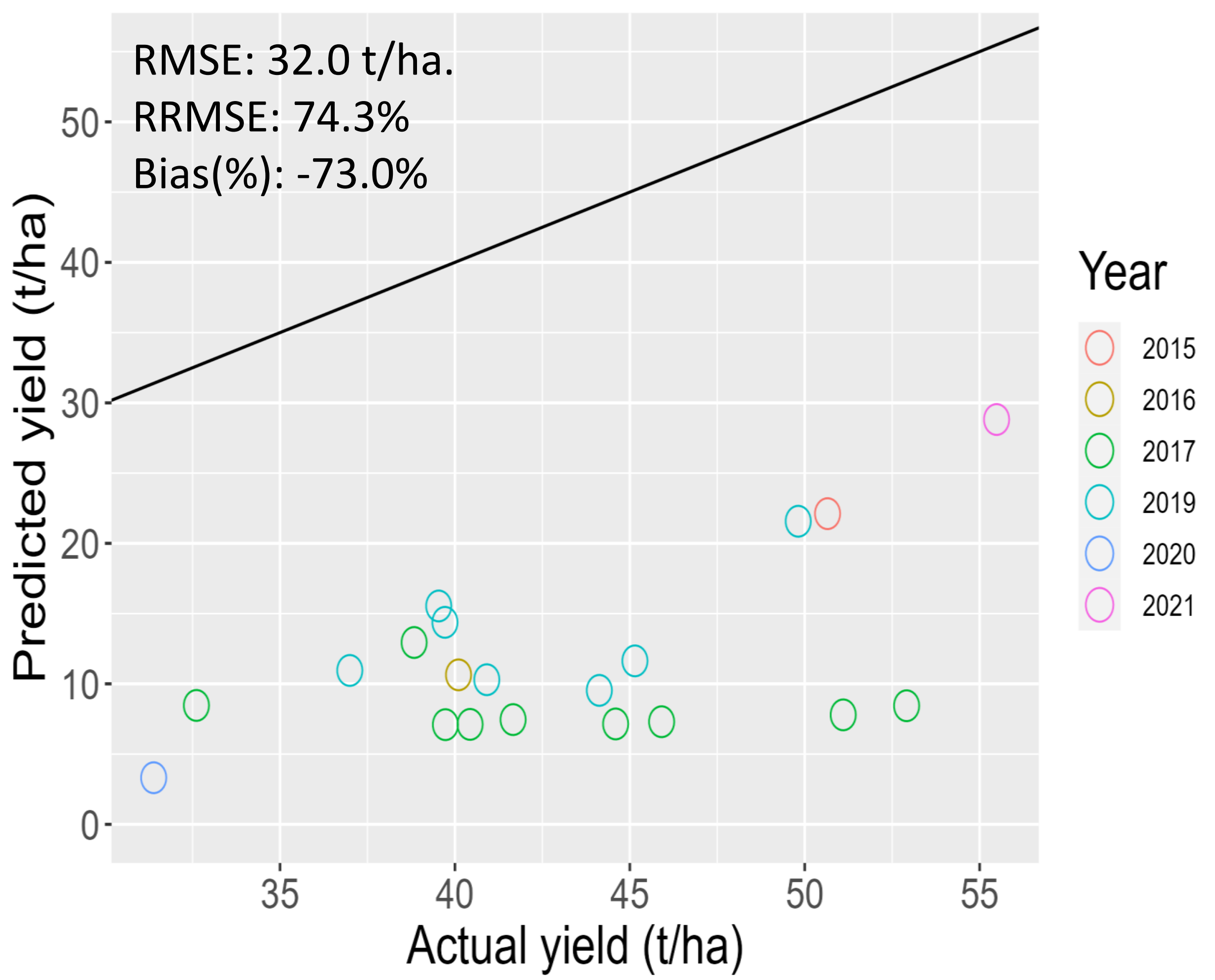
- Calibrated and evaluated on research fields
- Russet Burbank
- Ste. Foy, Quebec and Fredericton, New Brunswick
- 2012-2013
- Crop Coefficient

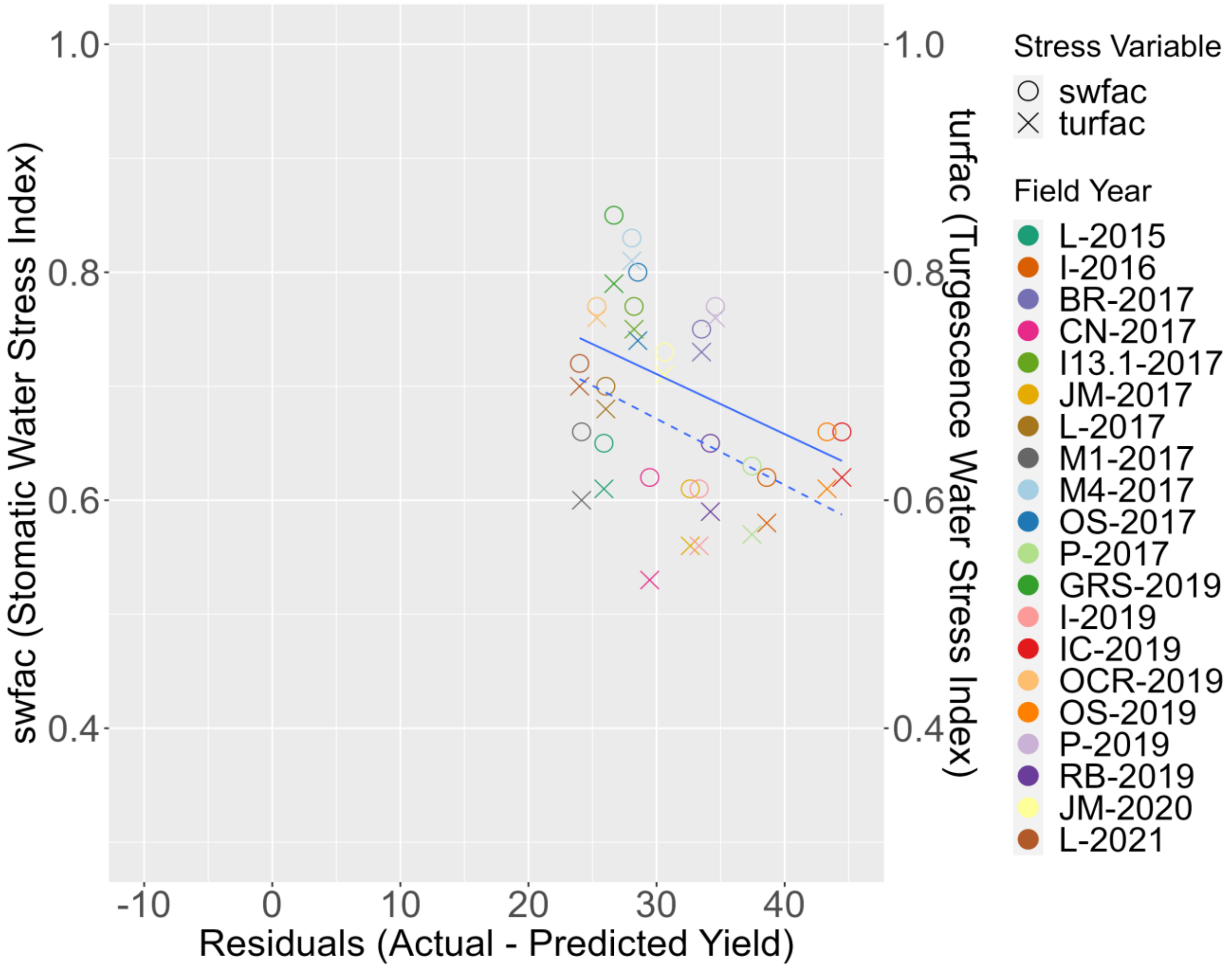
Two Options for Water Balance Calculation:

1. Crop Coefficient
   - Simplified
   - Reference crop evapotranspiration x Crop Coefficient

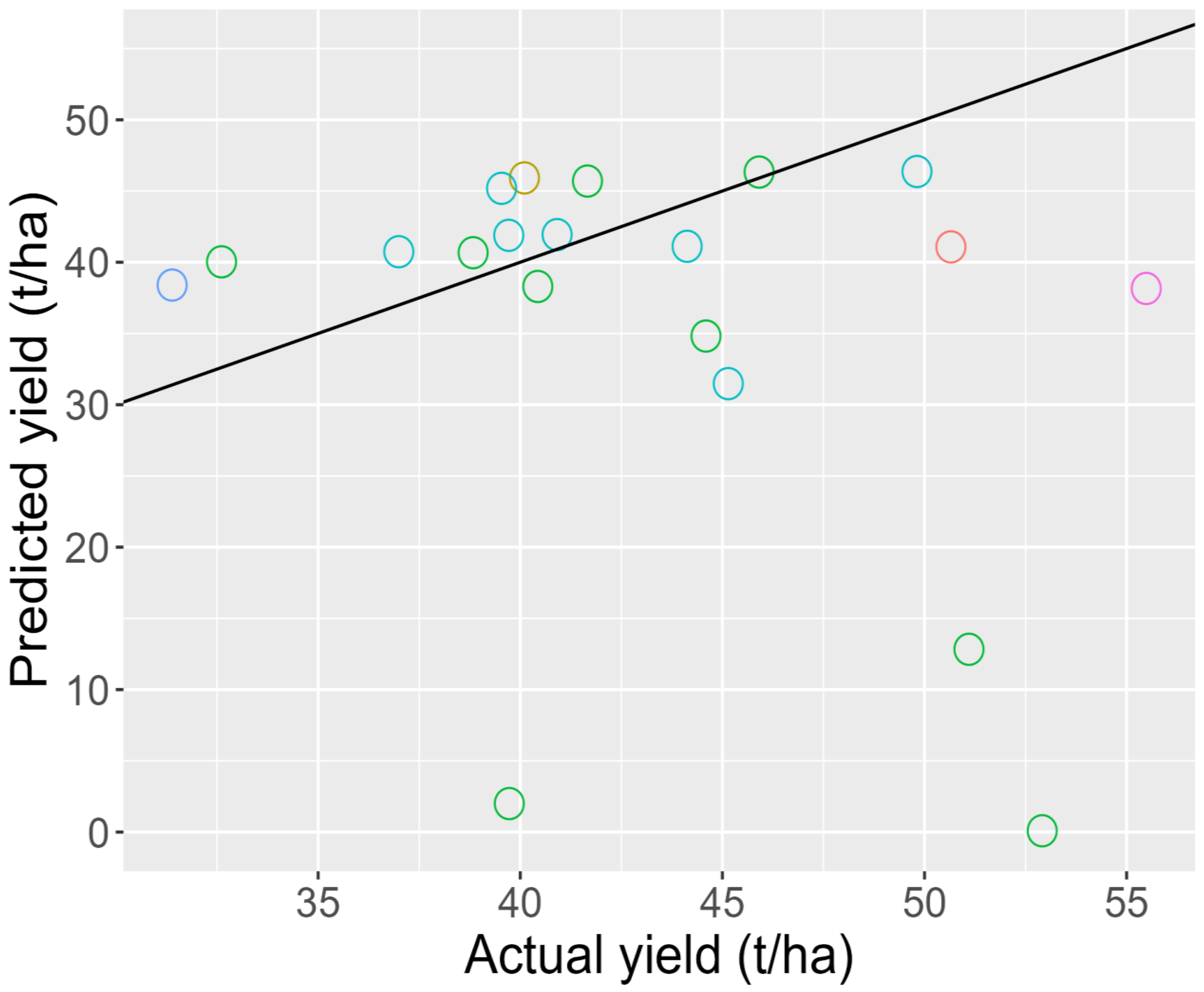
2. Resistance Approach
   - More sophisticated
   - Soil evaporation and crop water requirement calculated separately

STICS with Crop Coefficient on Industrial Farm Data

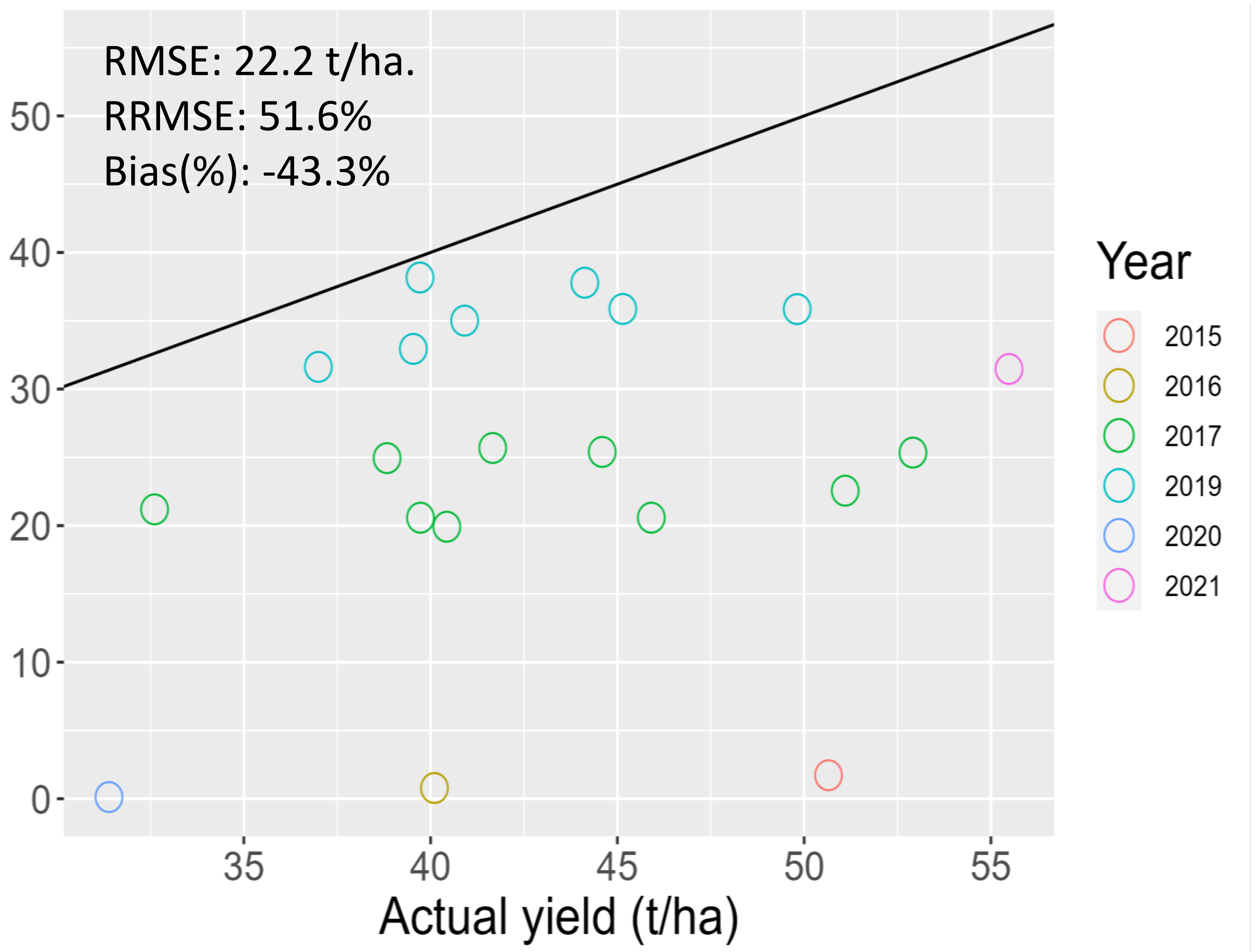




STICS with No Water Stress, Crop Coefficient



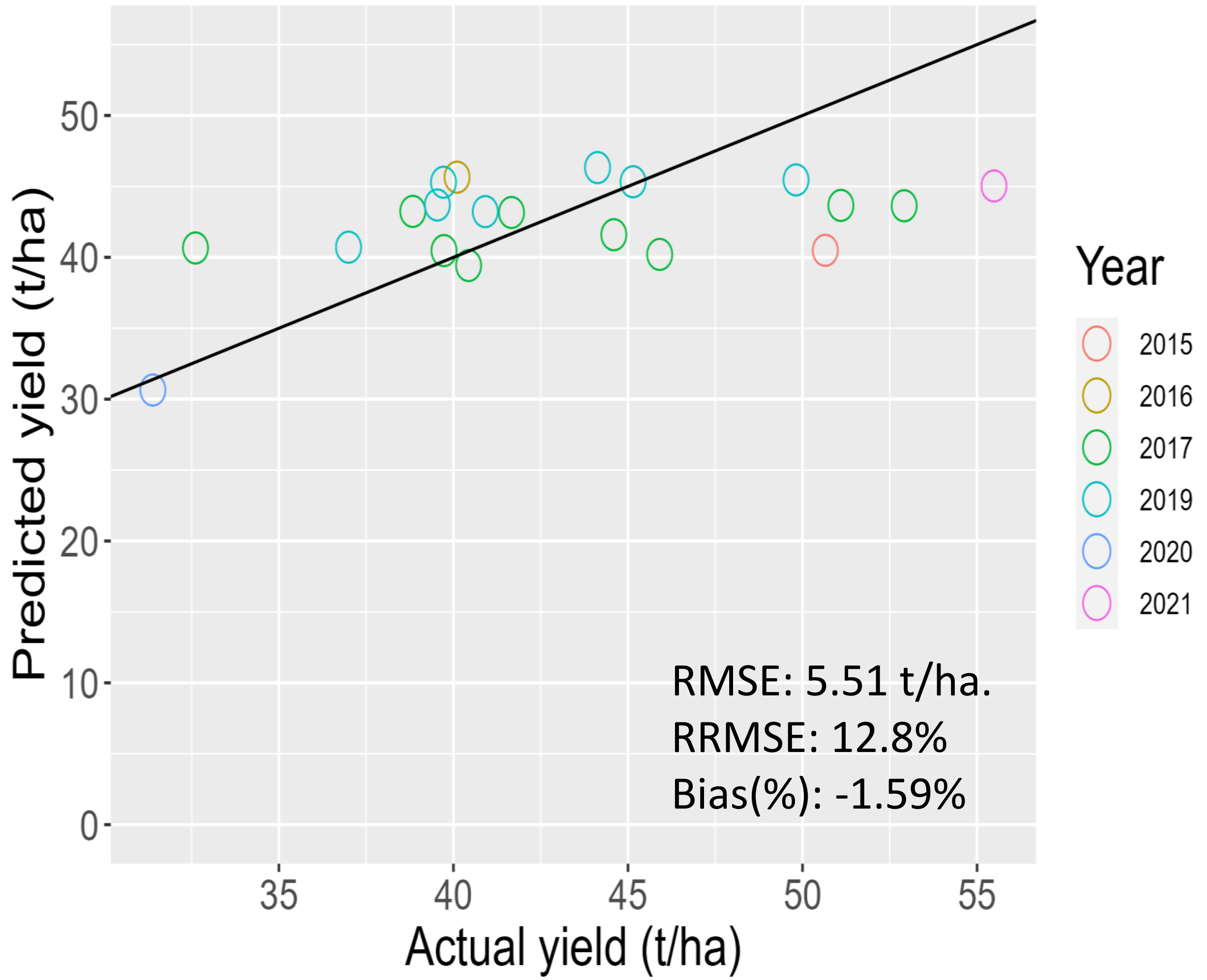
STICS with Uncalibrated Resistance Approach

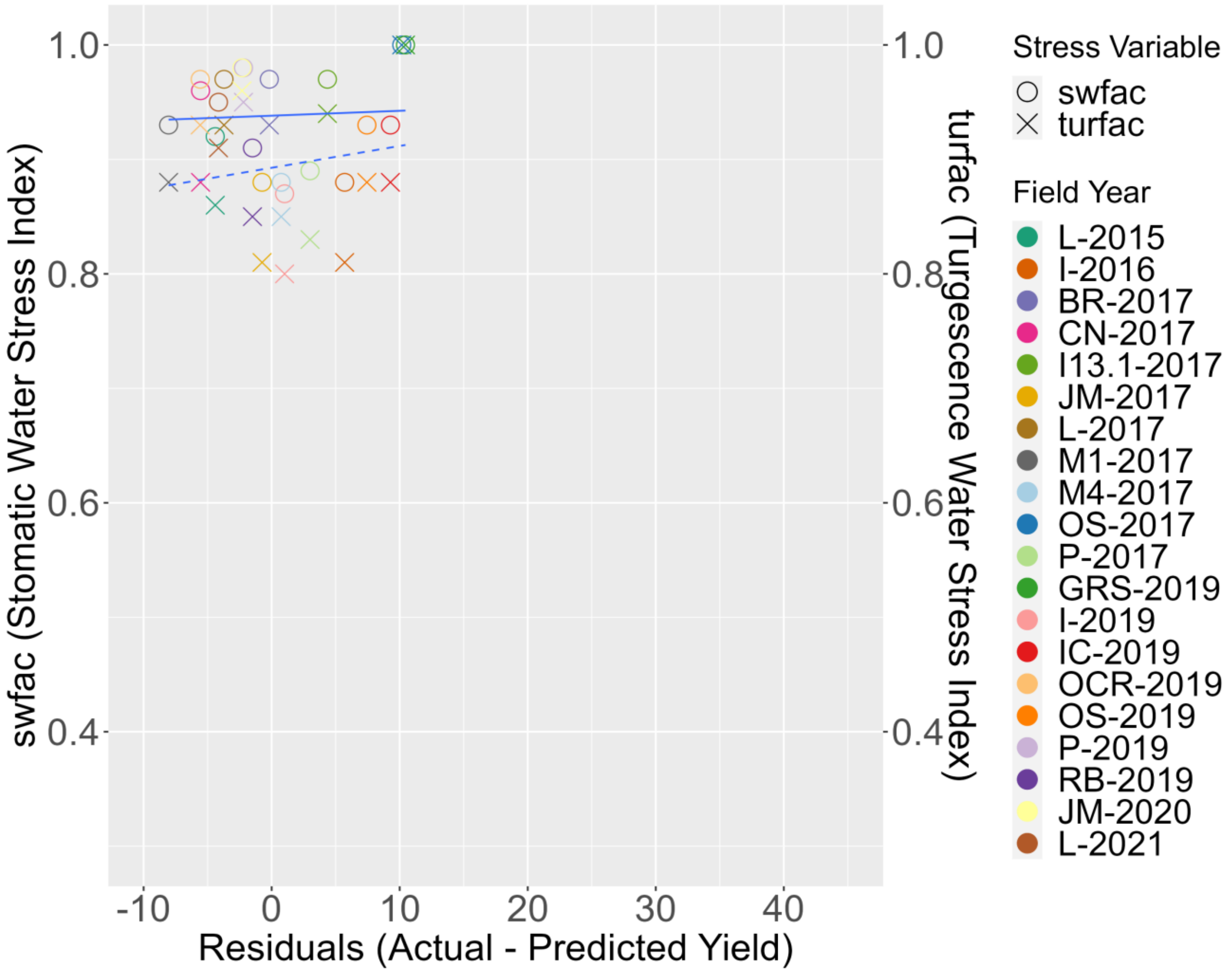


STICS Water Stress Parameters

Parameter	Description	Unit	Default	Adjusted
aclim	Climatic component for calculation of actual soil evaporation	mm	20	4
q0	cumulative soil evaporation above which evaporation rate is decreased	mm	4	3
rsmin	Minimal stomatal resistance of leaves	s/m	10	200

STICS with Calibrated Resistance Approach





Research Questions

- Can a crop model calibrated on research farm data be used for industrial fields?
- Can Random Forest predict the yield of unseen fields/years?

Random Forest

Top 10 out of 43 Total Features:

1. Cumulative GDD to Harvest Date
2. Cumulative Growing Season GDD (May to Harvest Date)
3. Total September Precipitation (mm)
4. Julian Harvest Date
5. Total Growing Season Radiation
6. DEM Slope
7. Julian Sowing Date
8. % Clay (Top 20 cm)
9. Soil pH (Top 20 cm)
10. % Organic Matter (Top 20 cm)

Train/Test Split:

Test on specific field year, train on the rest

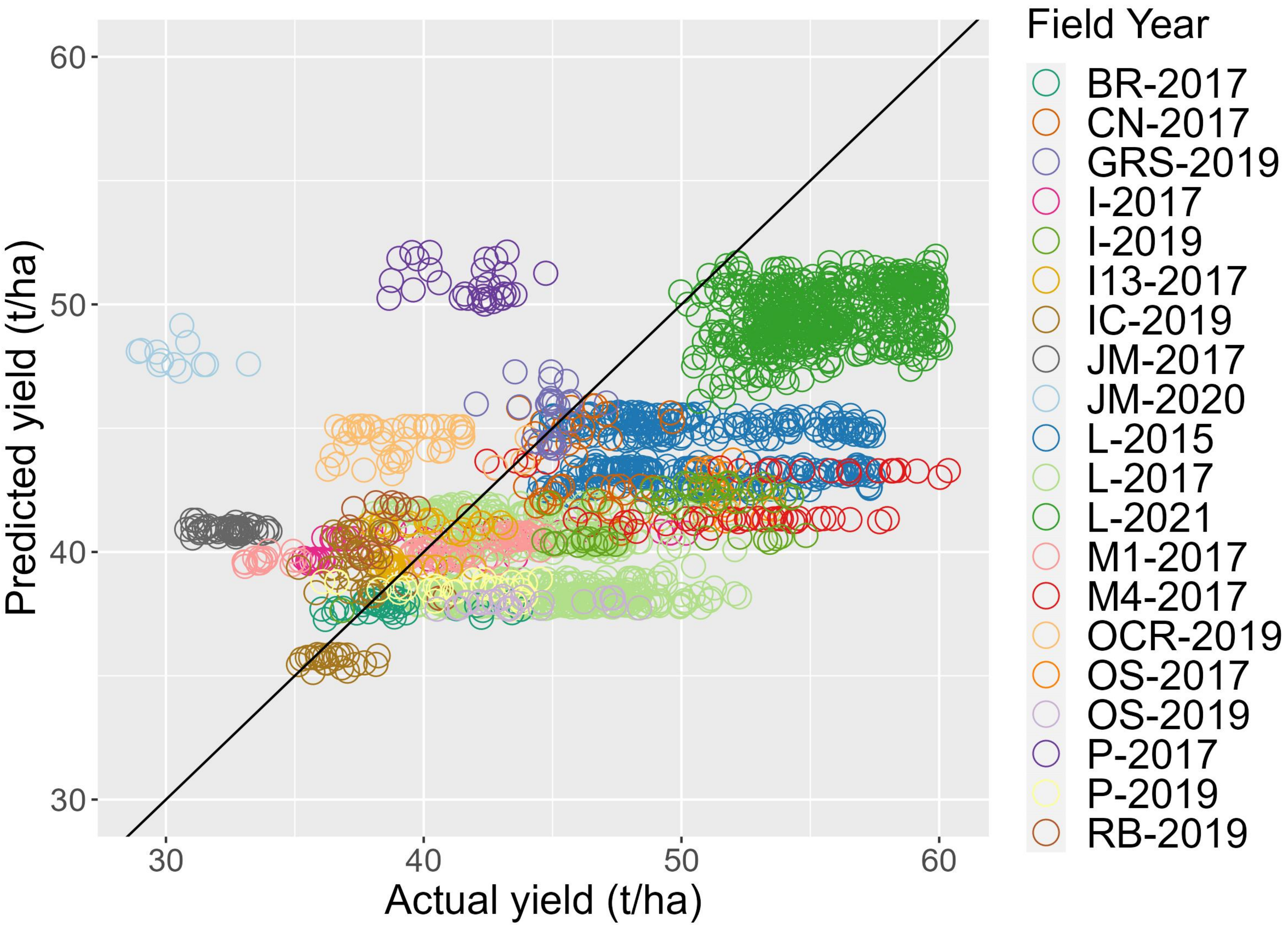
Number of Features Sampled at Each Node Split:

6

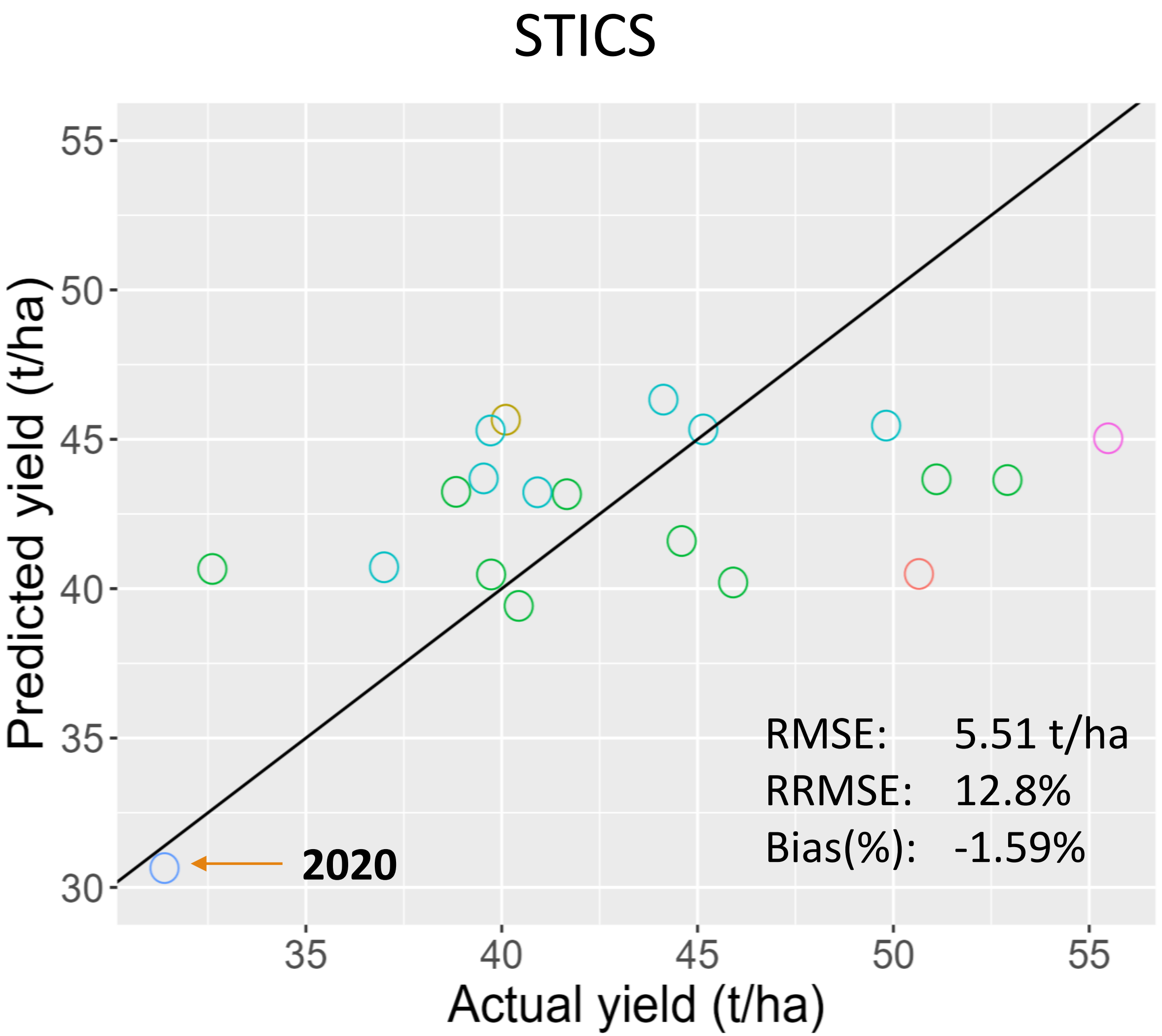
Number of Trees:

300

Random Forests for Unseen Field Years



Performance on Field Years



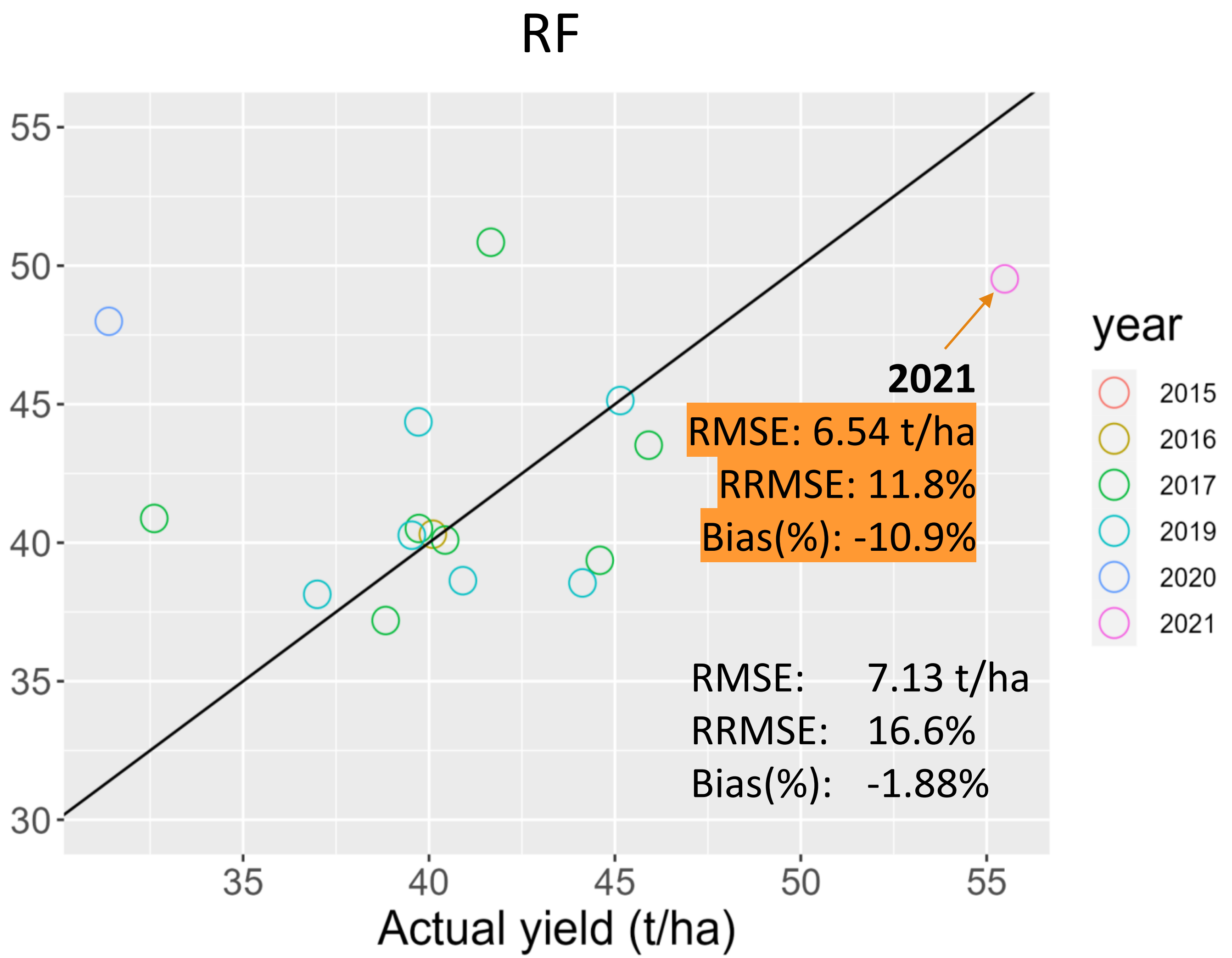


Image Source: https://www.cbc.ca/news/canada/prince-edward-island/pei-potato-u-s-border-closure-wart-planting-1.6393944

Conclusion

- Both models performed well
- STICS outperformed RF overall but the better model varied for individual field years
- Crop models calibrated using crop coefficient need to be recalibrated for resistance approach
- RF can produce good predictions at the field scale

Next Steps

Testing RF Performance

- Increase dataset size including CANSIS data
- With cross validation

Testing "Hybrid Approach"

- Combining crop model outputs with ML

References

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–32. https://doi.org/https://doi.org/10.1023/A:1010933404324

Brisson, N., Gary, C., Justes, E., Roche, R., Mary, B., Ripoche, D., Zimmer, D., Sierra, J., Bertuzzi, P., Burger, P., Bussière, F., Cabidoche, Y. M., Cellier, P., Debaeke, P., Gaudillère, J. P., Hénault, C., Maraux, F., Seguin, B., & Sinoquet, H. (2003). An overview of the crop model STICS. *European Journal of Agronomy*, *18*(3–4), 309–332. https://doi.org/10.1016/S1161-0301(02)00110-7

Boote, K. J., Jones, J. W., & Pickering, N. B. (1996). Potential uses and limitations of crop models. Agronomy Journal, 88(5), 704–716. https://doi.org/10.2134/agronj1996.00021962008800050005x

Crépeau, M., Jégo, G., Morissette, R., Pattey, E., & Morrison, M. J. (2021). Predictions of soybean harvest index evolution and evapotranspiration using STICS crop model. *Agronomy Journal*. *113*, 3281−3298. https://doi.org/10.1002/agj2.20765

Everingham, Y., Sexton, J., Skocaj, D., & Inman-Bamber, G. (2017). Accurate prediction of sugarcane yield using a random forest algorithm. *Agronomy for Sustainable Development*. https://doi.org/10.1007/s13593-016-0364-z

FAO. (2021). *World Food and Agriculture - Statistical Yearbook 2021*. Rome. https://doi.org/10.4060/cb4477en

Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., Timlin, D. J., Shim, K. M., Gerber, J. S., Reddy, V. R., & Kim, S. H. (2016). Random forests for global and regional crop yield predictions. *PLoS ONE*, *11*(6). https://doi.org/10.1371/journal.pone.0156571

Maqsood, J., Farooque, A. A., Wang, X., Abbas, F., Acharya, B., & Afzaal, H. (2020). Contribution of climate extremes to variation in potato tuber yield in Prince Edward Island. *Sustainability (Switzerland)*, *12*(12). https://doi.org/10.3390/SU12124937

Morissette, R., Jégo, G., Bélanger, G., Cambouris, A. N., Nyiraneza, J., & Zebarth, B. J. (2016). Simulating potato growth and nitrogen uptake in eastern Canada with the STICS model. *Agronomy Journal*, *108*(5), 1853–1868. https://doi.org/10.2134/agronj2016.02.0112

Shahhosseini, M., Martinez-Feria, R. A., Hu, G., & Archontoulis, S. v. (2019). Maize yield and nitrate loss prediction with machine learning algorithms. *Environmental Research Letters*, *14*(12), 124026. https://doi.org/10.1088/1748-9326/ab5268

Shahhosseini, M., Hu, G., & Archontoulis, S. v. (2020). Forecasting Corn Yield With Machine Learning Ensembles. *Frontiers in Plant Science*, *11*, 1120. https://doi.org/10.3389/fpls.2020.01120

Thank you

mariaelisa.polsinelli@mail.mcgill.ca

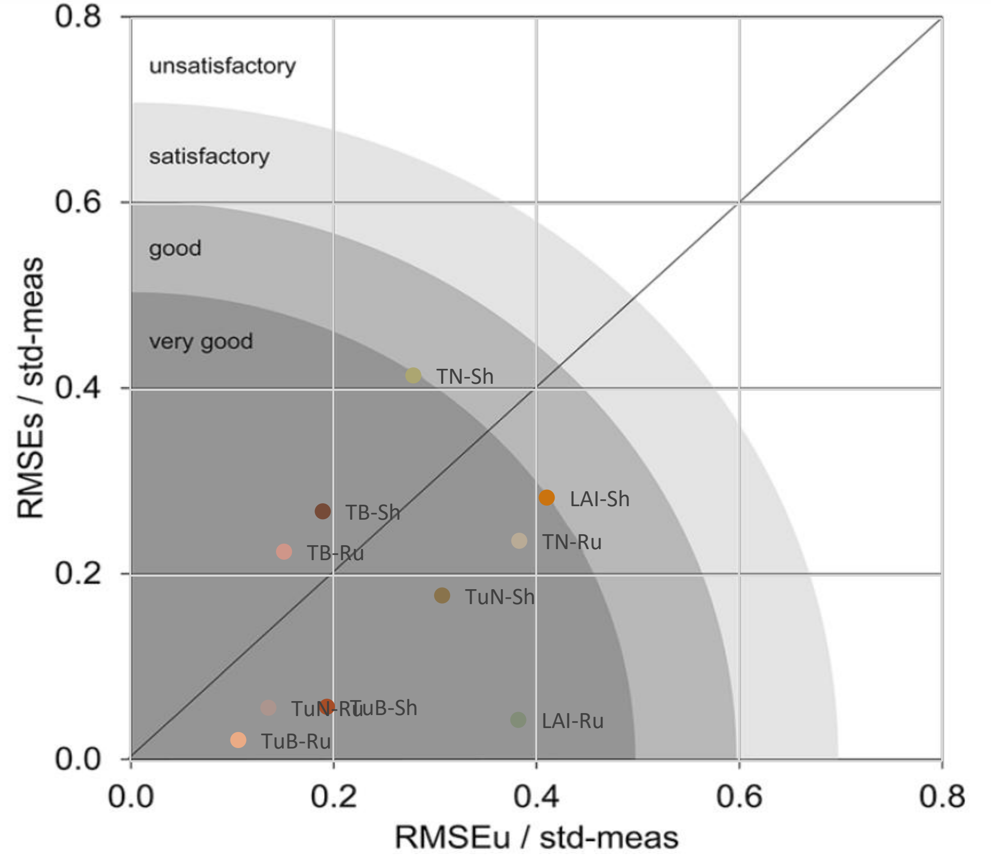
mariaelisa.polsinelli@agr.gc.ca

Acknowledgments

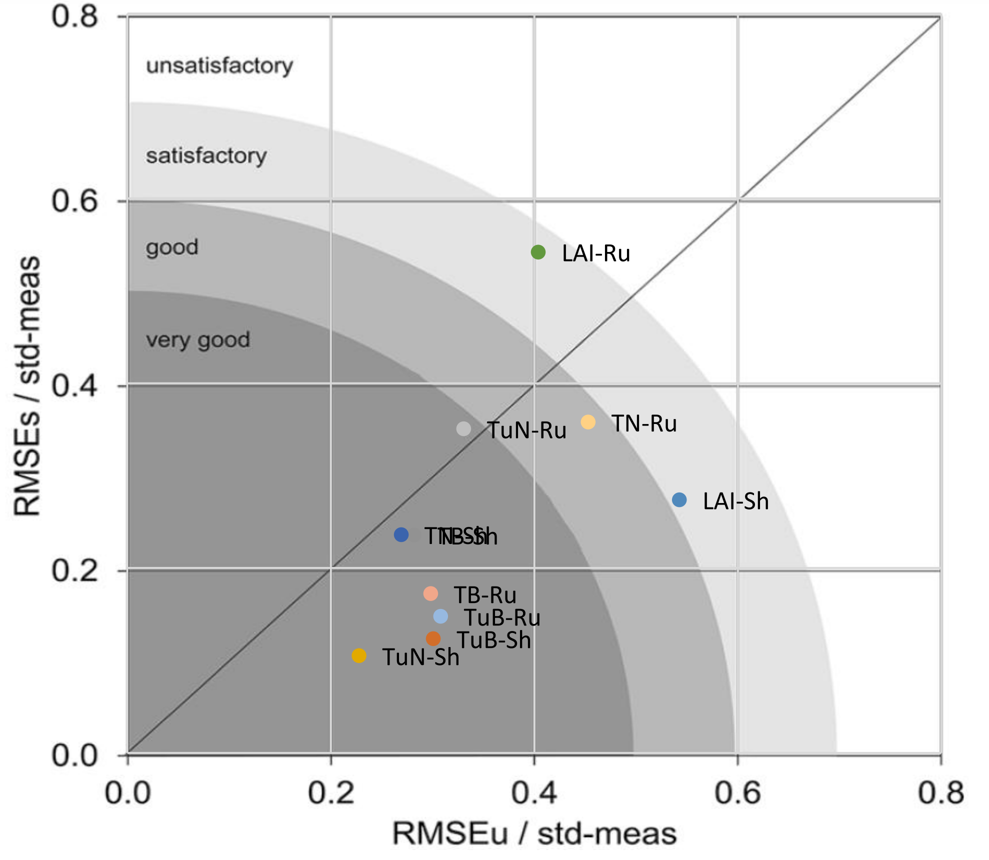
The funding for this project was provided by Agriculture and Agri-Food Canada
René Morissette and Guillaume Jégo for their help and assistance with STICS
Kristen Murchison for climate data processing, gap filling and technical support

Morissette et. al (2016) Original Calibration and Evaluation

Calibration

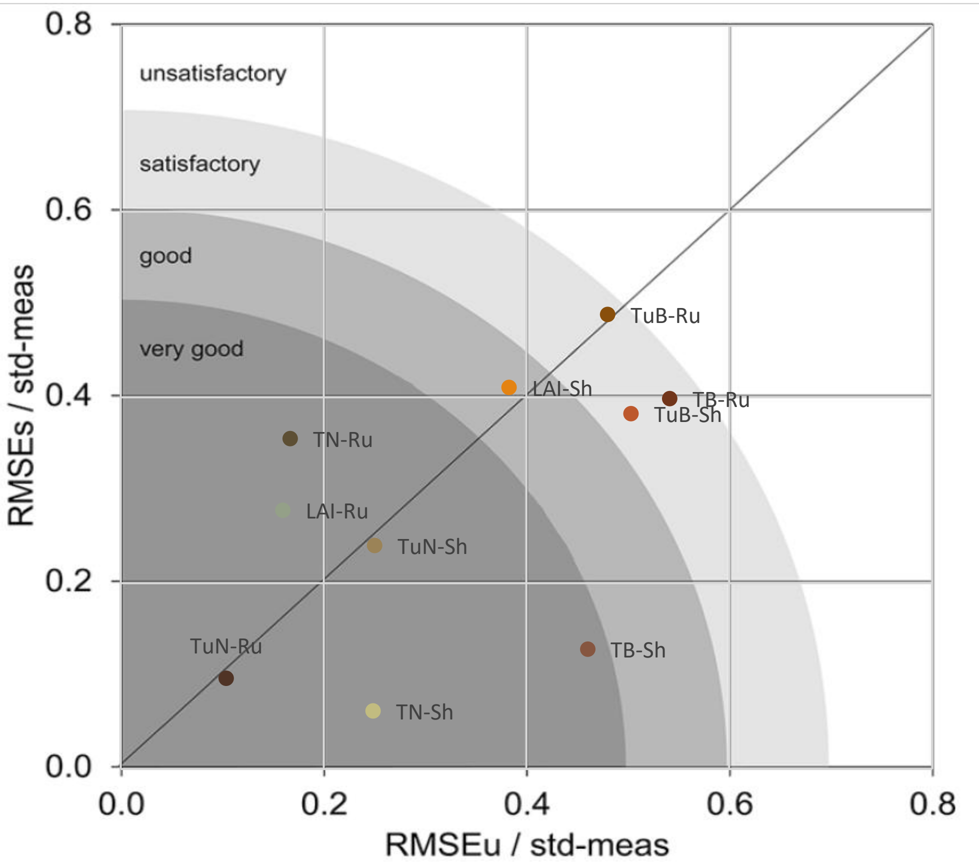


Evaluation

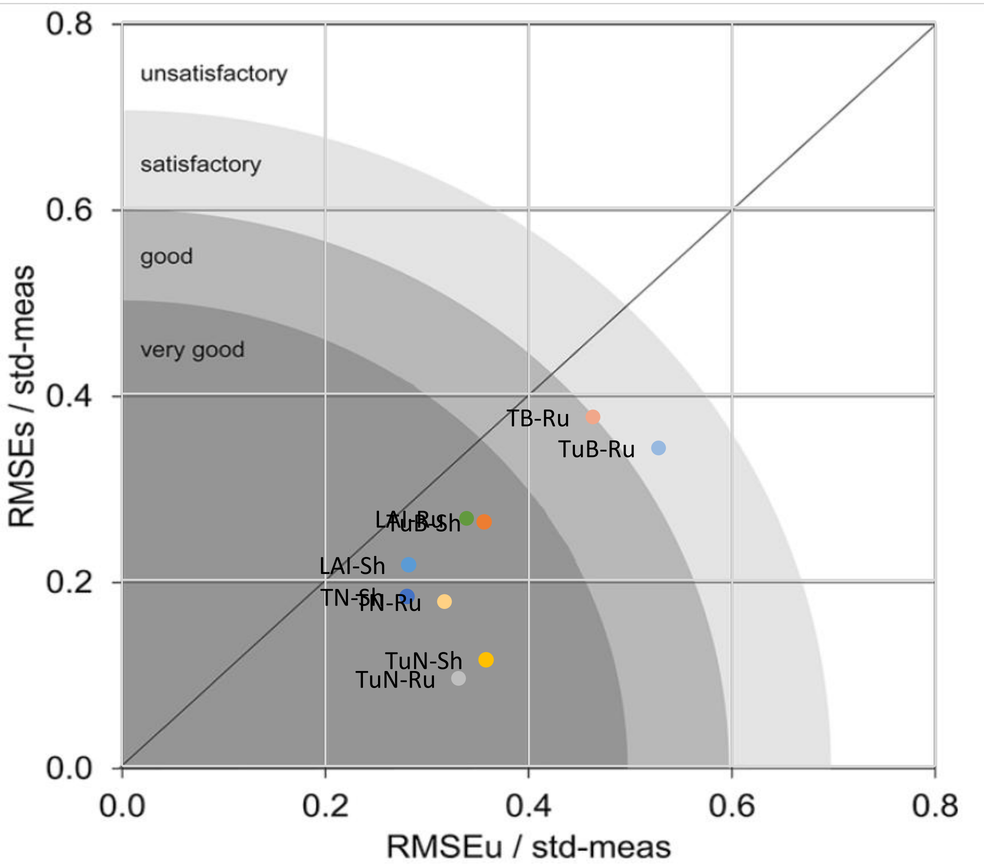


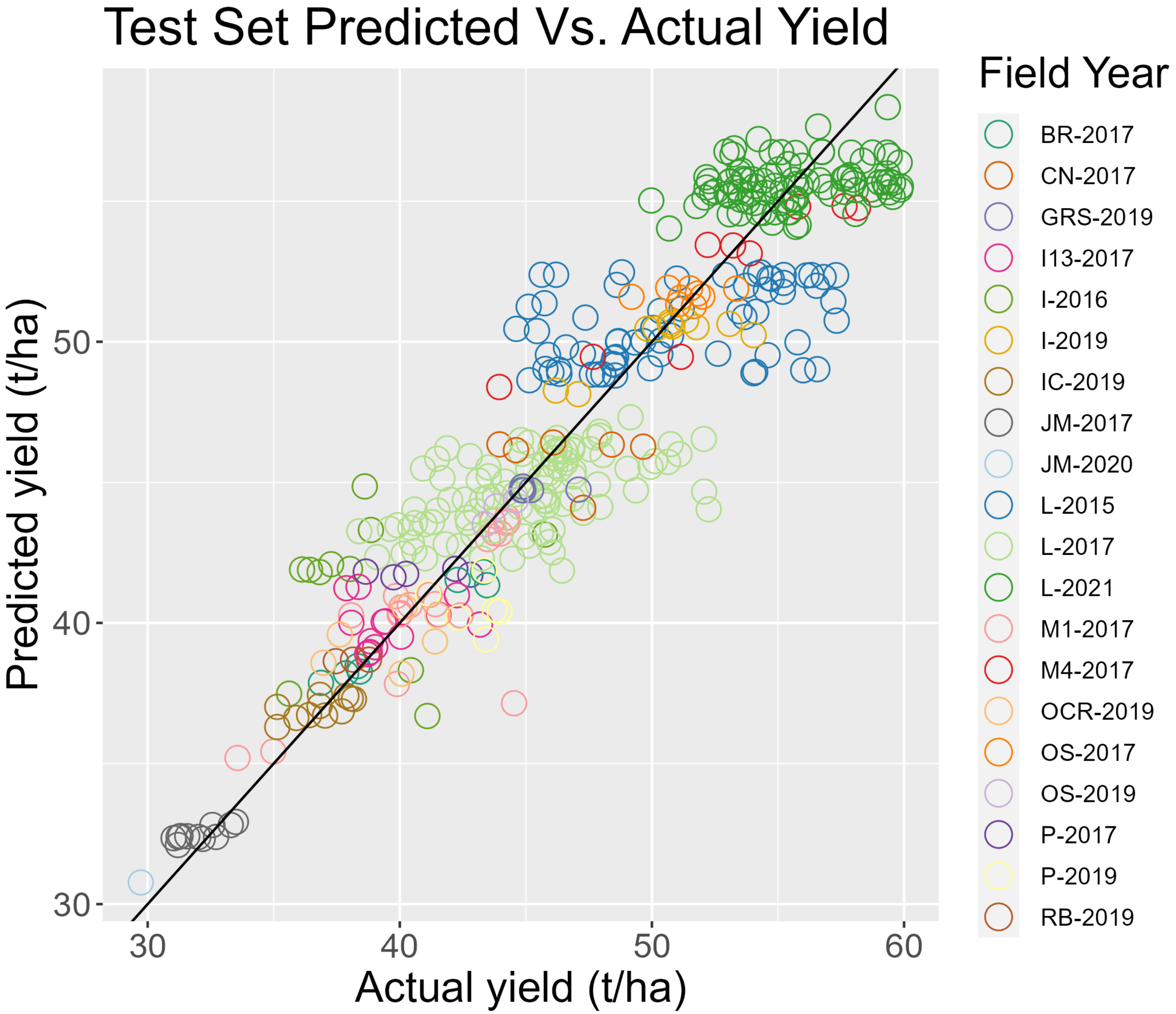
Morissette et. al (2016) Calibration and Evaluation after Resistance Approach Adjustments

Calibration



Evaluation





Random Forest Fine-Tuning

RMSE

2.54 t/ha

RRMSE

5.35%

Bias(%)

0.28%

Test Sets for Random Forests for Individual Field Years

Field Year	Number of Datapoints	Field Year	Number of Datapoints
L-2015	347	P-2017	32
I-2016	57	GRS-2019	28
BR-2017	38	I-2019	54
CN-2017	40	IC-2019	56
I13-2017	47	OCR-2019	47
JM-2017	45	OS-2019	25
L-2017	518	P-2019	32
M1-2017	77	RB-2019	17
M4-2017	55	JM-2020	12
OS-2017	25	L-2021	516

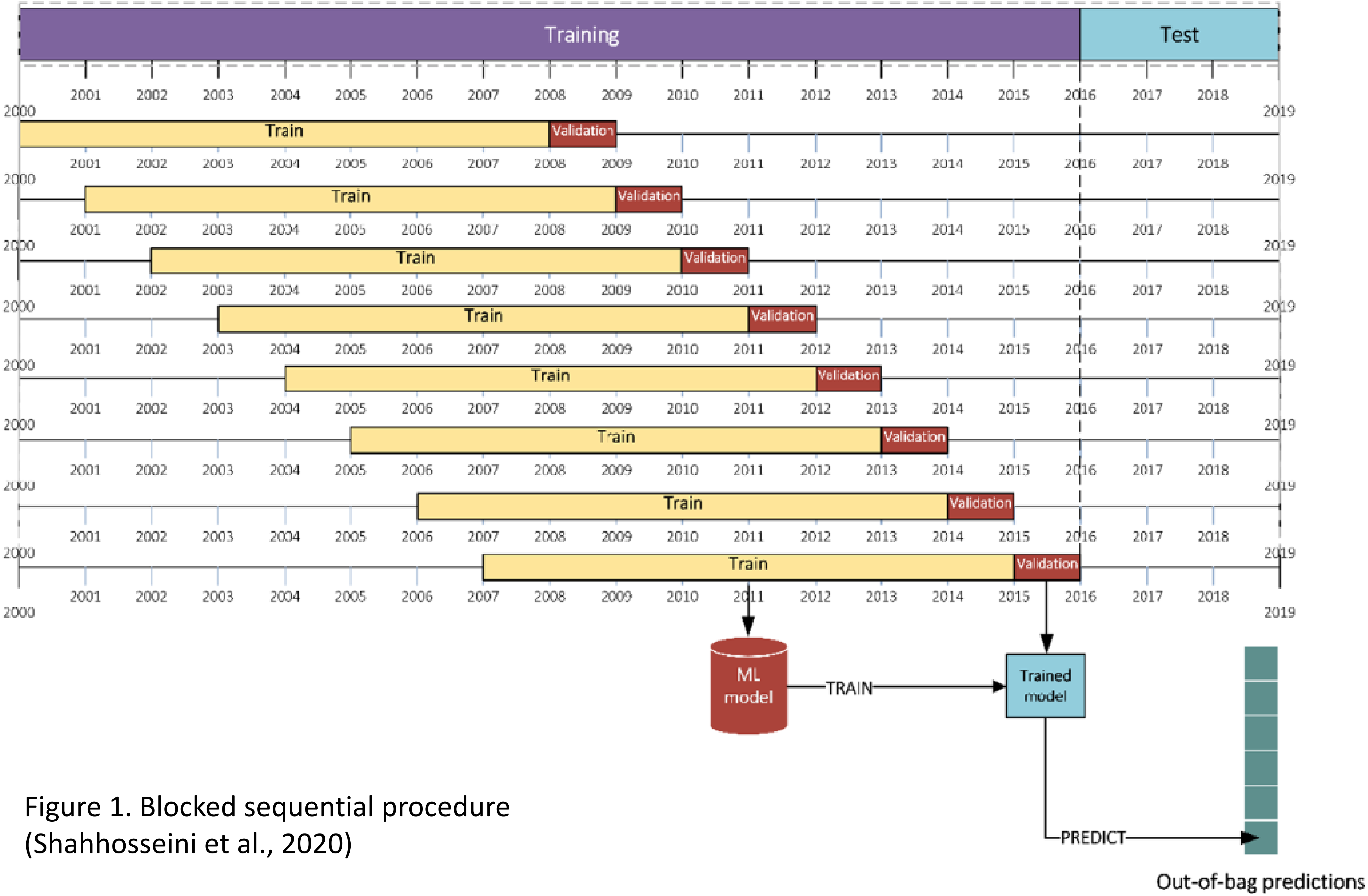


Figure 1. Blocked sequential procedure (Shahhosseini et al., 2020)