Promises and pitfalls of big data for agronomy

Towards decomposing yield gaps and benchmarking resource-use efficiencies at local level

João Vasco Silva (PhD)

Agronomy-at-scale Data Scientist

Sustainable Intensification Program

CIMMYT-Zimbabwe

October 2, 2020

Outline

1. **Introduction**
   - Grand challenges
   - Sustainable intensification
   - Big data
2. **Decomposing yield gaps**
   - Concepts and definitions
   - Application to contrasting farming systems
3. **From small to big data**
   - Data and methods
   - Preliminary results
4. **Take-home messages** and **future plans**

Grand challenges for the 21st century

1. Ensure food and nutrition security for all
2. Avoid land expansion and biodiversity loss
3. Climate change adaptation and mitigation
4. **Diverging paradigms**

Sustainable Intensification (SI)

- **Narrowing yield gaps on existing land while increasing resource-use efficiency**
- SI is contentious and trade-offs between sustainability and intensification need to be made explicit (Struik et al., 2014)
- Scale matters when talking about SI – opportunities at field, farm and regional level differ per farming system
- Prioritization of research agenda on sustainable intensification for staple crops (Cassman and Grassini, 2020)
- Big hope for ‘big data’ from farmers to deliver agronomic yield gains and environmental standards at scale

Big data, the end of traditional agronomy?

- **Big data → “**high volume, velocity and variety of information to require specific analytical and technological methods for its transformation into value”

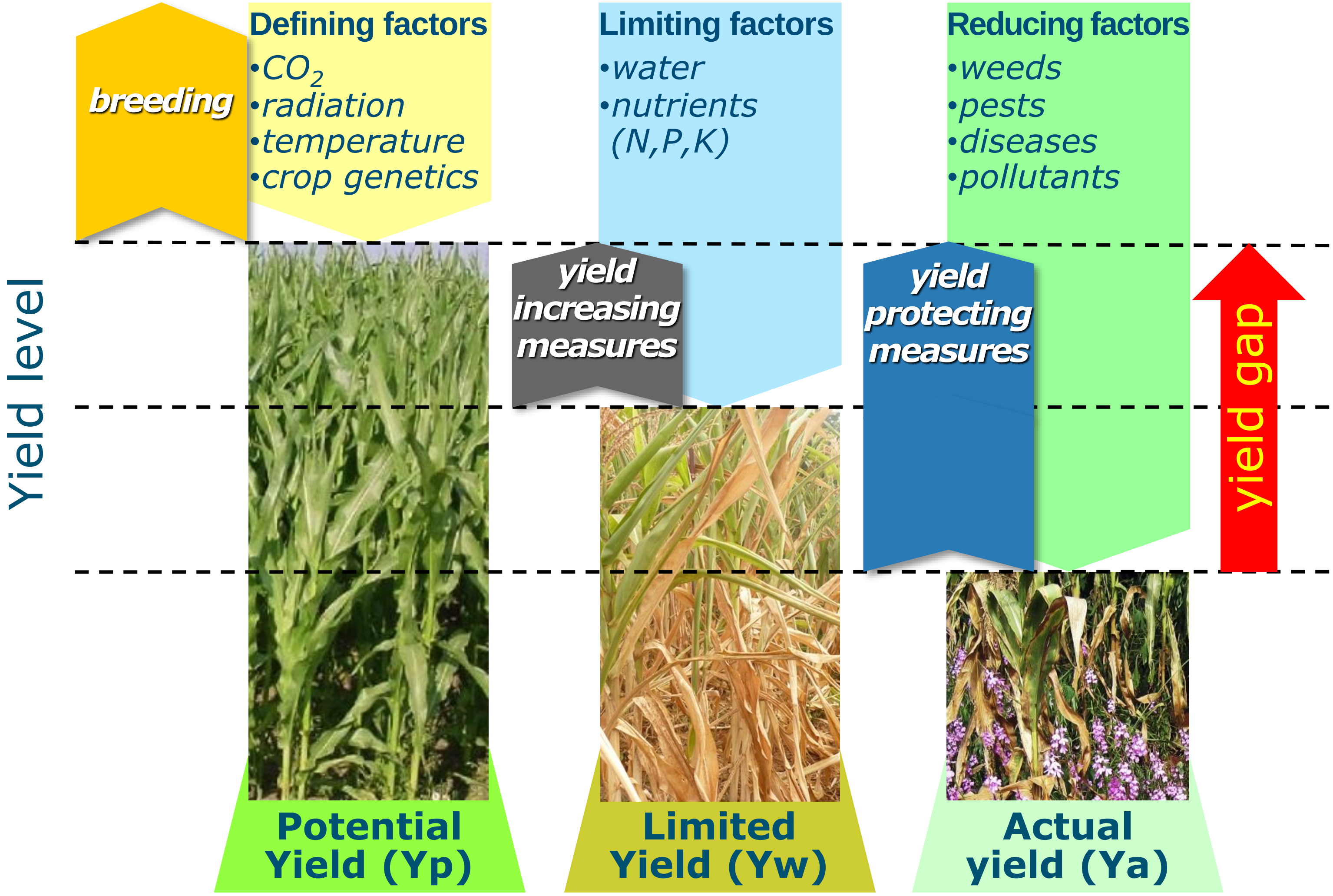
- **The opportunities:**
  - Large amounts and more complete data available from individual farms
  - Spatial explicit weather and soil data widely available
  - Equivalent to run hundreds of trials to evaluate M x E interactions
  - Benchmarks for resource-use efficiency and environmental quality

- **The challenges:**
  - Ensure data quality without simplifying farmers’ reality
  - Scattered information, ownership and privacy issues
  - Agronomists need to master many different algorithms and tools

My research focuses on...

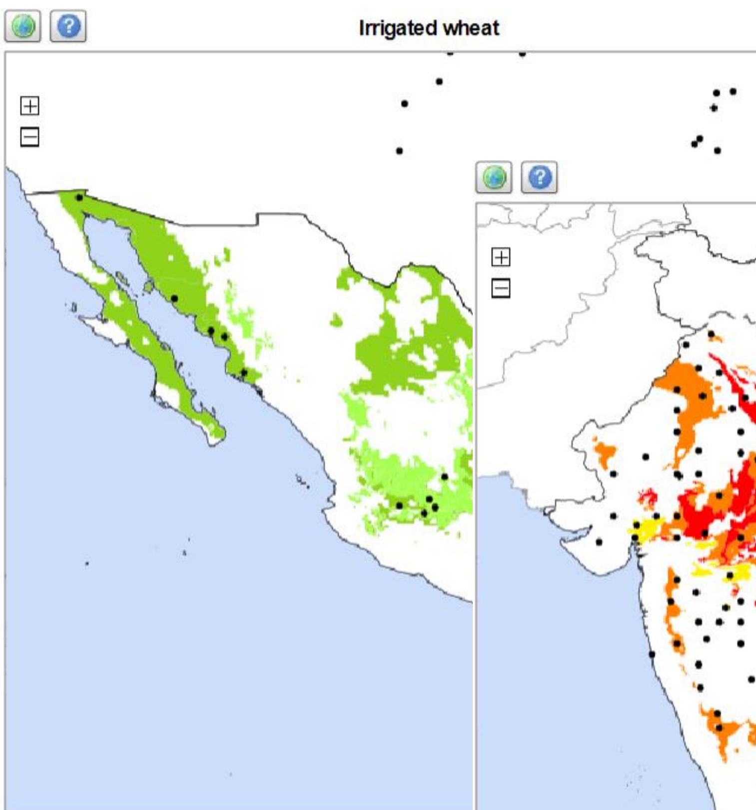
- **Decomposing & explaining yield gaps at systems level**
- Benchmarking resource use efficiencies at field level
- Crop model parametrization, improvement and application
- **‘Big data’ analysis across contrasting farming systems**

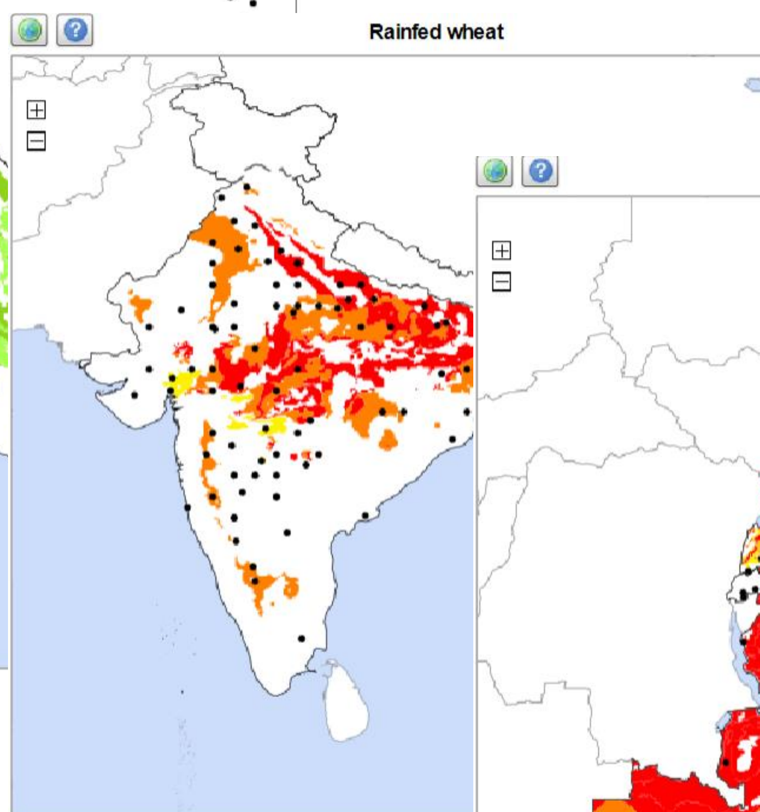
Concepts of production ecology

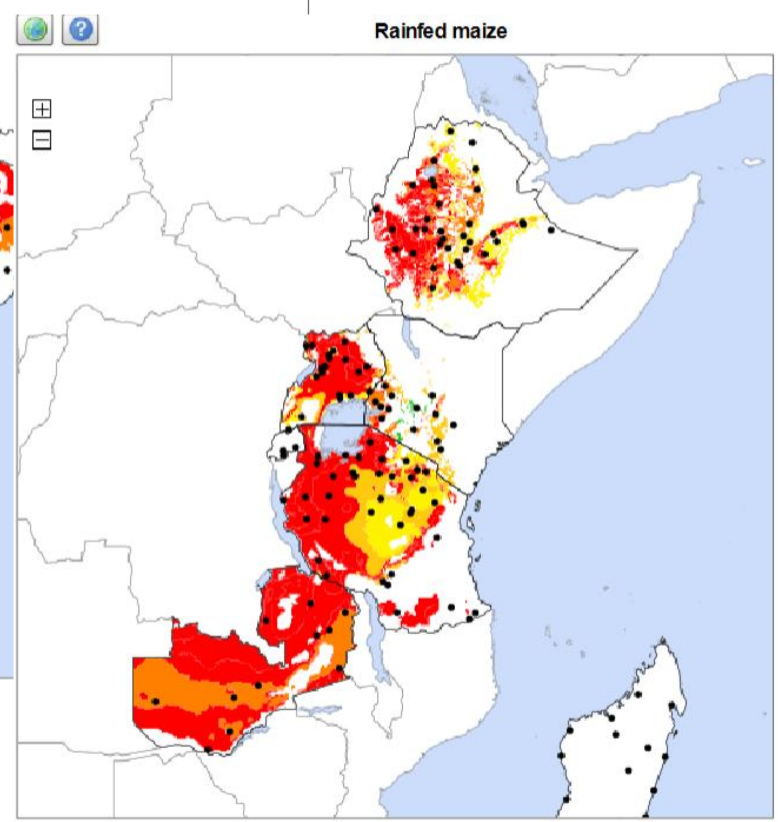


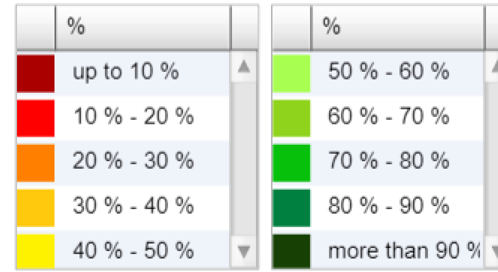
van Ittersum & Rabbinge (1997)
Field Crops Research

Yield gaps in CIMMYT research areas





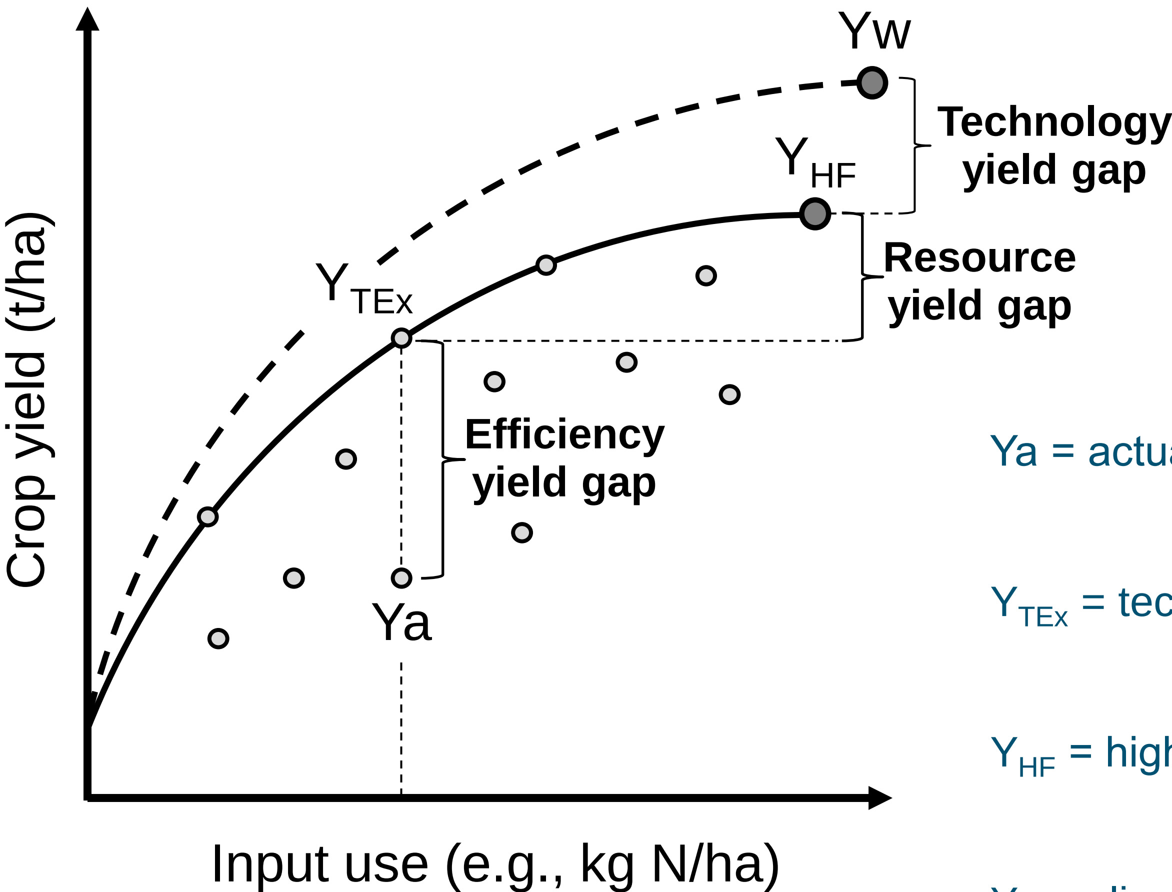




www.yieldgap.org

Decomposing yield gaps

Silva et al. (2017)
Eur. J. Agronomy



Ya = actual farmers' yields

- farm surveys: field level

Y_{TEx} = technical efficient yields

- stochastic frontier analysis

Y_{HF} = highest farmers' yields

- top 10^{th} percentile of Ya

Yp = climatic potential yield

- www.yieldgap.org

Contrasting farming systems

Mixed farming in Southern Ethiopia

Sample: 200 farms
Year: 2012
Farm size: < 2.5 ha
Crops: Maize in Hawassa and wheat in Asella

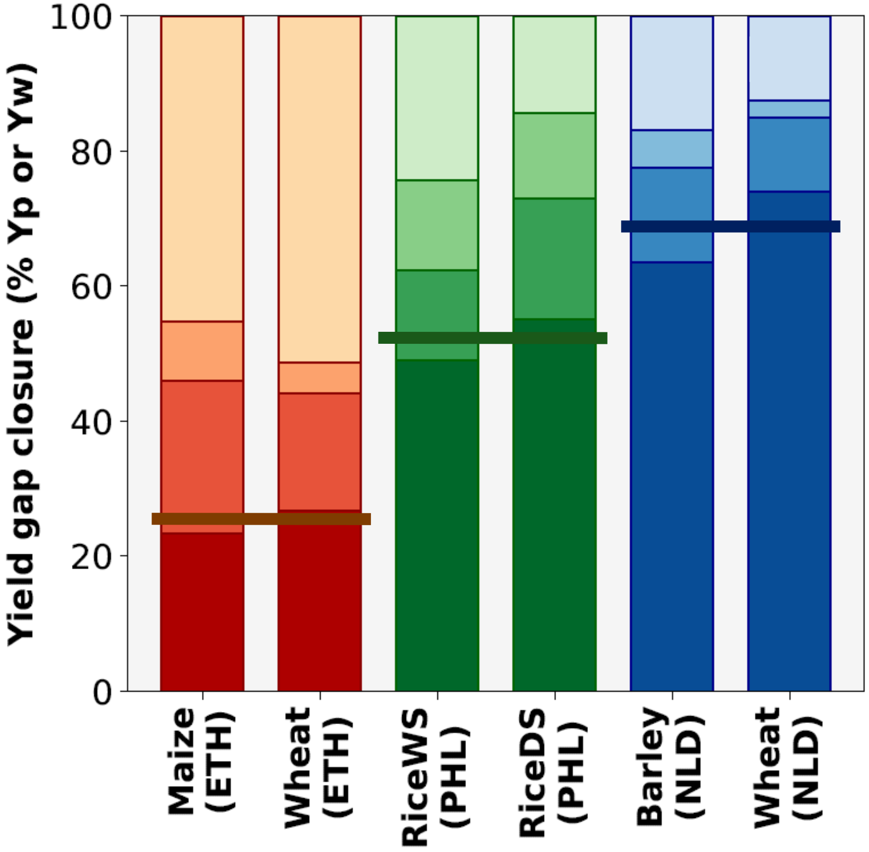
Rice farming in Central Luzon, Philippines

Sample: 100 farms
Year: 1966-2012
Farm size: 1.7 ha
Crops: Rice (wet season and dry season)

Arable farming in the Netherlands

Sample: 175 farms
Year: 2008 - 2012
Farm size: ~60 ha
Crops: Wheat, barley, potato, sugar beet, onion

Causes of yield gaps



Southern Ethiopia

Large yield gap attributed to technology yield gaps.

Silva et al. (AgSys, 2019)

Central Luzon, Philippines

Medium yield gap due to efficiency, resource and technology yield gaps.

Silva et al. (2017a, EJA)

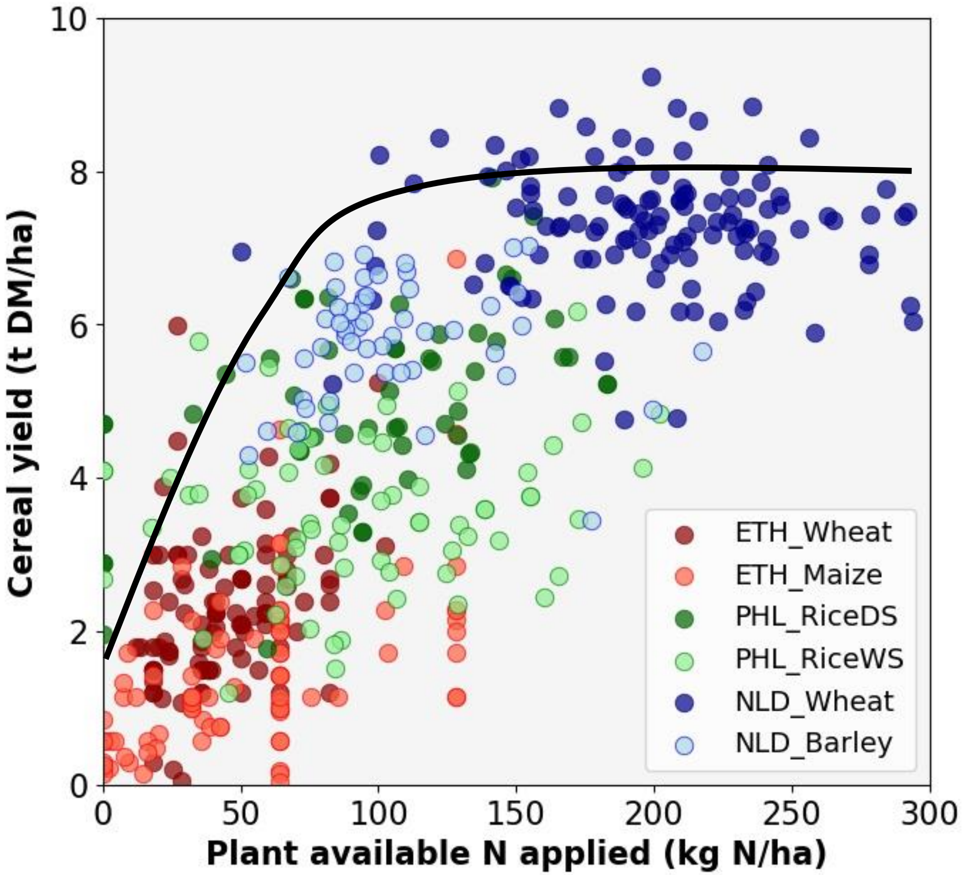
The Netherlands

Small yield gap attributed to efficiency yield gaps.

Silva et al. (2017b, AgSys)

Silva et al.
To be submitted

Sustainability vs. Intensification



Intensification

"More with more"

Ethiopia / Sub-Saharan Africa

"More with the less"

Philippines / Southeast Asia?

"Same with less"

Netherlands / Northwest Europe

Sustainability

Silva et al.

To be submitted

Other examples

Wheat (& maize) in Ethiopia

- Fine-tuning current practices can deliver the production needed to reach wheat and maize self-sufficiency;
- Reaching Yw requires seed rates, N rates and weeding beyond amounts currently used in highest yielding fields.

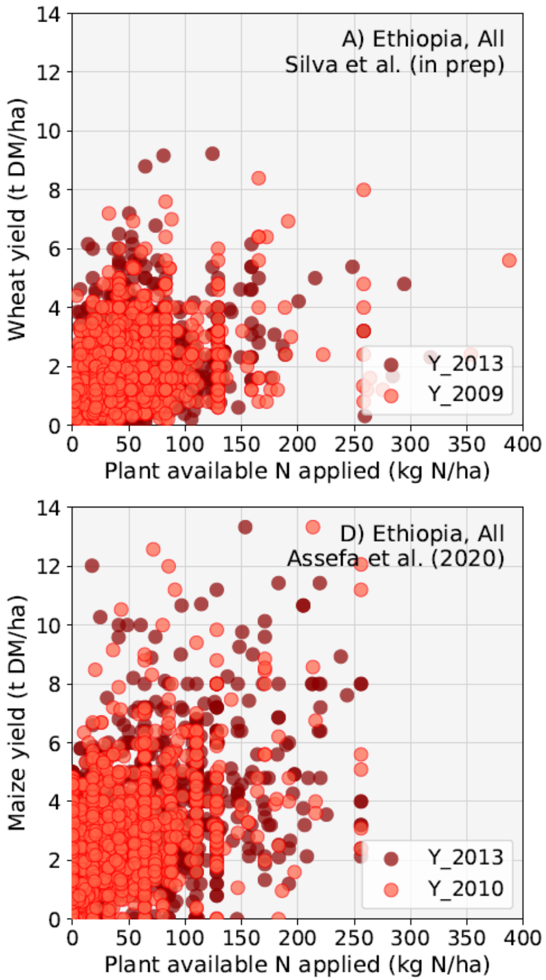
Assefa et al. (2020); Food Sec.

Silva et al. (under review); AgSD

- Framework expanded for economic & policy analysis (van Dijk et al., 2020)
- Wheat yield gaps in the Rwandan highlands (Baudron et al., 2019)
- Rice yield gaps in major rice-bowls of SE Asia (Stuart et al., in prep.)

From ‘small’ to ‘big data’

> 10**k** field x year combinations



Silva et al.
In preparation

From ‘small’ to ‘big data’

With which accuracy and precision can we predict crop yields in space and time?

Silva et al.
In preparation

Delaune (2018)
MSc thesis, PPS-WU

Preliminary results (R^2)

	Ethiopia Wheat	Ethiopia Maize	Philippines Rice WS	Philippines Rice DS	Netherlands Barley	Netherlands Wheat
Linear mixed model						
Full model	22.7 +- n.a.	27.3 ± n.a.	11.9 ± n.a.	27.7 ± n.a.	27.3 ± n.a.	41.0 ± n.a.
Cross-validation: Zone	20.5 ± 7.1	23.5 ± 4.3	12.3 ± 8.1	12.7 ± 7.8	14.7 ± 20.7	32.5 ± 13.2
Cross-validation: Farm	21.0 ± 4.3	25.7 ± 2.7	17.8 ± 3.8	13.8 ± 4.4	18.5 ± 11.7	37.7 ± 5.9
Cross-validation: Year	18.2 ± n.a.	26.0 ± n.a.	n.a.	n.a.	22.7 ± 1.3	6.6 ± 6.9
Random forest						
Full model	33.2 ± n.a.	34.0 ± n.a.	18.4 ± n.a.	35.3 ± n.a.	48.5 ± n.a.	57.8 ± n.a.
Cross-validation: Zone	13.9 ± 4.9	22.6 ± 5.1	3.8 +- 7.3	10.7 ± 12.9	29.7 ± 47.7	5.7 ± 4.1
Cross-validation: Farm	22.6 ± 3.0	28.4 ± 2.7	14.8 ± 3.3	33.7 ± 4.3	6.5 ± 5.6	40.6 ± 5.8
Cross-validation: Year	18.9 ± n.a.	25.4 ± n.a.	n.a.	n.a.	15.0 ± 1.9	2.0 ± 3.1

- Conclusions supported by RMSE and ME

Silva et al.
In preparation

Take-home messages

1. **Sustainable intensification** has different meanings in different farming systems and provides different opportunities at local level.

2. Technology **yield gaps** explain the largest share of the yield gap for smallholders in Africa. But, narrowing efficiency and resource yield gaps can deliver the production needed for self-sufficiency at national scale.

3. **Big data** are useful to describe cropping systems at regional scale, and derive benchmarks for farm performance, but not to predict and explain yield variability in time and space.

Future research activities

1. **Assemble databases and methods for doing ‘Agronomy-at-Scale’**
   - Databases with biophysical and socio-economic information
   - Returns on investment, technology targeting, sampling frames
2. **Decompose maize and wheat yield gaps in CIMMYT’s research sites**
   - Capitalize on existing datasets to provide global picture
   - Establish data collection tools and workflows
3. **Benchmark maize and wheat RUEs for smallholder farming systems**
   - Data-driven analysis of experimental (breeding) data
   - Crop model improvement and parametrization

With special thanks to

Martin van Ittersum (WUR)

Pytrik Reidsma (WUR)

Ken Giller (WUR)

Alice Laborte (IRRI)

Frederic Baudron (CIMMYT)

João Vasco Silva

CIMMYT-Zimbabwe

j.silva@cgiar.org

WAGENINGEN UR
For quality of life

IRRI International Rice Research Institute

CIMMYT International Maize and Wheat Improvement Center

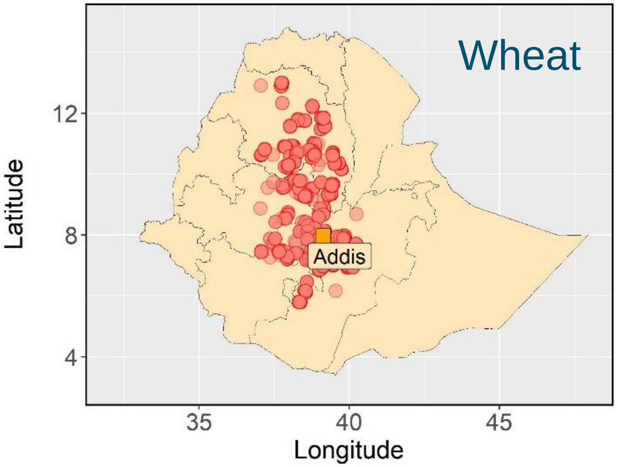
Other examples: RUEs in NW Europe

- 7 major arable crops in the Netherlands (>4000 fields 2015 – 2017)
- Yield gaps are ca. 30% of Yp and Yp achieved in some of those fields
- Actual water productivity is rather low due to large water surplus
- High NUE and high N surplus as a result of high N outputs and high N inputs

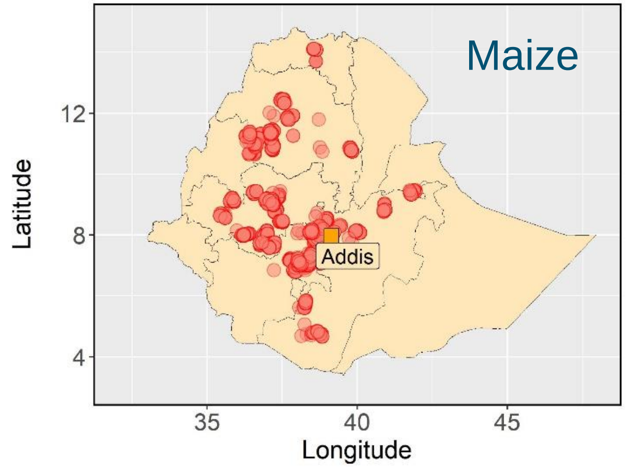
Silva et al. (2020, FCR)
Silva et al. (under review)

From ‘small’ to ‘big data’

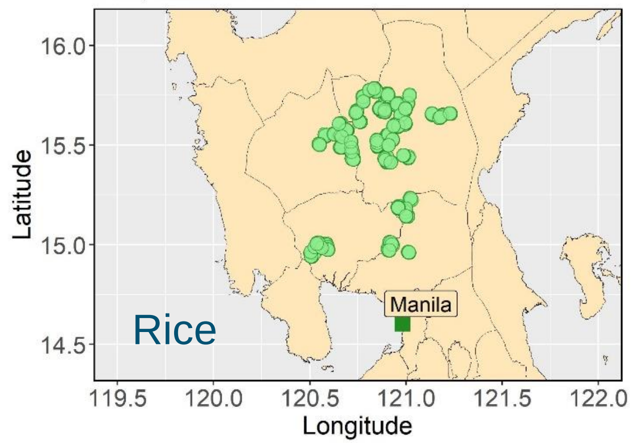
A) Wheat Adoption and Impact Surveys
CIMMYT, 2009/10 & 2013/14



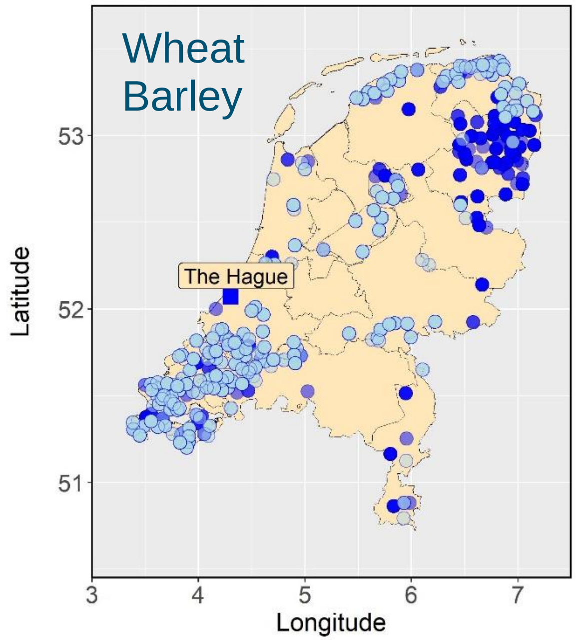
B) SIMLESA & DIVA Household Surveys
CIMMYT, 2011 & 2013



C) MISTIG Household Surveys
IRRI, 2014 DS & WS



D) Crop registration Agrovison B.V.
Agrovision B.V., 2015 - 2017



Silva et al.
In preparation