Food and Agriculture Organization of the United Nations

GLOBAL SYMPOSIUM ON
SOIL INFORMATION AND DATA

MEASURE
MONITOR
MANAGE

Insights from developing a global, harmonized, and normalized database of soil parameters for agricultural practices

Gaël Foëx

September 25-28, 2024
Nanjing, China

GLOBAL SOIL PARTNERSHIP

Outline

- Stenon: the need for high quality, normalized soil data
- Soil Database: fact and figures
- Data collection: methodology and logistical challenges
- Data standardization: extraction methods and laboratory testing
- Data harmonization: method transformations
- Conclusion

STENON

Overview

STENON

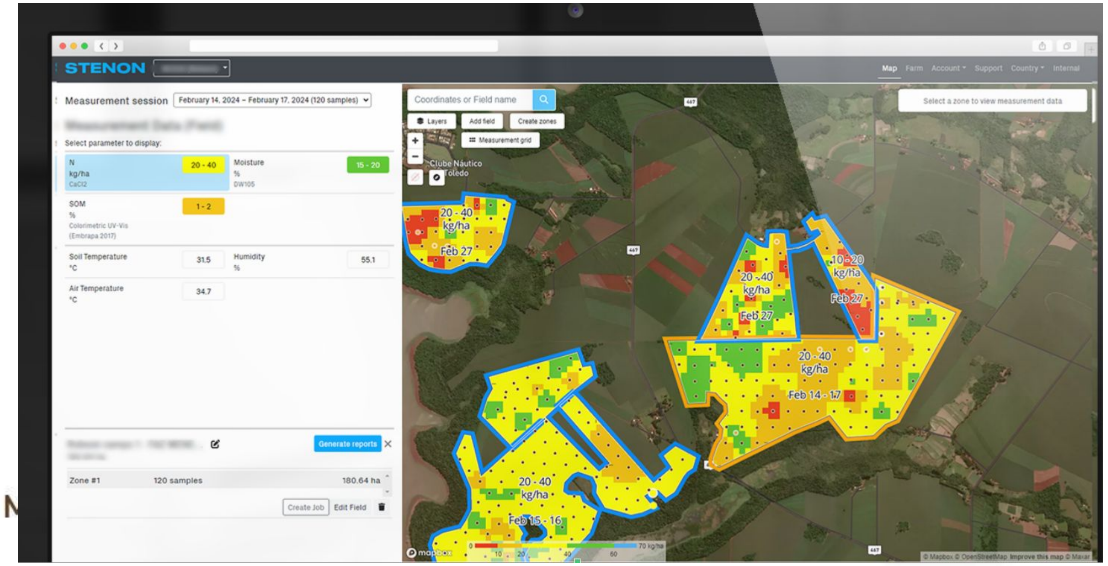
- German company (Potsdam), founded 2018, ~40 employees
- 3 core markets: Central Europe, KAZ, BRA
- Some partners: LAVORO, Eurasia Group (We feed the world!), syngenta, PEPSICO, McCain

Our vision: "Becoming the **global Soil Data Infrastructure** for Agri-business, creating value for our customers and building lasting relationships."

Our mission: "Being the **international leading technology for Soil Insights** and **Soil Fertility Data** making Agri-businesses more successful, improving sustainability, saving input costs for Growers, reducing over fertilization and optimizing yields over the full Ag value chain."

FarmLab

- **In situ real-time soil analysis**:
  - sensor data
  - ML models
  - online predictions
- Model training:
  - sensor vs. **physicochemical** data
  - model only as good as training data
  - **quality** & **homogeneity** crucial
- **<u>Building an extensive soil database</u>**:
  - soil sampling → **Stenon**
  - soil samples **analysis** → **laboratories**



Soil Database

Geographical Coverage

- Europe: **DACH** (~1000 fields), UK (~100 fields), RUS (~45 fields)
- Central Asia: **KAZ** (~100 fields in North + ~25 fields in South)
- South America: **BRA** (~100 fields in SP/PN + ~25 new ones in MT)
- North America: USA-CA (~80 fields)
- Africa: EGY (~25 fields)

Field selection:

- pre-selection on soil maps to maximize **diversity** (texture/SOC/pH/etc)
- base camps + local scouting + support by local farmers

Database Content

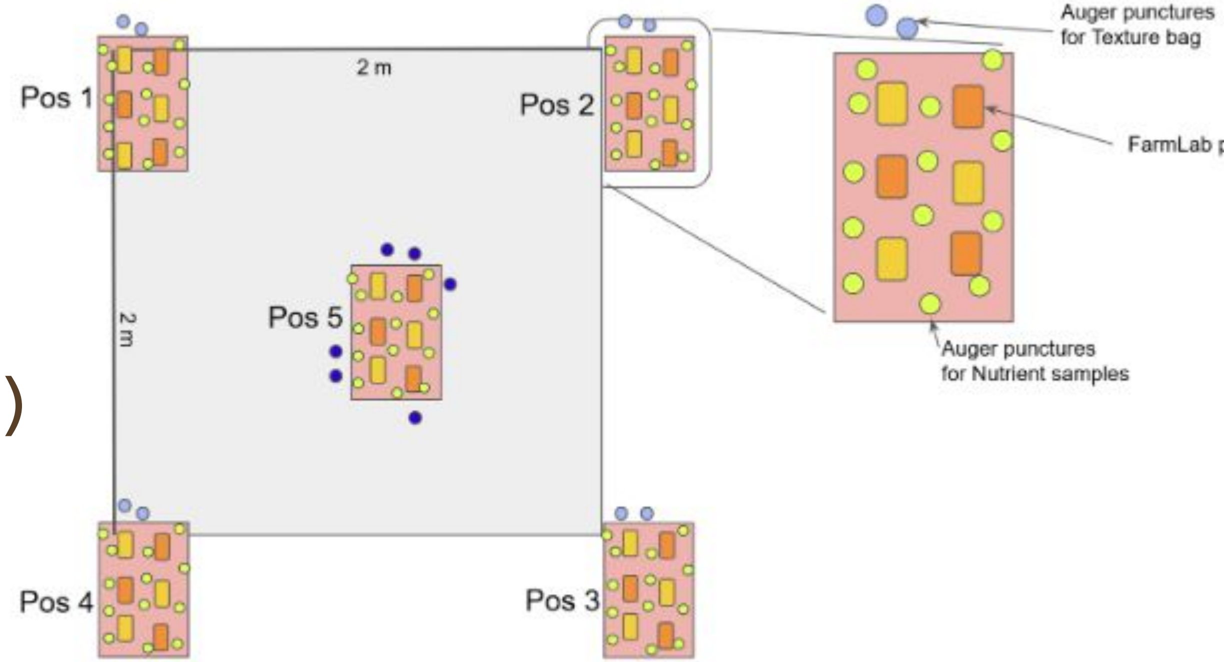
- **~25k soil samples**:
  - key nutrients for fertilization, e.g. **N**, **SOC**
  - soil texture: clay/sand/silt percentages → soil type
  - other: pH, moisture, etc
- Estimated worth: 2.5Mil€ (lab costs only!)

Data Collection

Standardized Methodology

Consistent approach to sampling:

- spatial design
- material (augers)
- sampling depth (0-30cm for DE)
- soil quantities (250g/bag)
- soil mixing (bucket+spatula)
- storage (plastic/cotton bags)
- data tracking (labels, QR codes)



Sampling in action

Logistical Challenges

- **Field work**:
  - carrying heavy material
  - difficult soils, e.g. dry compact, frozen, stones
  - weather, e.g. hot/humid in BRA
  - local fauna, e.g. mosquitos, snakes
  - crops, e.g. corn before harvest
- Soil **conditioning**, **storage**, and **transport**:
  - freezing at **-18°C** for NO3
  - local storage of back-up samples for re-analysis
  - shipment to **local** and **DE** labs (self-delivery or local/international couriers)

Data Collection at Scale

Our CEO currently in BRA double-hammering augers for extra efficiency!

Few samples on a nearby field on a nice day? Easy!

Scaling to 100 fields in a foreign country 1000s km away on a tight schedule?

- hard work
- time consuming
- expensive
- hazardous

Data Standardization

Laboratories Selection

1. **Search** for labs: internet, local partners, etc
2. **Contact** labs (language barrier!):
   - availability, capacity, processing time, price, etc
   - **extraction methods**
3. **Pre-select** labs (study local standards):
   - **compatibility** matrix lab vs. lab
   - **compatibility** matrix local vs. “global” standards
4. **Test** labs on dedicated soil samples

Extraction Methods

- **Multiple regional “standards”** (public/academic/private):
  - DACH: DIN ISO, VDLUFA, Agroscope, …
  - UK: LECO, RB247, NRM, …
  - KAZ/RUS: GOST
  - BRA: EMBRAPA, IAC, TEDESCO, …
  - USA: AOAC, USDA, QuikChem, TGW, …
- **~150 analysis protocols** in our database

→ local standardization and global harmonization **mandatory** but **challenging**!

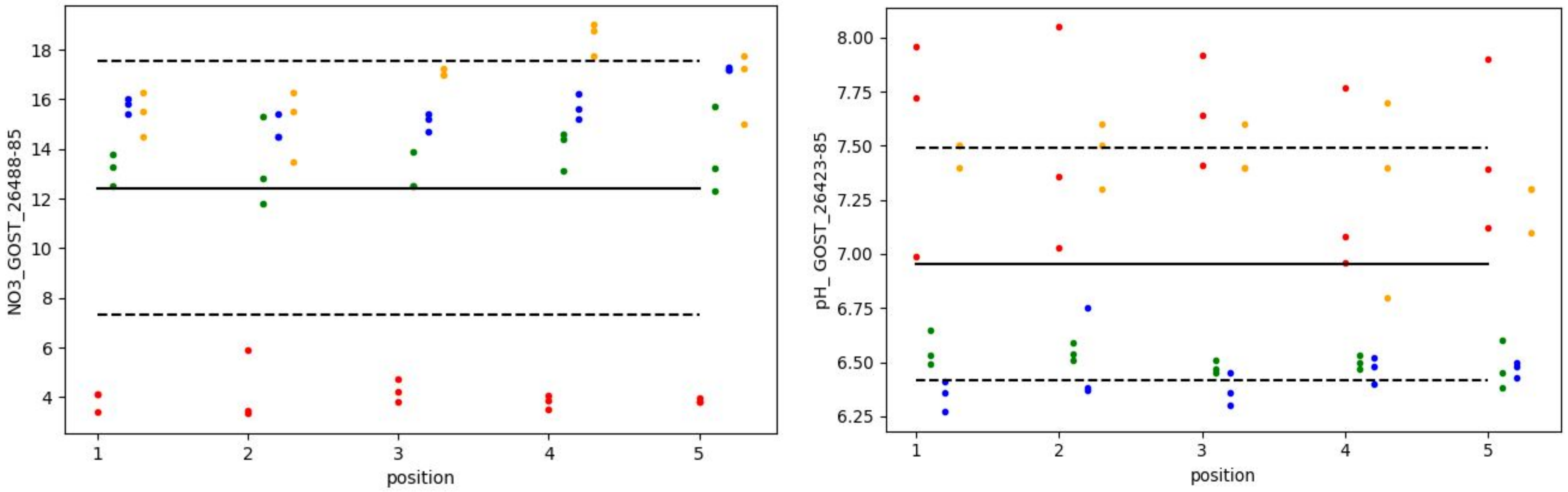
Laboratories Testing

same soil + same method = same results?

same soil by same lab?
~1%-10% differences

- Consistency on **same soil**
  - very small scatter ok (e.g. imperfections in soil mixing)
- Consistency between **spatially-close positions**
  - small scatter ok (small-scale variations)
  - global trends should be identical
- Consistency on **plot average**
  - very small scatter ok

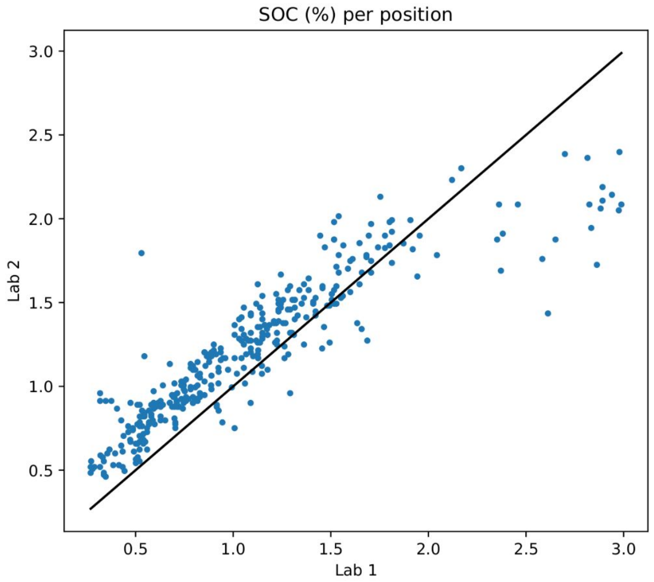
Laboratories Testing

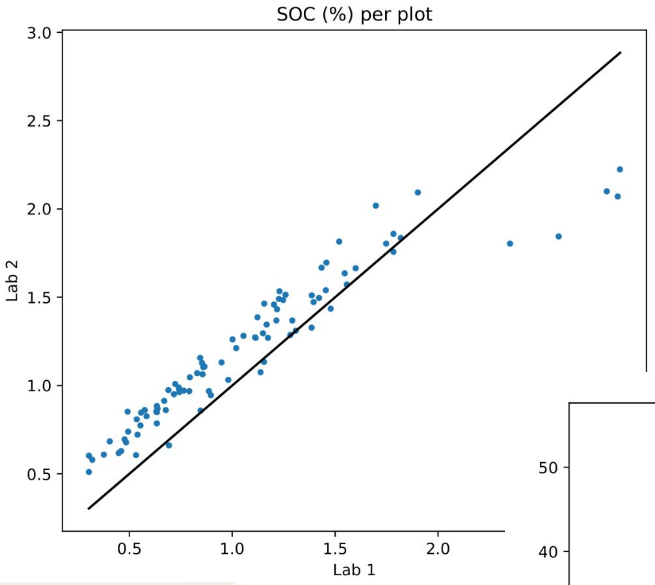


Pick Lab 1, maybe Lab 3 and Lab 4; exclude Lab 2

Pick Lab 1 and Lab 3; exclude Lab 2 and Lab 4

Post-Sampling Final Checks



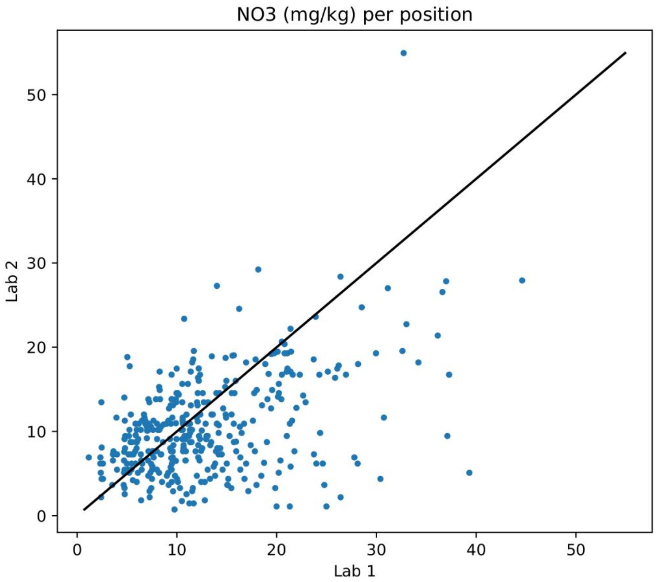


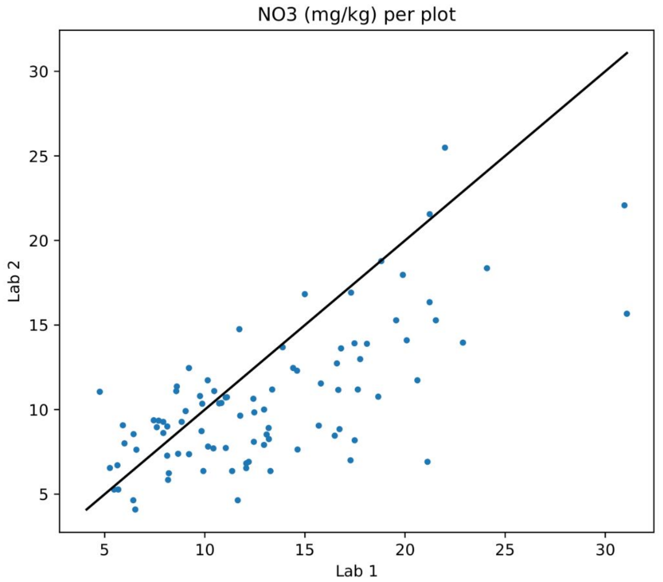
- limited scatter / few outliers
- small ~constant offset
- strong correlation

→ average labs :)
(reduce uncertainties by 1/√n; n=2→30%)

- large scatter / many outliers
- weak correlation

→ select single lab :(





Data Harmonization

Defining a Global Standard

many regional standards → no global standard → no harmonization?

Solution: use DE labs as standard ruler!

- Pros:
  - practical for us
  - high quality results, reliable, thoroughly vetted
- Cons:
  - send (frozen) soil across the globe ($, CO2)
  - DE standards
    - additional uncertainty from **mapping** back to local
    - methods not reliable for specific foreign soils? (e.g. low/high carbonate P/K GOST method)
  - sampling depth → additional mapping? (DE: 0-30cm, BRA: 0-20cm, UK/USA: 0-15cm)

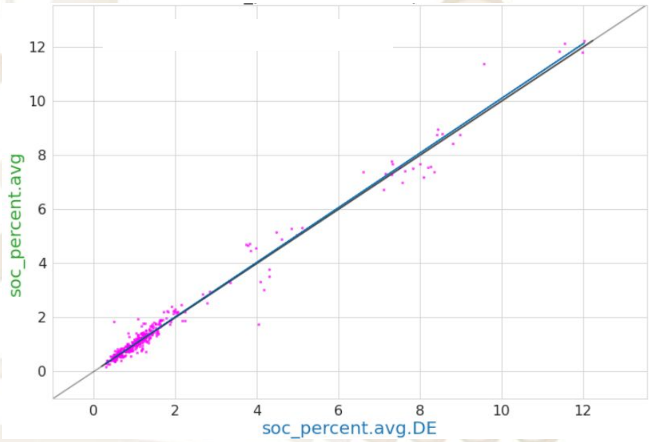
Implementing a Global Standard

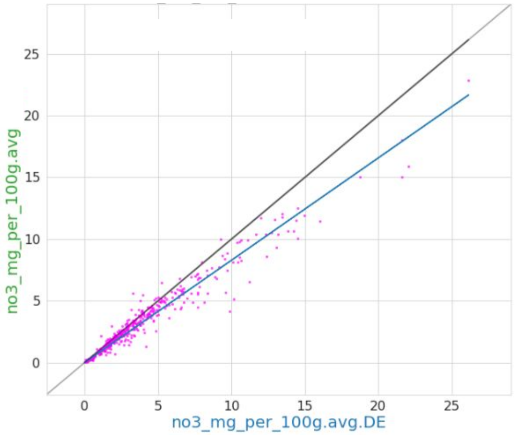
1. Obtain local and DE lab data (split soil between labs → extra $)
2. Train models on combined DE data
3. Build a **method transformation** DE→local
4. Convert DE predictions to local standards

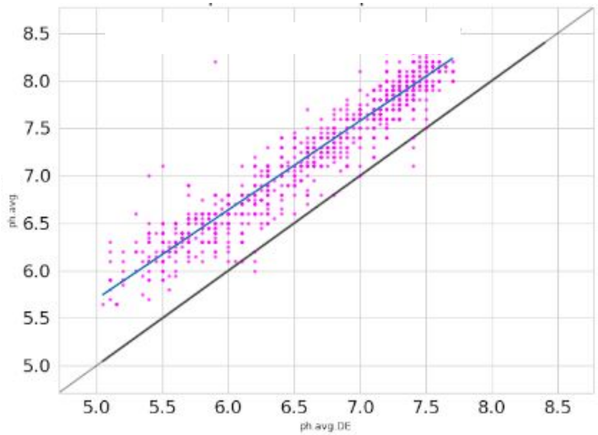
→ Leverage **harmonized global database** for **standardized local product**

Method Transformations

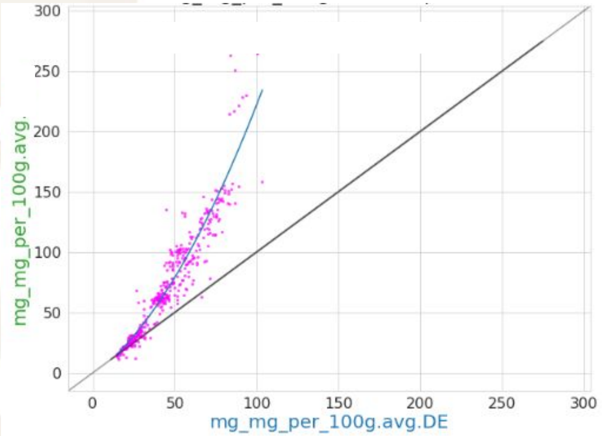
simple methods:

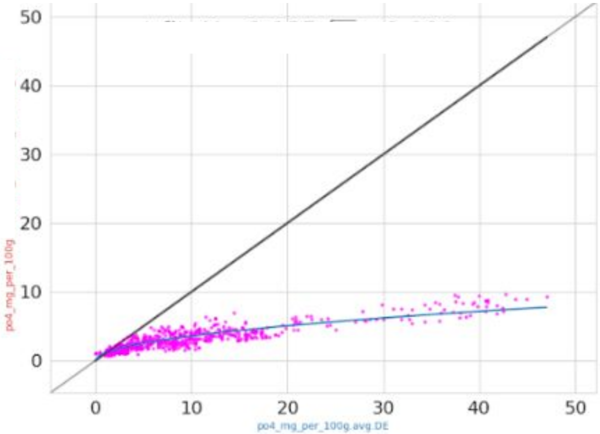






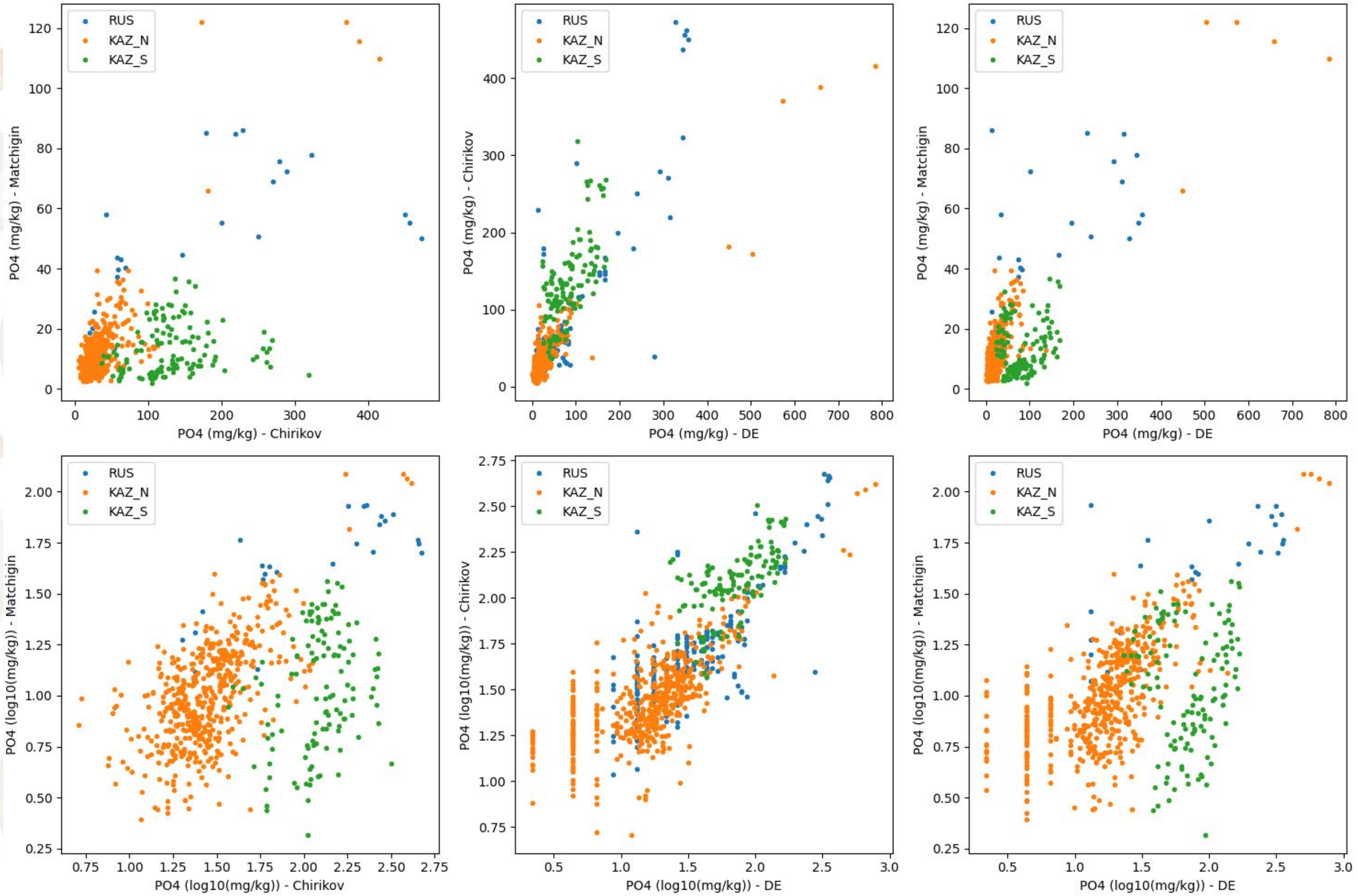
complex methods:





Method Transformations - PO4

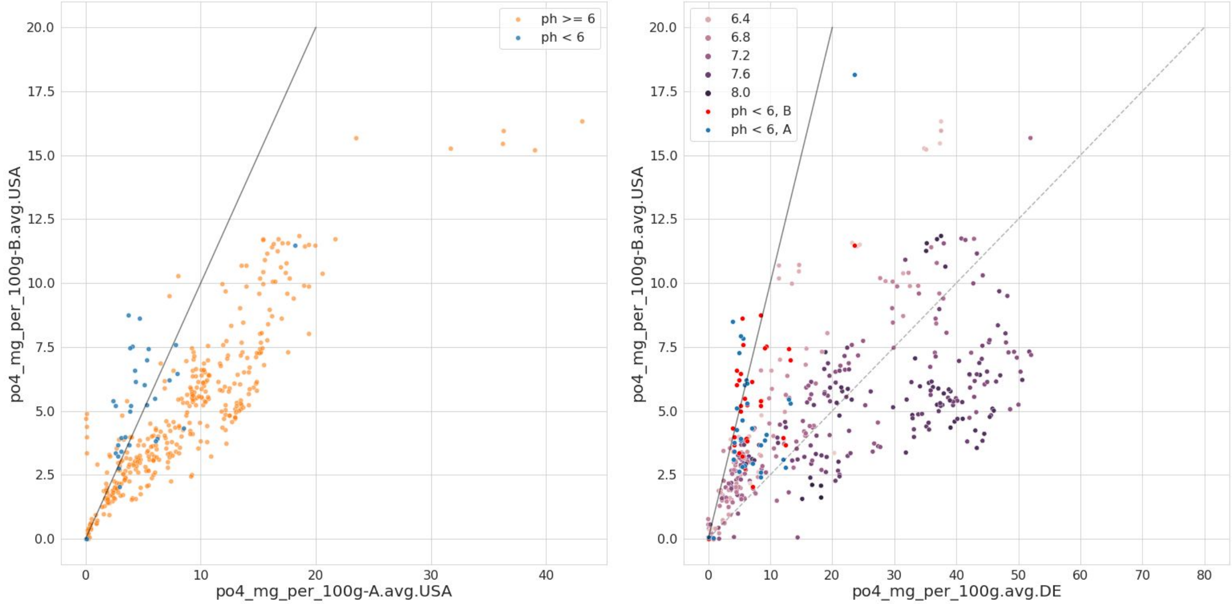
KAZ/RUS: Chirikov (low carbonate) vs. Matchigin (high carbonate)

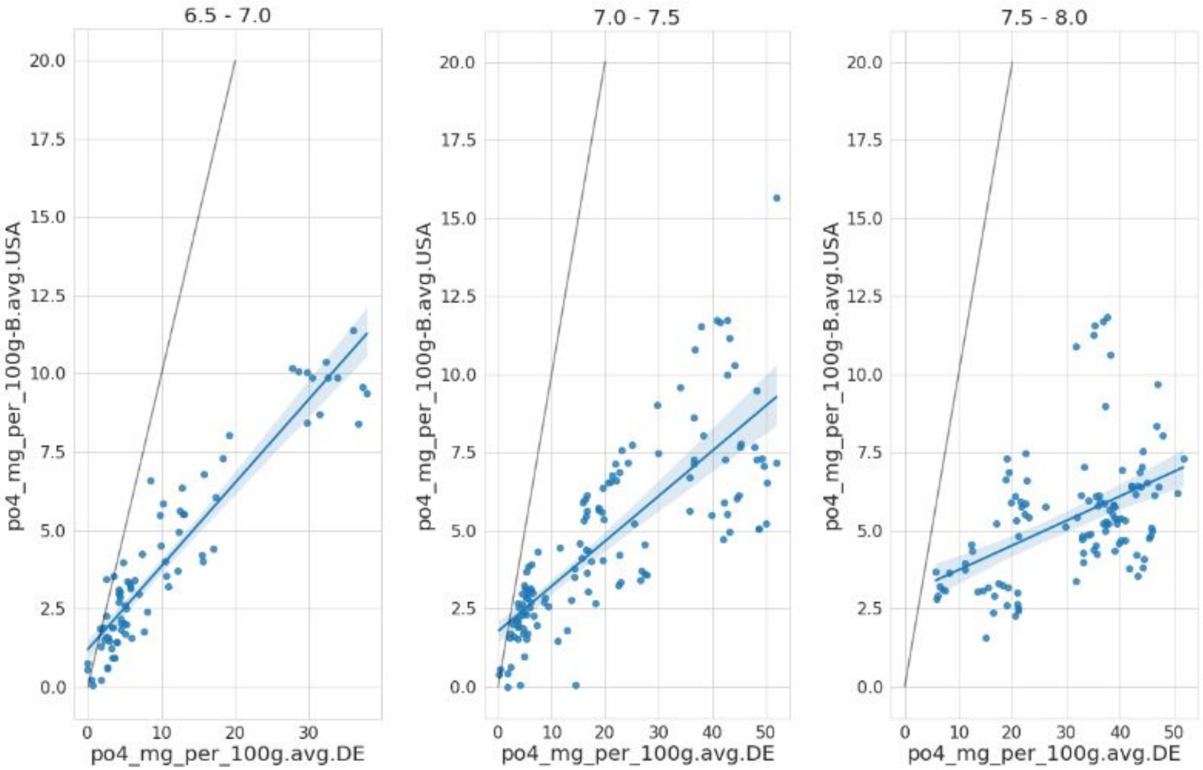


- Chirikov vs. Matchigin: weak/noisy
- Chirikov vs. DE: strong
- Matchigin vs. DE: strong but bimodal (KAZ_S)
- Chirikov ~ 10x Matchigin!
- Single DE method enough for all soils?

Method Transformations - PO4

USA: Bray (A; pH < 6) vs. Olsen (B; pH >= 6)





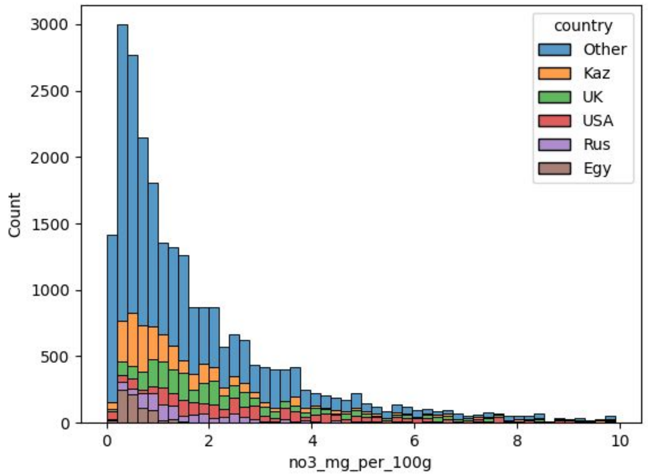
Harmonization in Practice

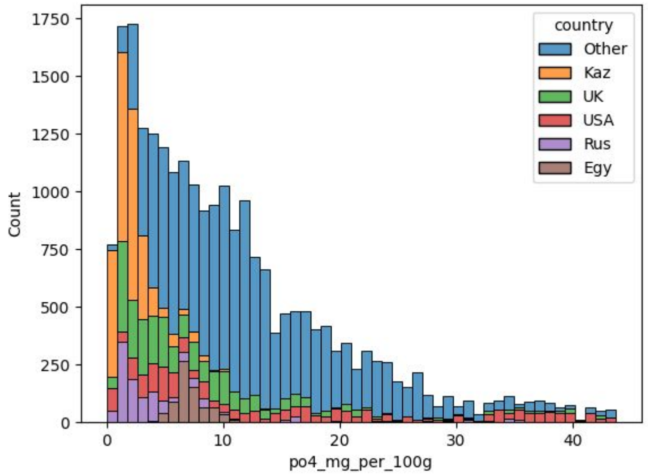
Harmonization is not straightforward!

- 1:1 correlations→ equivalent methods at global level
- complex mappings → extra source of uncertainty

Having DE lab data provides directly some global insights:

- California: highest share of high-N fields
  → lack of regulations?
- KAZ/RUS: peak of PO4 distribution at lower values
  → different P method recommendation systems?





Conclusion

Conclusion

- Building a soil DB is hard (field work, logistics) and expensive
- Working with many different labs is challenging (local standards)
- Lack of global standards make data harmonization difficult

→ hardware/technology is important but not most critical aspect for **scaling** a **global product** that is useful for **local applications**

→ fully **leveraging the potential of the data** requires to connect and relate it to **agronomic**, **political** and **economical** applications

Food and Agriculture Organization of the United Nations

THANK YOU

GLOBAL SOIL PARTNERSHIP