Improving the accuracy of K recommendations

effectiveness, utility?

S.M. Brouder, Purdue Univ.
T.S. Murrell, IPNI

What is an "accurate" recommendation?

Attempt at a definition:

An accurate recommendation is one that provides realistic estimates of costs and benefits, with associated levels of confidence, for a given K management option and a given set of conditions input by the user.

Note: we are not being accurate when we recommend one rate for a given set of conditions. A range of rates that fall into a prediction interval helps the farmer and the regulator understand that there is uncertainty in the rate estimate.

User want vs need

Credibility ("... a dark art." – Mike Bell, Rome, 1/26/2017)

Two Perspectives on Recommendations

Scientist perspective

- Theoretically proven in its components
- Sufficient research under representative conditions, environments, management scenarios
- Explicit in terms of precision / accuracy for a given "inference space"
- Complex enough to be meaningful, but user-interface is intuitive

**User (farmer, consultant, …) perspective**

- Credible: makes sense of what is being observed
- Key elements are visible, understandable, explainable
- Customizable
  - Local environment and management practices
  - Considers farm enterprise profitability, not just economics of one crop input
- Presents a set of options with associated confidence levels in the predicted outcomes
- Considers short- and long-term costs/benefits

Common ground? Data-driven solutions….

Have scientists under-utilized the K cycle in recommendation development...?

What do we consider in developing a K recommendation?

- K offtake in harvested portions

- K uptake requirements

- Estimate of phyto-available K in the soil
  - Soil testing
  - Plant uptake under unfertilized conditions

- Recovery efficiency of applied K (implicit)

- Partial mass balance (application vs. offtake)
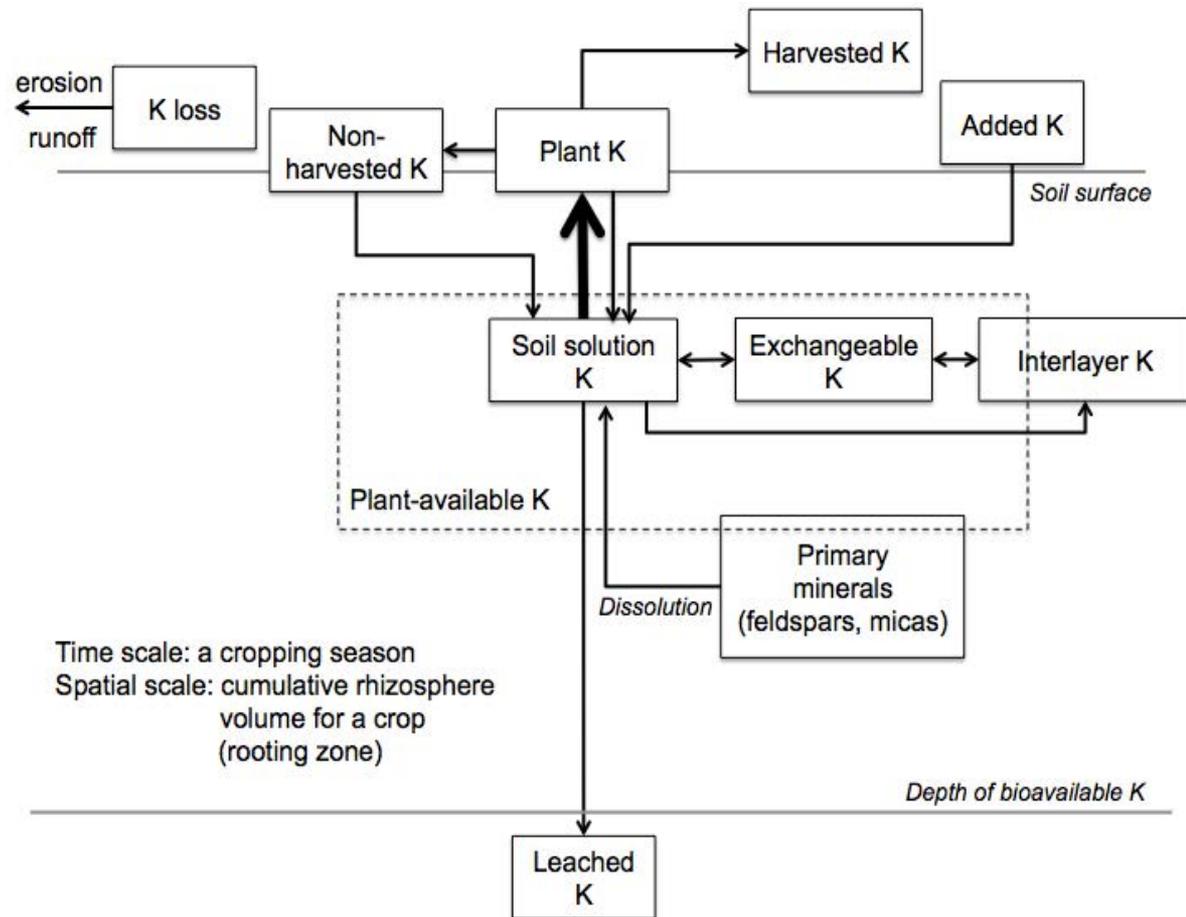
Factors Not Considered in Most K Recommendations

- Change in K availability with manure applications
- Impacts of interlayer-K and primary minerals
- Leaching and runoff losses
- Uplift from plant cycling
- System benefits, such as water use efficiency, salt tolerance, more efficient use of nitrogen, etc.
- Different plants' abilities to access interlayer-K and K from primary minerals
- Role of microorganisms (mycorrhizae) in increasing plant access to soil K

Existing recommendations are data-driven but are they credible with the end-user?

Problem: Model (over?)fitting... seeking the perfect calibration curve with sparse data
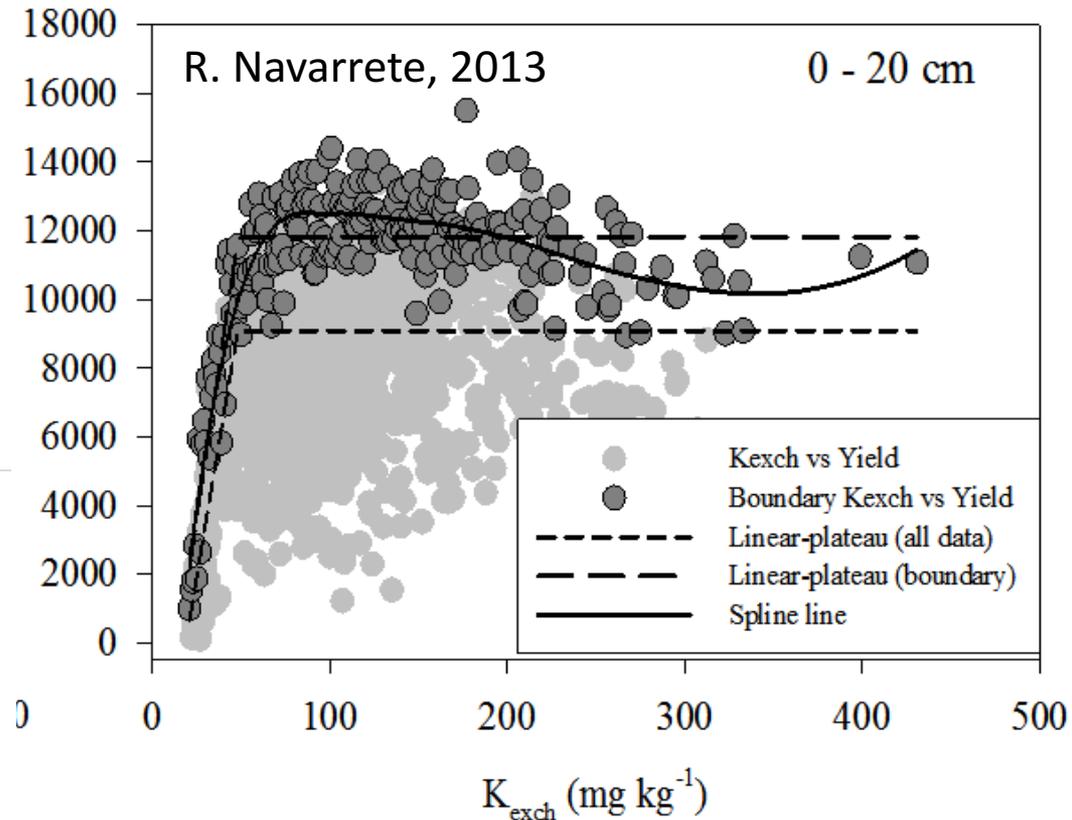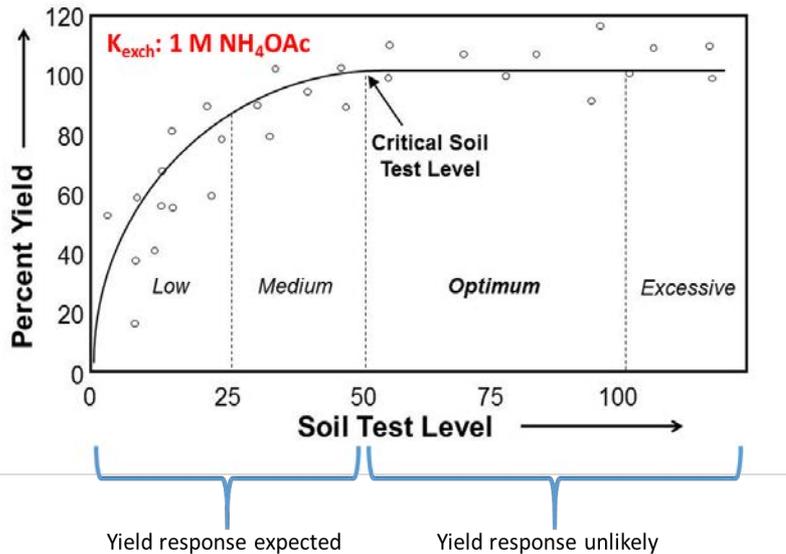
Existing recommendations are data-driven but are they credible with the end-user?

Problem: Not transparent to underlying data &/or incorrect quantitative representation of process(es)

R. Navarrete, 2013
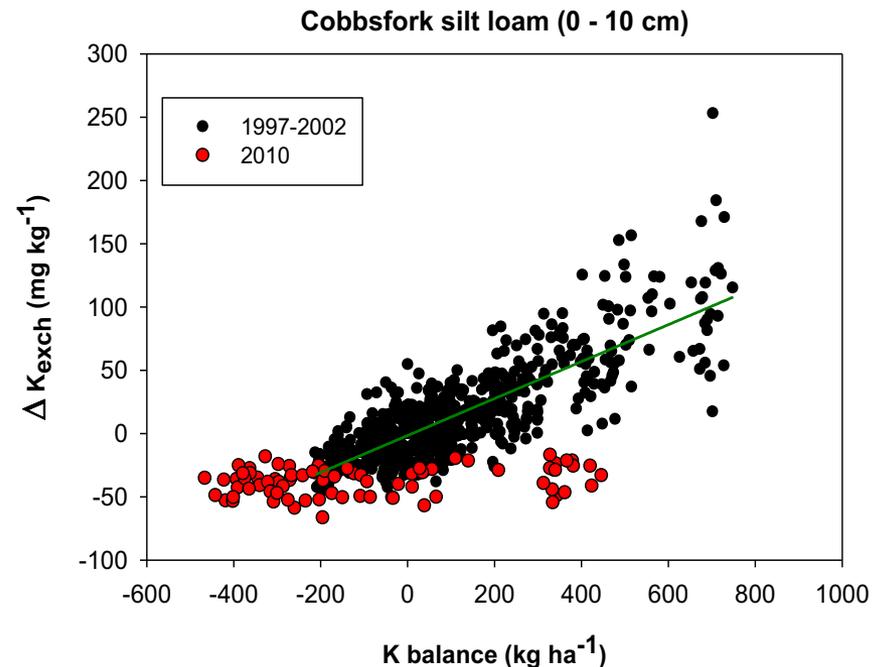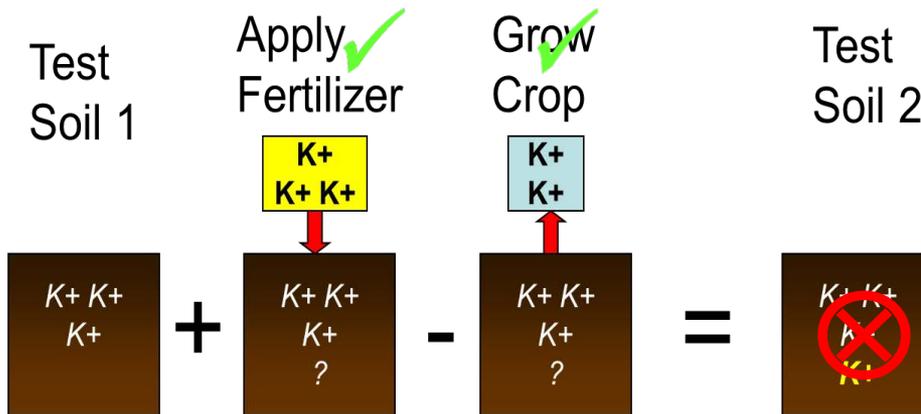
The outcome of K Frontiers 2015:
Improved accuracy via better soil assessment

- Pilot project proposed for benchmark soils

- Improved recommendation accuracy via better soil assessment

- Focus: develop soil categories for:
  - K holding capacity
  - K supplying power
  - Temporal aspects of K fluxes to plants
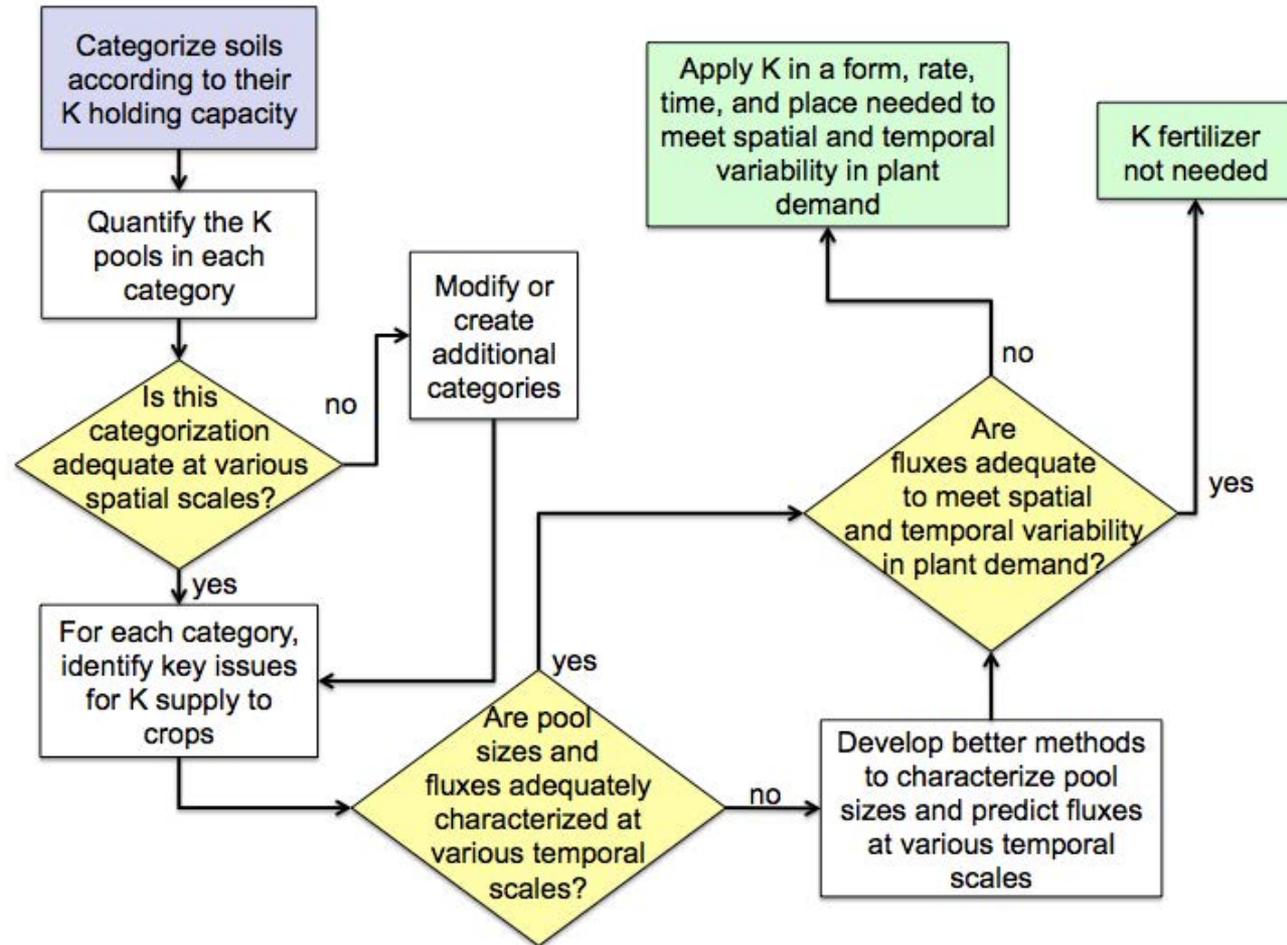
Question worthy of reflection: Too much "business as usual"

→ *Yes, if we proceed with*

- *Standard correlation-calibration protocols*

- *Disparate arrays of small studies*

- *"Over-fitted" models & nontransparent interpolation*

- *K cycle oversimplification*

→ *No, if we think broadly about*

- *Available data*

- *Proxy measures*

- *Plant K demand*

- *Open-source frameworks*

- *Machine learning*

- *Trans/Interdisciplinary collaborations*

Modeling Efforts Have Addressed Parts of the K Cycle

- Barber (U.S.):
  Movement of K in soil and root and shoot uptake

- Datta et al. (India):
  Modeling fixation and release, based on the concept of threshold levels

- OVERSEER (New Zealand):
  K balance, comparing K inputs with K outputs in pasture systems

- PROFILE (Sweden):
  Weathering rate of primary K minerals

- QUEFTS (Netherlands):
  Considers simultaneous impacts of N, P, and K soil supplies on nutrient uptake and yield

- EPIC (Brazil):
  Uses a simplified K cycle to simulate crop growth, K uptake, and yield

- DSSAT:
  Models plant biomass and K accumulation, K management

- MANNER-NPK
  Estimates available K from manure

Big Data Efforts for K Soil Fertility

- Better Fertilizer Decisions for Cropping Systems in Australia – ongoing
    - Customizable soil test calibration relationships
- IRRI's Reversing Trends of Declining Productivity in Intensive Irrigated Rice Systems (RTDP)
    - Large database of indigenous soil K supplies, measured from on-farm omission plots

Soil test-crop response trials

The database holds 5760 trial treatment series undertaken at 1227 distinguishable geographic locations, many being nearest town. The treatment series are collated from 1829 N, 2388 P, 365 K and 287 S trials.

Searching the database

Trial sites are plotted on the map as grey dots. Make a selection of trials based on the search criteria below and/or by drawing a polygon on the map around your region of interest. Always begin with a broad selection, then narrow the criteria to search the selection in more detail.

Nutrient:	K	Farming System:	All
From Year:	All	To Year:	All
State:	All	Season:	All

Crop:

All
cereal barley
cereal barley feed
cereal barley malting
cereal maize
cereal oats
cereal sorghum
cereal triticale

Australian Soil Class:

All
Calcarosol
Calcarosol calcic
Calcarosol hypercalcic
Calcarosol hypocalcic
Calcarosol lithocalcic
Calcarosol supracalcic
Chromosol

Select trials that satisfy the selection criteria above

Going "bigger" w/ data: Data from publicly-funded K research likely to remain sparse: What are other data sources?

Public Sector Data:

- Knit together existing large K databases (e.g. Australia BFDC, 4R Research Repository, etc.)

- Agronomic & breeding research on other topics by research & technical assistance entities (USDA-ARS online datasets)

- Public soil / tissue testing labs

- Emerging / Emerged soil mapping databases, "Yield Gap" projects, precipitation & soil moisture mapping initiatives
  - USDA SSURGO vs Africa Soil Information Service (AFSIS)
  - Global Yield Gap Atlas project accounts for H_2O deficit but crop models need specification for K

- Requires more "intentional" collaboration

Going "bigger" w/ data: Data from publicly-funded K research likely to remain sparse: What are other data sources?

Private Sector Data:

- Farmers (not typically shared)
- On farm research collaboratives (public-private or private-private)
- Commodity groups
- Commercial soil & tissue testing labs (IPNI Soil Test Summary North America)
- Industry & industry researchers

"Intentional" collaboration must address trust & privacy concerns, proprietary interests & ownership, & accessibility, governance & cost issues

Improving K Recommendations: Actively Borrow from Emerging Disciplines

- Past advancements have borrowed concepts from physics, chemistry, crop physiology, microbiology, mathematics, statistics, etc.
- Now and in the future, we must borrow concepts from emerging disciplines and disciplines we have previously paid little attention to:
  - Library science
  - Data & information sciences
  - Citizen science
  - Open-source software
  - Artificial intelligence
  - Non-traditional statistical methods
  - Many others

The Framework: 10 steps to real-time data uptake, analysis & customized recommendations

1. User enters data via web portal
2. Portal has imbedded workflows for ease of use & auto quality assurance/quality control (QA/QC)
3. Data anonymized at entry according to mutually acceptable terms & conditions
4. User data combined with existing research data
5. Data archived and preserved in a "trusted" repository

Infrastructure for ingesting & organizing data in real time

The Framework: 10 steps to real-time data uptake, analysis & customized recommendations

6. Automatic reanalysis w/ accruing data
7. Machine learning strategies to minimize human resources

8. Combination analytical strategies that are directed by scientist using proven theories & data mining ("unsupervised") strategies to surface overlooked linkages, drivers & proxy measures

9. Tools for "unpacking" the analytical result to explore new/unexpected results & discoveries

The Framework: 10 steps to real-time data uptake, analysis & customized recommendations

10. Customized, credible, K recommendation
A. Self-improving
B. References the users data
C. Can be modified for non agronomic priorities (risk consideration, time horizons, etc.)

Why artificial intelligence (AI) and machine learning (ML) for recommendation development?

Machines learn in the process of transforming data into actionable knowledge (Lantz, 2013)

Data mining requires machine learning

Can use machine learning for analyses that are not data mining…

Analytical Framework

ARTIFICIAL INTELLIGENCE
Let the games begin..

Dredging vs mining…?

Brett Lantz ~ Machine learning with R (open source)

Process (humans and machines) ~ 3 components

- **Data input**: observation, memory storage and recall used to provide "factual basis" for further learning

- **Abstraction**: translates data into broader representations

- **Generalization**: abstracted data used to form basis for action

- Machine learning algorithms use best guesses (heuristics) to select most important concepts
    - People do this
    - Both people & machines can make erroneous conclusions…
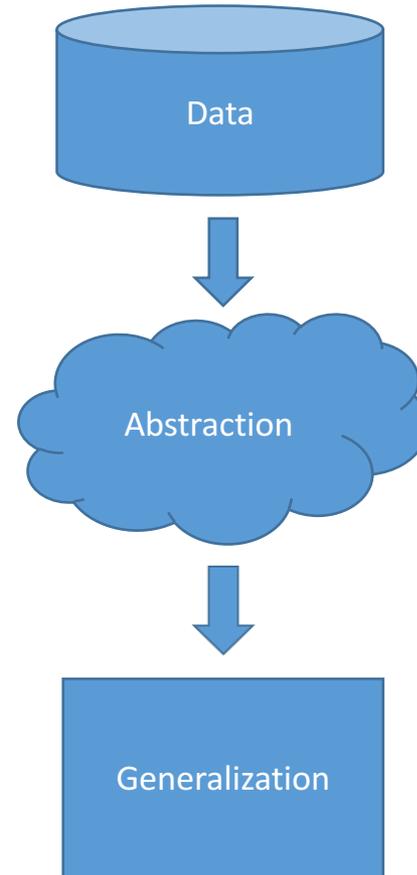
A flexible, useful, open-source framework

*Thinking about the how vs. the what of a recommendation…*

K Frontiers 2015 surfaced considering / implementing categorical analyses & decision trees ~ examples of potential binary decisions:

- Y/N: leaching routinely reduces plant-available K
- High/Low: CEC value
- Y/N: dominant / competitive cations
- Y/N: K fixing soil
- Y/N: plant can substitute Na for some K
- Y/N: crop for human consumption

K means clustering & binary logistic regression for decision trees

Alfalfa: Clustered into yield categories to try to understand P/K interactions in tissue tests

- Darkened symbols = acceptable yields

- Open symbols = yields reduced by P or K

- BLR relationships shown are based on analysis of May 2004 & June 2004 data
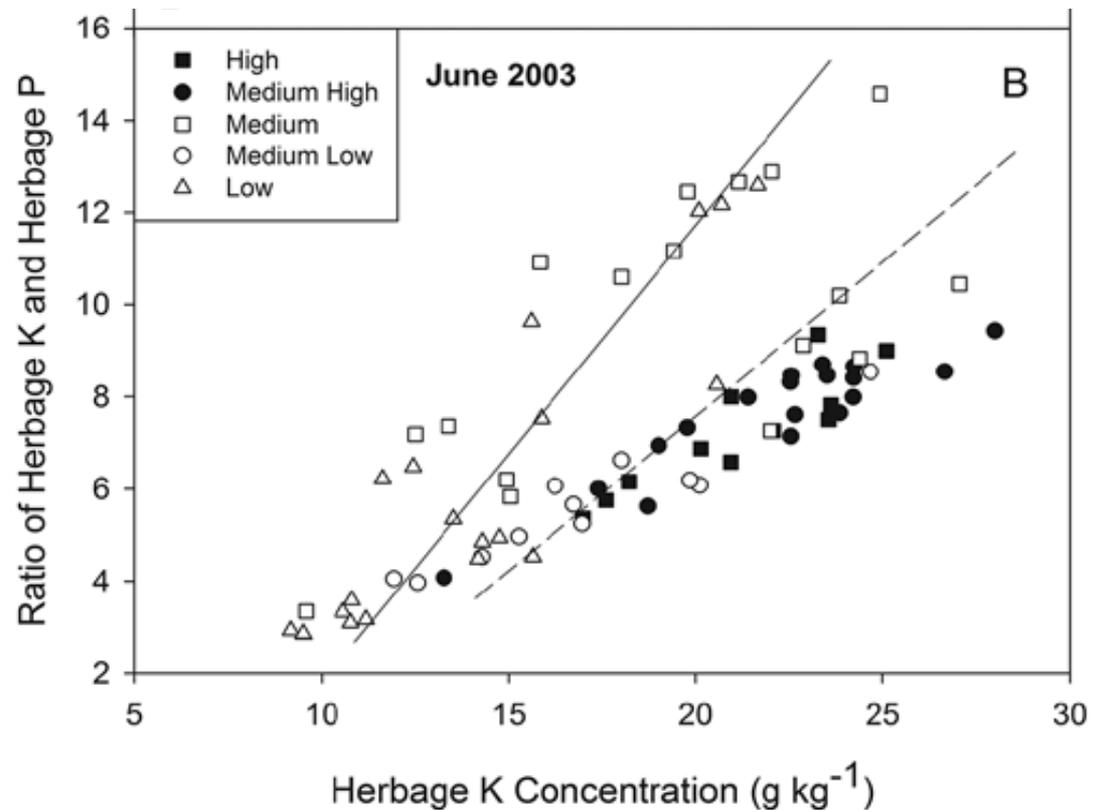
- Previous ANOVA analyses could not untangle P/K interactions BUT
  - Berg et al., 2005, 2007: 105, 33 citations
  - Lissbrant et al., 2010: 4 citations

Novel stats in N studies of more interest:
Cluster analysis for yield management zones (2005): 35 citations
Diagnosis N deficiency/sufficiency/excess w/ grain N conc.: 36 citations

Compelling reasons for using K recs as a test case for the framework & public / private partnership approach

Necessity is the mother of invention

- Limited funds / current priorities of public research agendas make it unlikely that all important K "unknowns" will be comprehensively and rapidly addressed

Beneficial attributes of K compared to other nutrients

- K is important to plant, animal, and human nutrition

- No known environmental problems: Inadvertent exposure of users to regulatory consequences is a low to non-existent risk
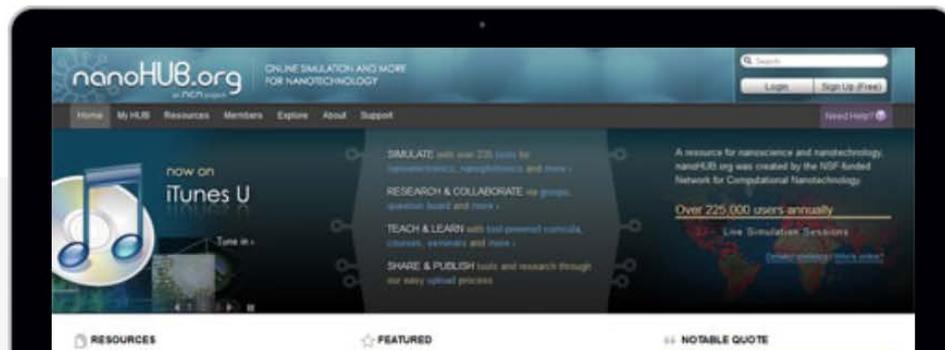
A Logical Start:
Collect and Centralize What We Already Have

https://hubzero.org/

HUBzero® is a powerful, open source software platform for creating dynamic web sites that support scientific research and educational activities

nanoHUB.org

Granddaddy of all hubs focused on

https://hubzero.org/getstarted

nanoHUB.org

ONLINE SIMULATION AND MORE
FOR NANOTECHNOLOGY

Home   My HUB   Resources   Members   Explore   About   Support

SIMULATE with over 220 tools for nanoelectronics, nanophotonics and more ›

RESEARCH & COLLABORATE via groups, question board and more ›

TEACH & LEARN with tool-powered curricula, courses, seminars and more ›

SHARE & PUBLISH tools and research through our easy upload process

A resource for nanoscience and nanotechnology. nanoHUB.org was created by the NSF-funded Network for Computational Nanotechnology.

Over 225,000 users annually

Live Simulation Sessions

RESOURCES   FEATURED   NOTABLE QUOTE

A resource for nanoscience nanotechnology, nanoHUB. created by the NSF-funded for Computational Nanotech
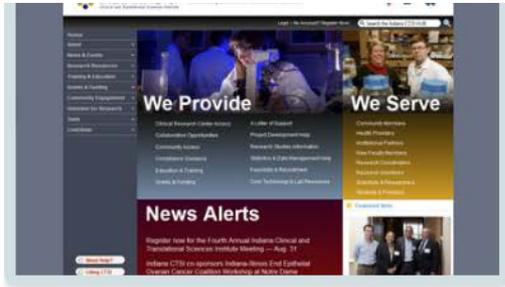
Hub communities….

IndianaCTSI.org

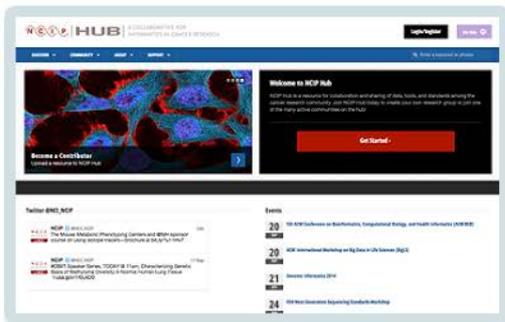accelerating clinical and
translational research in healthcare
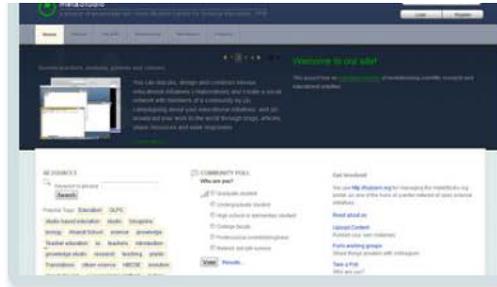
metastudio.org

Studio-based education initiative

MyGeoHUB

Community-driven geospatial data
modeling, analysis, visualization,
and publication

ncipHUB.org

A Collaborative for Informatics in
Cancer Research

pharmaHUB.org

pharmaceutical product
development and manufacturing

**Purdue University Research
Repository (PURR)**

Data management and dataset