Ecoinformatics for agricultural entomology: using data from farms to solve problems on farms

Jay A. Rosenheim

UC Davis

Collaborators:

- Cotton: Soroush Parsa, Billy Krimmel, Yao Hua Law, Frances Sivakoff, Michal Segoli, Moran Segoli, Tania Zaviezo, Rachel Goeriz, Emily Thacher, Kimberly Steinmann, Gail Langellotto, Andy Zink, Andrew Forbes, Peter Goodell, Jim Jones, Kevin Gross, Yves Carrière, Pierre Dutilleul, Matt Meisner

- Citrus: Beth Grafton-Cardwell, George Livingston, Lindsey Hack, Kim Steinmann, Bodil Cass, Karen Jetter, David Haviland, Craig Kallsen

Ecoinformatics:

- What is it?
- Why use it?
- Can it work?
  - Cotton pest management
  - Citrus pest management

Ecoinformatics: what is it?

1. Use of pre-existing data
2. Integration of data from multiple sources
3. Use of observational data
4. Large spatial and temporal scales
5. Large amounts of data
6. New tools for data management and analysis

Why use ecoinformatics?

Argument against: strengths of experimental methods:

Attribute	Experimental approaches	Ecoinformatics-based
1. Causal inference	**Stronger**	Weaker
2. Flexibility	**Higher**; any variable that the researcher can manipulate can be examined	Lower; only variables with pre-existing variation can be explored
3. Between-replicate	**Lower**, increasing statistical	Higher, decreasing statistical
4. Data uniformity, completeness, and perhaps quality	**Higher**; researcher has direct control of data collection	Lower; data collection is decentralized
5. Privacy concerns	**Lower**; data are collected by researchers away from the setting of the private farm	Higher; farmer willingness to share data may be variable

Why use ecoinformatics?

Argument for: strengths of ecoinformatics methods:

Attribute	Experimental approaches	Ecoinformatics-based approaches
1. Opportunity to integrate outreach with research	Lower; experiments are often conducted in off-farm settings	Higher; because data come from farmers, farmers are involved from the start
2. Study's spatial and temporal scale	Smaller; often much smaller than the scale of farming	Larger; matching the actual scale of farming
3. Applicability to the broad range of farming conditions	Lower; results may only apply to conditions under which the experiment was conducted	Higher; with suitable planning, data sets can embrace a large range of real farming conditions
4. Ability to evaluate many variables simultaneously	Lower; experiments are operationally difficult and costly for more than 4-5 variables at once	Higher; may be particularly valuable when many variables must be screened

Why use ecoinformatics?

Argument for: strengths of ecoinformatics methods:

Attribute	Experimental approaches	Ecoinformatics-based approaches
5. Ease of translating research results into farmer recommendations	Lower; researchers often use different sampling methods than farmers	**Greater**: using data from farmers means that research results translate naturally into recommendations
6. Ability to study farmer decision-making	Lower; farmers are typically excluded from the experimental research setting	**Higher**
7. Cost efficiency	Lower; labor costs of data collection are high	**Higher**; data can be mined inexpensively
8. Size of resulting data sets	Smaller	**Larger**; data sets may substantially larger, offering greater power

Do IPM experiments provide sufficient power?

Challenge: crops are valuable, insecticides are cheap

- Pima cotton is worth $2,000/acre
- An insecticide application may only cost $20/acre

So, farmers may be motivated to apply an insecticide even when only a small amount of yield is threatened

- *Lygus* on cotton: decision-point occurs when an insecticide protects 1% of yield

Can we measure such small yield effects?

- Literature survey: 27 yield-impact studies

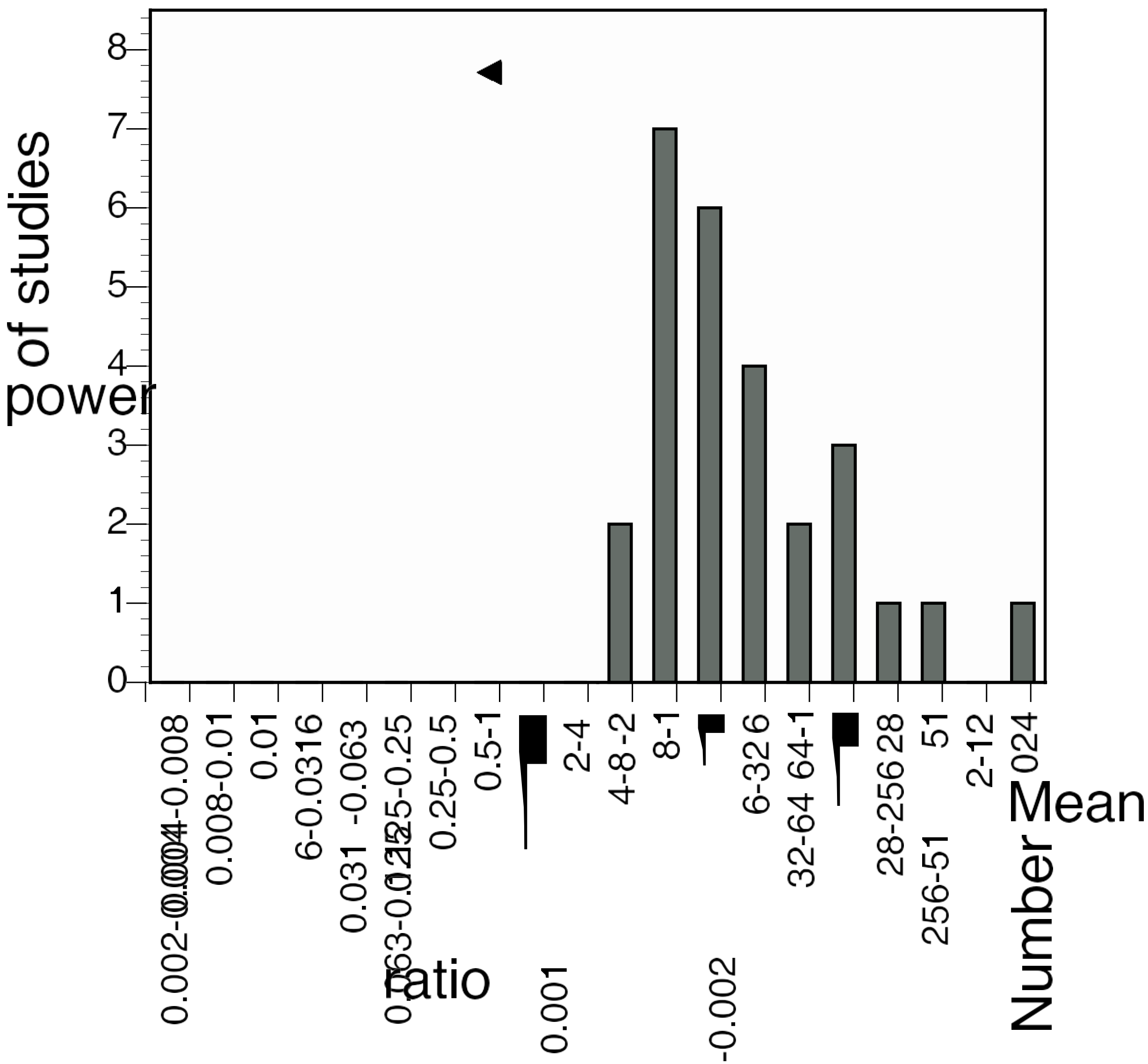
Do IPM experiments provide sufficient power?

- Collated crop value, cost of insecticide application, statistical power (variance, replication)

$$\text{Power ratio} = \frac{(\text{smallest proportional yield loss that can be detected with 50\% probability})}{(\text{smallest proportional yield loss that would motivate a farmer to spray})}$$

- Desired outcome: Power ratio $\leq$ 1.0

IPM experiments lack sufficient power:



. . . Consequence:

a disconnect between effects experimentalists can measure, and effects that matter to farmers

Ecoinformatics: can it work? Example 1: cotton

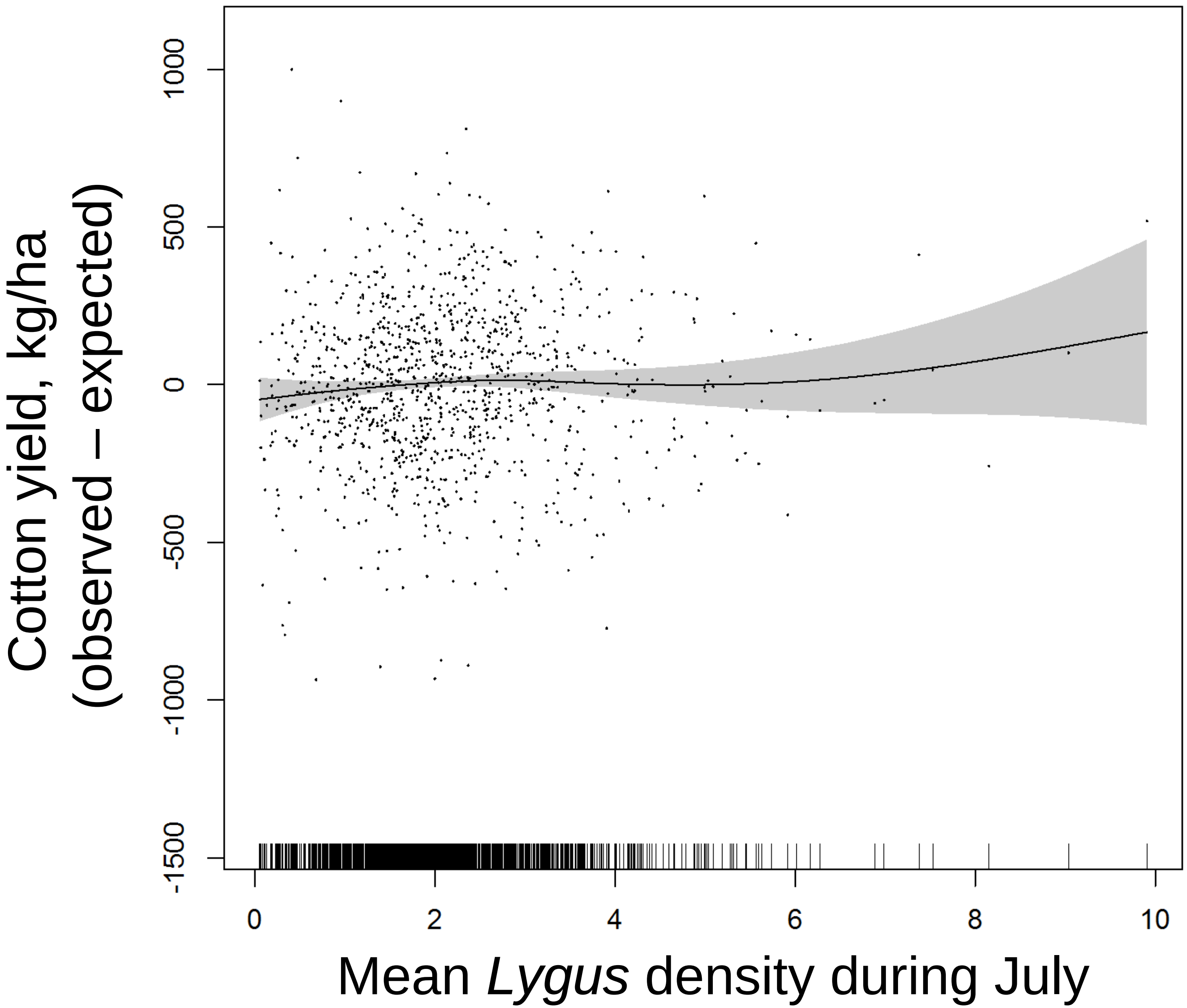
History

- Long controversy regarding at what densities *Lygus* depress cotton yield
- Many experimental studies (1968-present) conducted; none has shown significant yield effects at densities < 10 *Lygus*/sweep sample
- Nevertheless, growers are aggressive (treat at 3-4 *Lygus*/sweep sample)
- Experiments are *VERY* difficult to conduct (*Lygus* are mobile, often resistant)

Approach: large observational data set

1. *Many* factors influence yield; we knew a large data set would be critical
2. Building a database:
   - work with independent pest control consultants
   - all consultants sample *Lygus* the same way
   - Data streams: *Lygus* densities, crop yield, and as many supplementary variables as we could:
   - Larger data set: *N* = 1118 fields
3. Impact of Lygus on yield (mid-season, early-season)

Cotton compensates well for *Lygus* damage in July

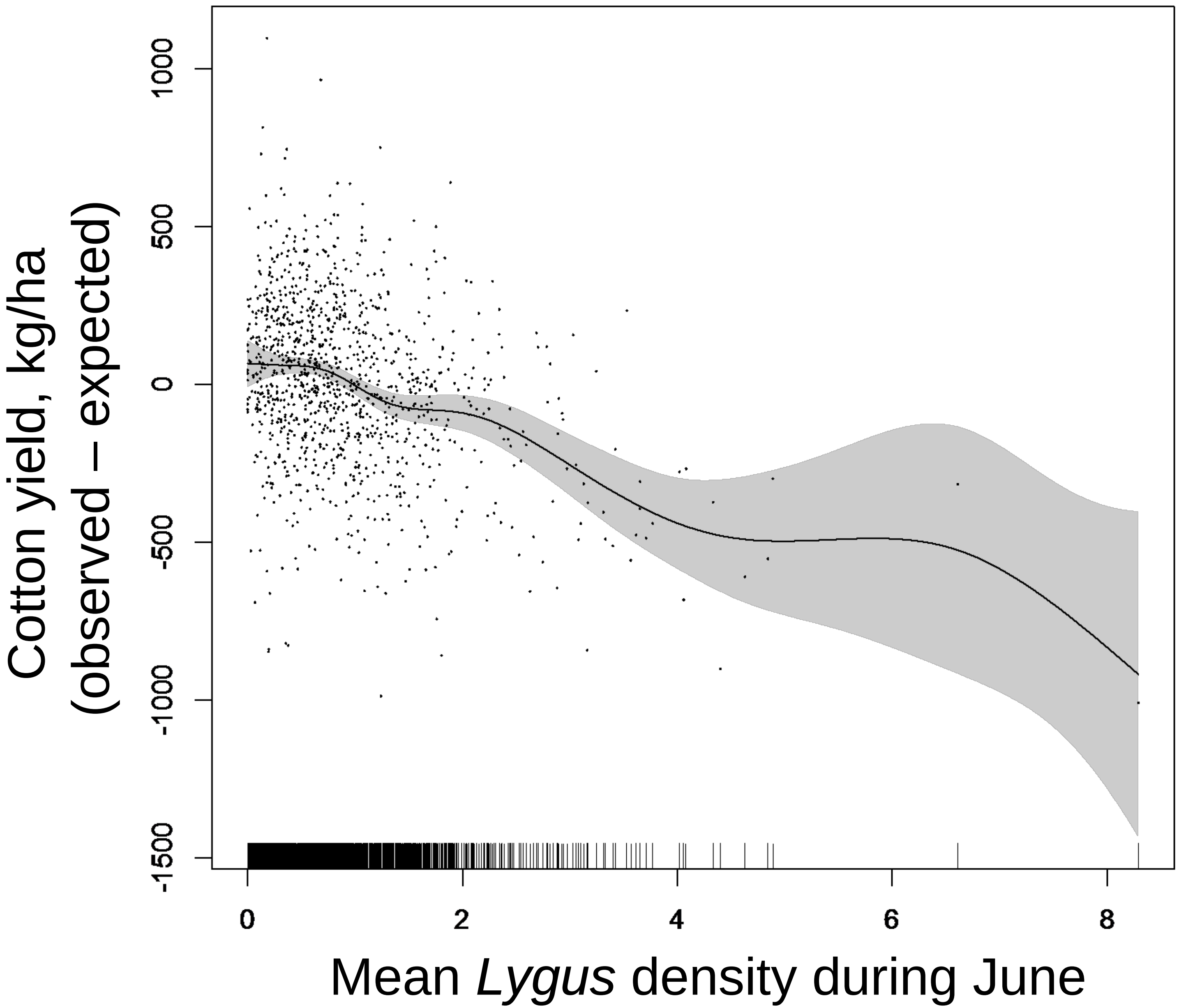


N = 1118
P = 0.51

→Farmers are over-using insecticides during July

Weakness: we can't explore effects of higher *Lygus* densities

Cotton compensates poorly for *Lygus* damage in June



N = 1118
P < 0.0001

→ *Farmers can increase yield by controlling early* Lygus

Can it work?

An 'ecoinformatics' data set resolved *Lygus*-cotton interactions that had been recalcitrant to experimentation

- farmers are losing yield (June herbivory)
- farmers are incurring needless costs (July insecticide applications)

The larger amount of data provided greater power (despite the greater 'noisiness')

Ecoinformatics: can it work? Example 2: citrus

History

- Proud tradition of IPM research on California citrus from UC Riverside
- IPM: use insecticides only if “economic injury level” has been exceeded
- Using experiments to define the economic injury level is *very difficult* for perennial crops; done for only 1 pest
- Most newly planted citrus acreage is in mandarins, for which we have no pest management research

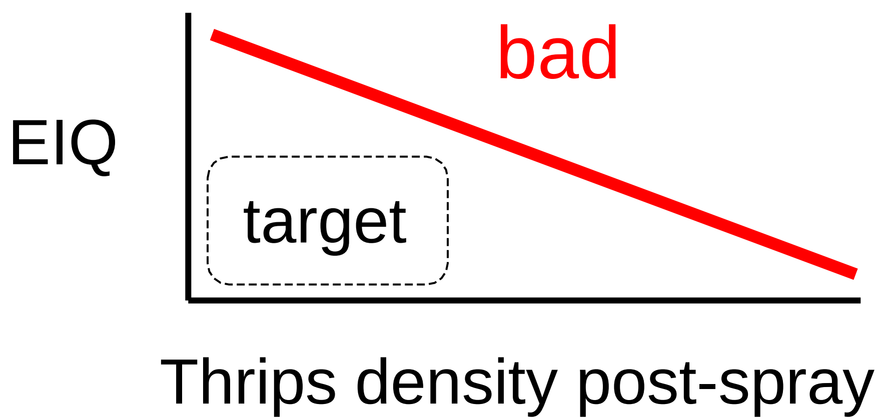
Citrus dataset

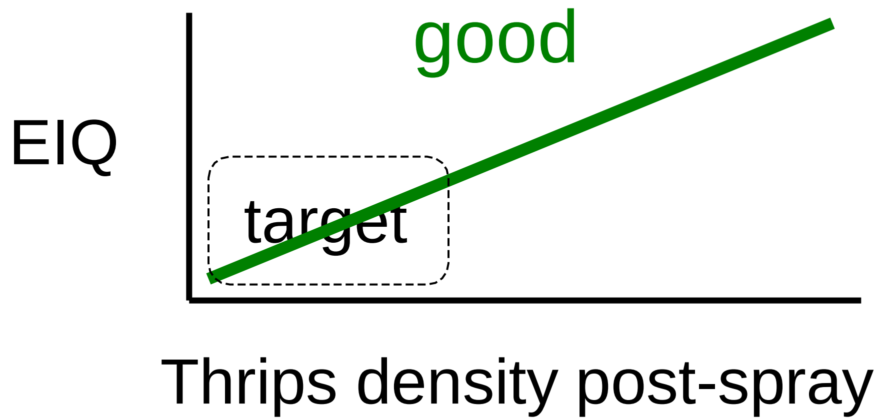
- With CDPR support, built a dataset for 1500 field•years, including:
  - pest densities
  - beneficial insect/mite densities
  - pesticide use
  - plant nutrient status
  - fruit damage/infestation
  - full harvest data
- today: early example, management of citrus thrips

Question 1: are more effective insecticides always worse for the environment?

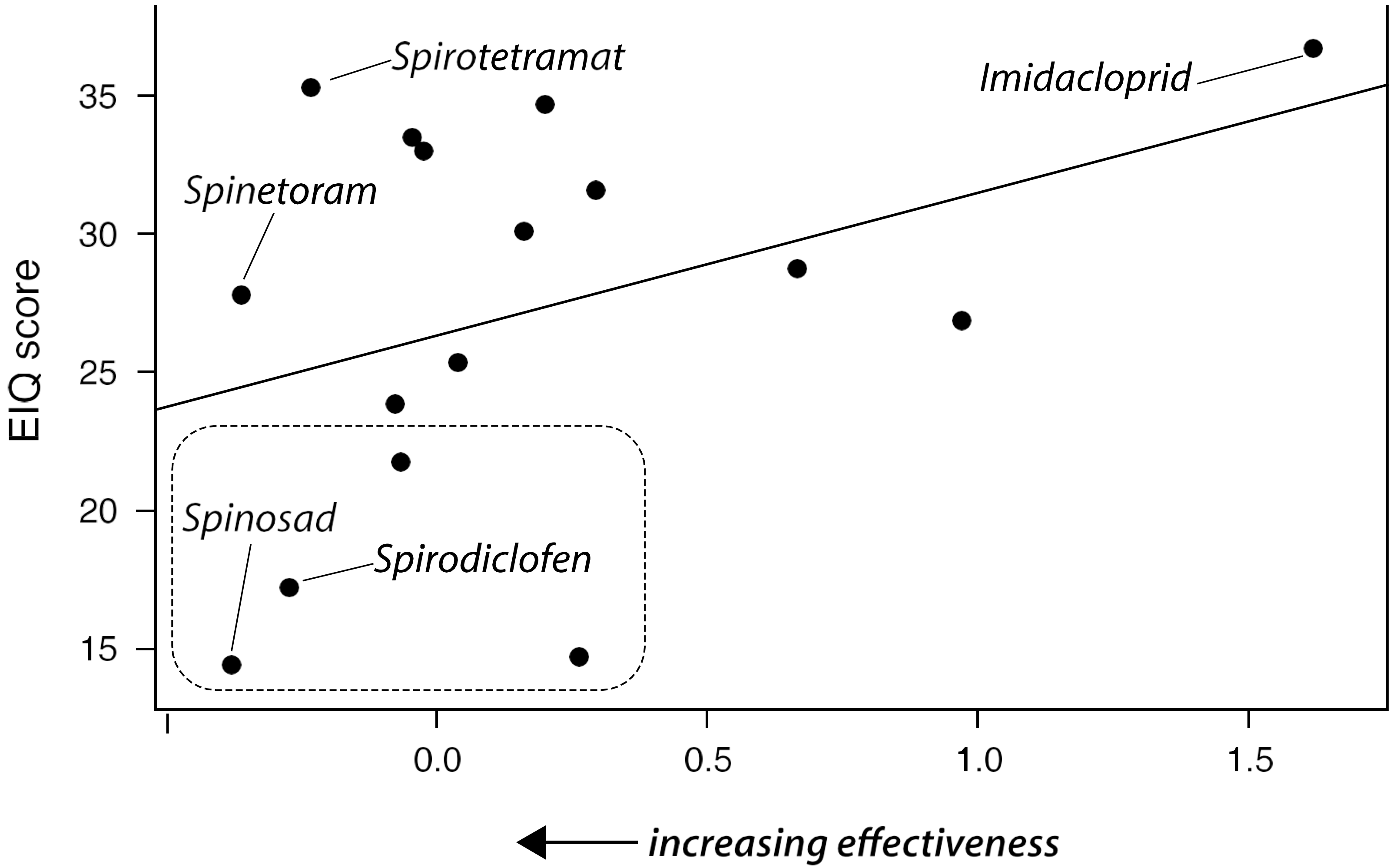
Environmental Impact Quotient (EIQ):

- toxicity to humans, birds, fish, beneficial insects
- leaching and surface runoff potential
- persistence in soil and plant surfaces

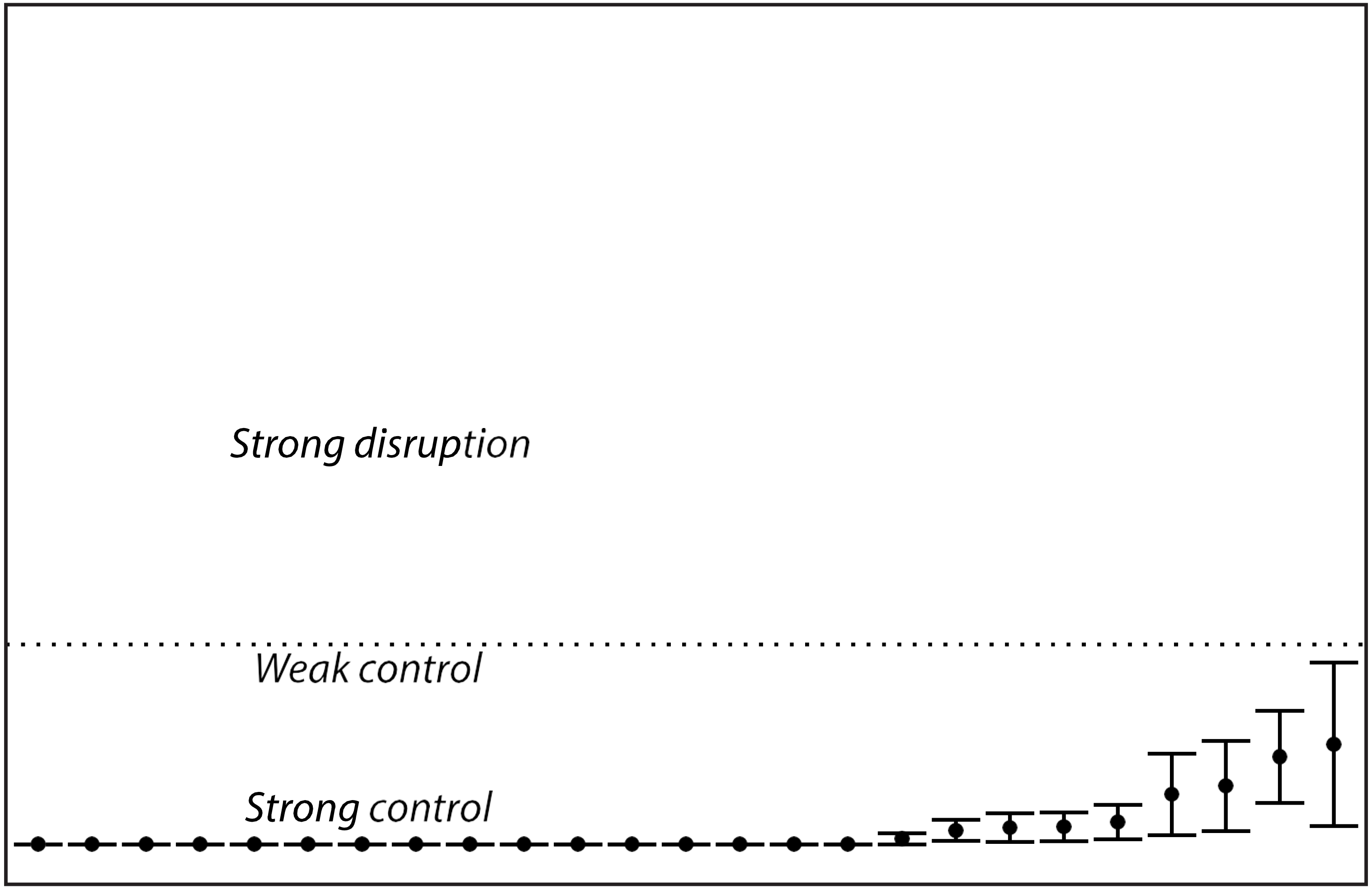




Some low EIQ compounds are highly effective



Some spatial variation in efficacy (Carzol)



Experimentation has both strengths and weaknesses

Observational approaches can complement traditional experimentation

Ecoinformatics can produce low-cost, large, and flexible datasets that can address many IPM questions

A more inclusive approach holds the promise of accelerating progress in agricultural entomology