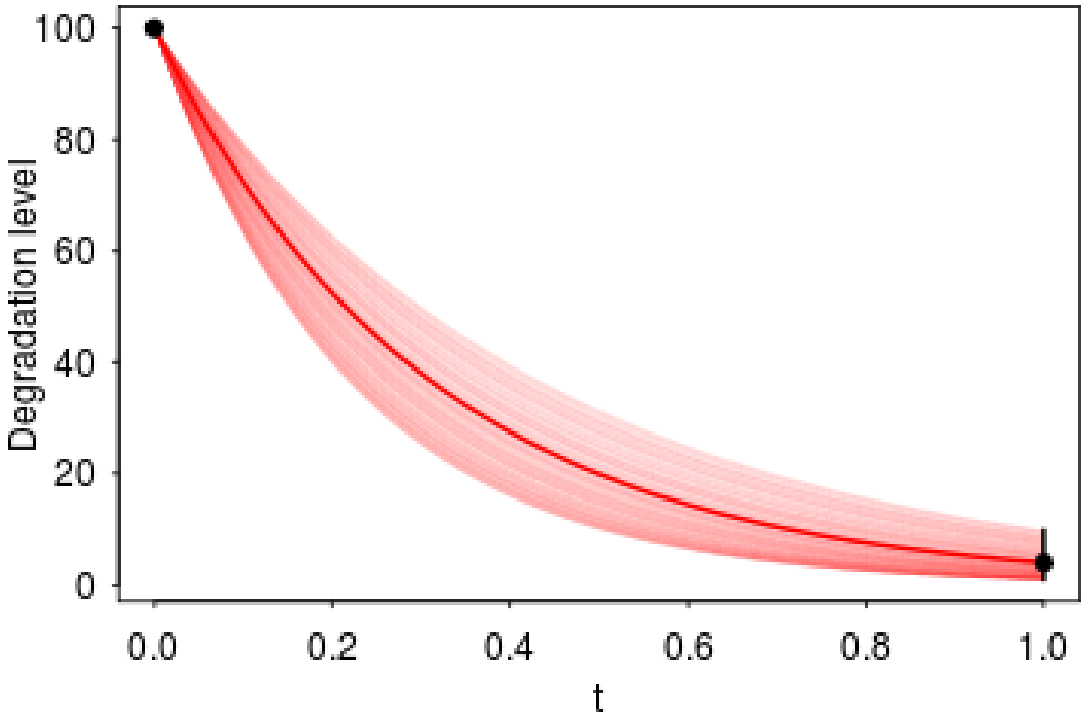
Statistics problems from Syngenta

Degradation of chemicals



Data: Sparse and dense measurements of degradation levels.

Questions:

- Characterise the error in the early measurements.
- When to measure? Experimental design problem.
- How to compare curves? Functional data.

Ranking of chemicals

Response: Kill score (0-100) for several chemicals from several experiments.

Explanatory:

- Dosage
- Plant species
- Application timing
- Application rate

Problem: Rank the chemicals.

Features: Missing data.

What do they mean by chem A is better than chem B?

Current approaches:

- Non-linear regression, rank by ED50. Each chem modelled separately.
- Elo ranking. Not order invariant (depends on screening order).

From chem space to chem ranking

Response: Kill score (0-100)

Explanatory:

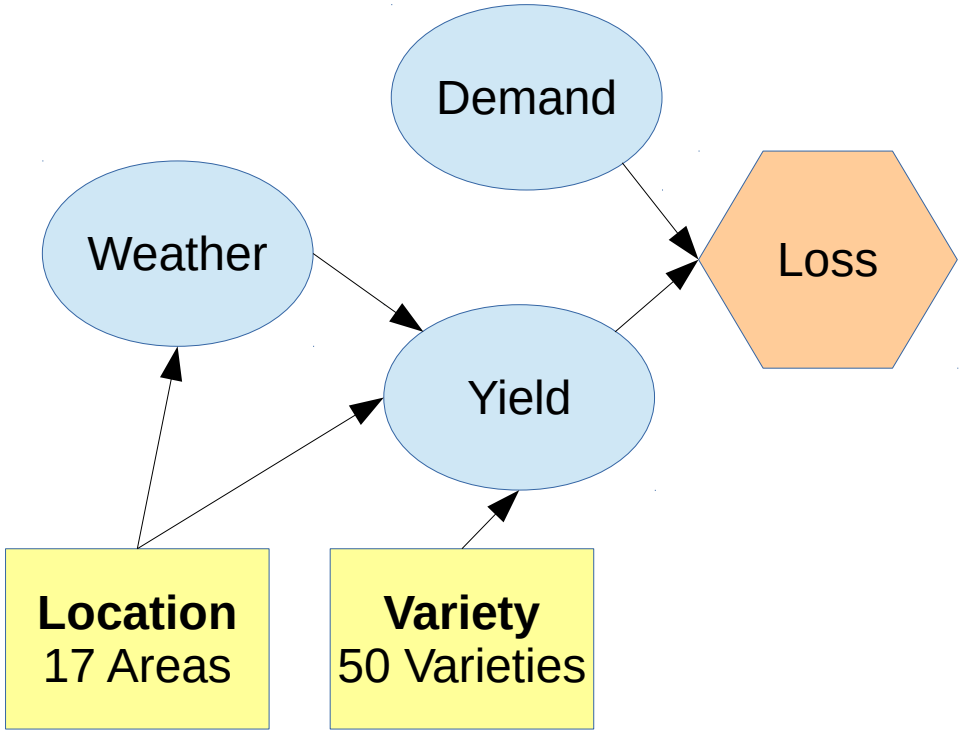
- Chem composition
- Everything in previous slide.

Question: Predict kill score (y) from composition ($\underline{x}$)

$$y = f(\underline{x}) + \varepsilon$$

Non-smooth fcn

Scheduling seed production



$$Yield: \quad Y=(Y_1,\ldots,Y_{50})$$
$$Demand: \quad D=(D_1,\ldots,D_{50})$$
$$Loss: \quad L(Y,D)=\|Y-D\|$$

Decision theory problem: minimise expected loss.

Questions:

- Model the whole system
- Conditional dist'ns?
- Loss fcn?

Formulation toxicity

Response: Compound toxicity (0-100 or categorised).

Explanatory variables: Chem composition.

Problem: Estimate individual toxicities. Interactions?

$$y = f(x,\theta) + \varepsilon$$