Cap Dev Lectures series:
Breeding autogamous cereals - a complete lecture from
*Parents to Farms*

Filippo M Bassi
senior durum wheat breeder

@fillobax
• Stage 1 and Stage 2 yield trials
  • Differences
  • Secrets of success
  • Logistics vs goals

• Target population of environments
  • What is a good station
  • Effect of climate on traits
  • New models to define TPE

• Stability analysis
  • What is GxE
  • How to control it
The breeder equation

Genetic gain_{year} = Accuracy \times Heritability \times Selection intensity

Accuracy: how close to the actual value

Heritability: how likely to move a trait from parents

Selection intensity: what proportion of the progenies is advanced each cycle

Recycling time in years: how many years between crossing and re-crossing
Yield trials: why?

- Why do we do yield trials?
Yield trials: why?

- Why do we do yield trials?

We need to test all traits declared in the product profile
Yield trials: common aspects

• What information would you give to describe a yield trial?
Yield trials: common aspects

• What information would you give to describe a yield trial?

Size, plot size, entries, experimental design, location, ..
Yield trials: Stage 1 vs Stage 2

- Is there any difference between Stage 1 and Stage 2?
Yield trials: common aspects

- Is there any difference between Stage 1 and Stage 2

The main difference is seed availability and number of progenies.
It is then critical to conduct a Stage 1 that enables Stage 2.

```
Single plant
  F5 or F4 rows
  multiplication (no/little selection)
  Stage 1
  1/few loc, small plot, not replicated
  Stage 2
  Many loc, large plot, replicated
```
Stage 1 yield trial: What is it?

<table>
<thead>
<tr>
<th>Stage</th>
<th>Entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crosses</td>
<td>100 pop</td>
</tr>
<tr>
<td>F1&gt;F4</td>
<td>100 pop</td>
</tr>
<tr>
<td>F5 (pedig.)</td>
<td>100 x 24 progenies</td>
</tr>
<tr>
<td>Stage 1</td>
<td>2 400</td>
</tr>
<tr>
<td>Stage 2</td>
<td>240</td>
</tr>
<tr>
<td>Stage 3</td>
<td>24</td>
</tr>
</tbody>
</table>

- Stage 1 is when most of **intensity** happens
- Small plot size
- No reps/loc
- Lots of “bad” genotypes
### Stage 2 yield trial: What is it?

<table>
<thead>
<tr>
<th>Stage</th>
<th>Entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crosses</td>
<td>100 pop</td>
</tr>
<tr>
<td>F1&gt;F4</td>
<td>100 pop</td>
</tr>
<tr>
<td>F5 (pedig.)</td>
<td>100 x 24 progenies</td>
</tr>
<tr>
<td>Stage 1</td>
<td>2 400</td>
</tr>
<tr>
<td>Stage 2</td>
<td>240</td>
</tr>
<tr>
<td>Stage 3</td>
<td>24</td>
</tr>
</tbody>
</table>

- Stage 2 is when most of **accuracy** happens
- Large plots
- Across environments / reps
- Mostly “valuable” genotypes

It is then an equilibrium between discarding in Stage 1 and assessing in Stage 2 with accuracy.
Stage 1 yield trial: secrets of success

Accuracy of selection

- **Type I error**
- **Type II error**

*top 10%*
Stage 1 yield trial: secrets of success

- Reach it with as much variation as possible
- Reach it with as much inbreeding as possible
- Avoid testing any “wrong” progeny
- Dedicate to it lots of your budget
- Avoid losing any useful lines before going to Stage 2
- Maximize intensity at decent accuracy and heritability
Stage 1 yield trial: secrets of success

- Reach it with as much *variation* as possible
- Reach it with as much *inbreeding* as possible
- Avoid testing any “wrong” *progeny*
- Dedicate to it lots of your budget
- Avoid *losing any useful lines before going to Stage 2*
- Maximize *intensity* at decent accuracy and heritability

![Accuracy of selection diagram](image-url)
Stage 1 yield trial: intensity

- 10% pressure is often used
- What would be the risk of raising it to 5%? Or increasing the pop to 4 800?

### Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder’s equation

Joshua N. Cobb, Roselyne U. Juma, Partha S. Biswas, Juan D. Arbelaez, Jessica Rutkoski, Gary Atlin, Tom Hagen, Michael Quinn, Eng Hwa Ng

<table>
<thead>
<tr>
<th>Ratio (%)</th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>STD differential</th>
<th>Genetic gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.0%</td>
<td>2 400</td>
<td>240</td>
<td>1.750</td>
<td>1.00</td>
</tr>
<tr>
<td>5.0%</td>
<td>2 400</td>
<td>120</td>
<td>2.063</td>
<td>1.18</td>
</tr>
<tr>
<td>1.0%</td>
<td>2 400</td>
<td>24</td>
<td>2.665</td>
<td>1.52</td>
</tr>
<tr>
<td>0.5%</td>
<td>2 400</td>
<td>12</td>
<td>2.892</td>
<td>1.65</td>
</tr>
<tr>
<td>0.1%</td>
<td>2 400</td>
<td>2-3</td>
<td>3.253</td>
<td>1.86</td>
</tr>
</tbody>
</table>
10% pressure is often used.

What would be the risk of raising it to 5%? Or increasing the pop to 4,800?

<table>
<thead>
<tr>
<th>Ratio (%)</th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>STD differential</th>
<th>Genetic gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.0%</td>
<td>2,400</td>
<td>240</td>
<td>1.750</td>
<td>1.00</td>
</tr>
<tr>
<td>5.0%</td>
<td>2,400</td>
<td>120</td>
<td>2.063</td>
<td>1.18</td>
</tr>
<tr>
<td>1.0%</td>
<td>2,400</td>
<td>24</td>
<td>2.665</td>
<td>1.52</td>
</tr>
<tr>
<td>0.5%</td>
<td>2,400</td>
<td>12</td>
<td>2.892</td>
<td>1.65</td>
</tr>
<tr>
<td>0.1%</td>
<td>2,400</td>
<td>2-3</td>
<td>3.253</td>
<td>1.86</td>
</tr>
</tbody>
</table>
Stage 2 yield trial: secrets of success

Accuracy of selection

Type I error

Type II error

R² = 0.1526
Stage 2 yield trial: secrets of success

- Most expensive step: it determines “numbers”
- Include only top genotypes
- Dedicate to it lots of your budget
- Select environments carefully

- Maximize accuracy and heritability at decent price
Stage 2 yield trial: secrets of success

- Most expensive step: it determines “numbers”
- Include only top genotypes
- Dedicate to it lots of your budget
- Select environments carefully

- Maximize **accuracy and heritability** at decent price

- Once lines are not advanced to Stage 3 they are out..
- Stage 2 are often the elites used as parents for crossing
Yield trials: decisions and logistics

- Which one would you like as home station?

<table>
<thead>
<tr>
<th></th>
<th>Env 1</th>
<th>Env 2</th>
<th>Env 3</th>
<th>Env 4</th>
<th>Env 5</th>
<th>Env 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance (Km)</td>
<td>5</td>
<td>20</td>
<td>200</td>
<td>300</td>
<td>1000</td>
<td>2500</td>
</tr>
<tr>
<td>Cost</td>
<td>High</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>Yield</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Disease</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Quality</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Drought tol</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>TPE</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>C</td>
</tr>
</tbody>
</table>
Yield trials: decisions and logistics

- Which one would you like as home station?
- Where would you do yield trials for TPE B?

<table>
<thead>
<tr>
<th></th>
<th>Env 1</th>
<th>Env 2</th>
<th>Env 3</th>
<th>Env 4</th>
<th>Env 5</th>
<th>Env 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance (Km)</td>
<td>5</td>
<td>20</td>
<td>200</td>
<td>300</td>
<td>1000</td>
<td>2500</td>
</tr>
<tr>
<td>Cost</td>
<td>High</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>Yield</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Disease</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Quality</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Drought tol</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>TPE</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>C</td>
</tr>
</tbody>
</table>
Yield trials: decisions and logistics

<table>
<thead>
<tr>
<th></th>
<th>Env 1</th>
<th>Env 2</th>
<th>Env 3</th>
<th>Env 4</th>
<th>Env 5</th>
<th>Env 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance (Km)</td>
<td>5</td>
<td>20</td>
<td>200</td>
<td>300</td>
<td>1000</td>
<td>2500</td>
</tr>
<tr>
<td>Cost</td>
<td>High</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>Yield</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>Disease</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Quality</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Drought tol</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>TPE</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>C</td>
</tr>
</tbody>
</table>

- Which one would you like as home station?
- Where would you do yield trials for TPE B?
- What is the value of Env 6?
Yield trials: decisions and logistics

- Which one would you like as home station?
- Where would you do yield trials for TPE B?
- What is the value of Env 6?

<table>
<thead>
<tr>
<th>Env 1</th>
<th>Env 2</th>
<th>Env 3</th>
<th>Env 4</th>
<th>Env 5</th>
<th>Env 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance (Km)</td>
<td>5</td>
<td>20</td>
<td>200</td>
<td>300</td>
<td>1000</td>
</tr>
<tr>
<td>Cost</td>
<td>High</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Yield</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Disease</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Quality</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Drought tol</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>TPE</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
</tbody>
</table>
Target population of Environments: TPE

- It defines how well a station represents an agro-ecology
- How would you define it?
Target population of Environments: TPE

- It defines how well a station represents an agro-ecology
- How would you define it?

Adaptation and stability analysis of ICARDA durum wheat elites across 18 countries
Bassi & Sanchez-Garcia. Crop Science (cover), 57:2419
Effect of climate on traits: climatic regressions

- There is an inherent issue in using traits to define environments, that will then be used to define traits
- Climatic factors are then better as they are “independent”
- Which climatic factors affect plant growth?
Effect of climate on traits: climatic regressions

- There is an inherent issue in using traits to define environments, that will then be used to define traits.
- Climatic factors are then better as they are “independent”
- Which climatic factors affect plant growth?

<table>
<thead>
<tr>
<th>Stage</th>
<th>Max temp</th>
<th>Min Temp</th>
<th>Moisture</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GY</td>
<td>DtH</td>
<td>Spk.m2</td>
</tr>
<tr>
<td>Overall</td>
<td>(0.81)</td>
<td>(0.40)</td>
<td>(0.32)</td>
</tr>
<tr>
<td>Pre-planting</td>
<td>0.41</td>
<td>0.42</td>
<td>0.69</td>
</tr>
<tr>
<td>Vegetative grow</td>
<td>(0.62)</td>
<td>(0.64)</td>
<td>(0.85)</td>
</tr>
<tr>
<td>Flowering</td>
<td>(0.44)</td>
<td>(0.22)</td>
<td>(0.38)</td>
</tr>
<tr>
<td>Grain filling</td>
<td>(0.76)</td>
<td>(0.26)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>Maturity</td>
<td>(0.59)</td>
<td>(0.23)</td>
<td>(0.26)</td>
</tr>
</tbody>
</table>
Defining TPE: climatic models

- Once the climatic boundaries of each station are defined (i.e. moisture, temperature, ..) these can be used to assess how much of the cultivation area that station represents.
Genotype x environment: what is it?

- It represents the rank variation of genotypes performances between environments
Genotype x environment: what is it?

- It represents the rank variation of genotypes performances between environments
- It is a factor that needs to be controlled, but that can also be exploited: how?
The secret factors of GxE

Performance of ‘Karim’

Two ways of controlling GxE:
1. Stability studies
2. Trait-based research
Stability analysis: principles and solutions

- Methods to assess stability?

**Finlay–Wilkinson regression**

The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis

Stability analysis: principles and solutions

• It is critical to have a quantification of “how stable” a line is vs another line
  • It is not always “the most stable” that we want

Finlay–Wilkinson regression

• This model provides:
  • $b'$ which is the angle of the slope (GxE)
  • $r^2$ response of genotypes to inputs ($\approx G$)
  • Among G25 and G45, which one is better?

The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis

Stability analysis: principles and solutions

- It is critical to have a quantification of “how stable” a line is vs another line
  - It is not always “the most stable” that we want

AMMI wide adaptation index (AWAI)

\[ AWAI = \sum_{i} S_i \times |PC_i| \]

- AWAI is a single value:
  - G2 has little GxE for PC1 and strong GxE for PC2
  - G9 has strong GxE for PC1 but little for PC2
  - Which one is more stable?
Controlling GxE: stability

- Pairing AWAI for GxE with BLUP for G ensures ideal selection

Adaptation and stability analysis of ICARDA durum wheat elites across 18 countries
Bassi & Sanchez-Garcia. Crop Science (cover), 57:2419
Genetic gain for G + GxE

- Since 2015 we have incorporated a selection index for stability (GxE) + yield potential (G)
- These graphs show it was possible to make a rapid strong gain for stability, while increasing yield potential
• The genetic gain equation drives all breeding decisions
• Stage 1 yield trials need to be designed to serve the exact purpose, keeping in mind logistics
• Stage 2 are vast investments that need to match the breeding goals (not one station just because I can)
• Defining TPE is fundamental to deliver on the set goals
• GxE and stability analysis are the bread and butter of breeders.. but not all models are useful
• Genetic gain for yield stability is possible