Principles of Genetic Coefficient Calibration

Concepts
How to obtain or determine genetic coefficients

• Querying modeler colleagues, literature, etc.
  – Heat units, seed size, growth habit
  – Pedigrees (related variety?)

• Calibrating to data
  – Measure specific traits
  – Fitting parameters to predict them
How to obtain or determine genetic coefficients?

• Querying modeler colleagues, literature, etc.
  – Heat units, seed size, growth habit
  – Pedigrees (related variety?)

• Calibrating to data
  – Measure specific traits
  – Fitting parameters to predict them
114. Wheat genetic coefficients

From: "Dr. Mohamed" - May 27, 2002

Dear List Members:

I am testing modifications to CERES-Wheat and found a very complete set of field data that could serve my purpose. But the experimental data corresponds to cultivars for which I do not have the genetic coefficients. The cultivars in which I am interested are:

- Norstar
- Brule
- TAM 101
- Colt

Does any one have these genetic coefficients and is willing to share them? Do you have any suggestion?

I appreciate your valuable reply,
## Published information:

![Western Beef Development Centre](image)

**Table 1. Silage Corn Varieties**

<table>
<thead>
<tr>
<th>VARIETY</th>
<th>HEAT UNITS</th>
<th>SEEDING RATE (seeds/acre)</th>
<th>AREA (acres)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Super Leafy Silage</td>
<td>2200</td>
<td>30,000</td>
<td>8</td>
</tr>
<tr>
<td>Hyland 2007</td>
<td>2100</td>
<td>30,000</td>
<td>2</td>
</tr>
<tr>
<td>Hyland 2428</td>
<td>2100 - 2250</td>
<td>30,000</td>
<td>2</td>
</tr>
<tr>
<td>Hyland 2093</td>
<td>2200 - 2350</td>
<td>30,000</td>
<td>2</td>
</tr>
<tr>
<td>Pioneer 39M27</td>
<td>2200 - 2350</td>
<td>30,000</td>
<td>2</td>
</tr>
<tr>
<td>Pioneer PB85</td>
<td>2175</td>
<td>30,000</td>
<td>2</td>
</tr>
<tr>
<td>Pioneer 39W54</td>
<td>2150</td>
<td>30,000</td>
<td>2</td>
</tr>
<tr>
<td>Canamaize</td>
<td>2100</td>
<td>60,000</td>
<td>7.5</td>
</tr>
<tr>
<td>RR OKC27-11</td>
<td>2250</td>
<td>30,000</td>
<td>30</td>
</tr>
</tbody>
</table>

---

Also, research journals

[www.wbdc.sk.ca/publications/factsheets](http://www.wbdc.sk.ca/publications/factsheets)
Online pedigree databases: e.g., GRIN

Similar cultivar? Ask. Varieties differ less than plant breeders will admit.
How to obtain or determine genetic coefficients?

• Querying modeler colleagues, literature, etc.
  – Heat units, seed size, growth habit
  – Pedigrees (related variety?)

• Calibrating to data
  – Measure specific traits
  – Fitting parameters to predict them

• Gene-based estimation
Measuring specific traits

• Environment Control Chamber Experiments
  – Control one or more environmental variables
    • Temperature, daylength, etc.
  – Isolate factors, e.g., temperature, photoperiod, etc.
  – Expensive, time consuming
  – Conditions often differ from those in experimental plots or farmers fields
    • Light intensity/spectrum; Relative humidity and dewpoint; Wind speed
    • Pot studies and root restrictions
    • Unreliable for biomass & yield (border-dependent!!!
How to obtain or determine genetic coefficients?

• Querying modeler colleagues, literature, etc.
  – Heat units, seed size, growth habit
  – Pedigrees (related variety?)

• Calibrating to data
  – Measure specific traits
  – Fitting parameters to predict them
Estimating Cultivar Coefficients Using Field Data

• Most commonly used method
• Referred to as “inverse modeling”, fitting coefficients, or “calibrating” in some cases
  – You have a model but do not know the cultivar coefficients
  – You have field data that carefully measured traits that are influenced by the unknown coefficients, and also measured model inputs
  – The model is solved many times, searching for the coefficient(s) that are the “best” coefficients or are most likely to be the correct ones
  – Objective criterion for determining “best-fitting” values
  – Measures of “fit”, similar to statistical regression
Various statistical methods to estimate parameters:
- Maximum Likelihood
- Ordinary Least Squares (RMSE)
- D-statistic (Wilmott)
- Using observations with different units or variances
- Frequentist (deterministic)
- Bayesian (GLUE=Generalized Likelihood Uncertainty Estimation)
- Coefficients, their variances and correlations

Measures of model fit and uncertainties

Practically, if data is too scarce, just go for the means: anthesis, maturity date, grain size, etc.
Fitting coefficients to field data

• Data needed (minimum data set):
  – Frequent observations of timing of vegetative and reproductive events
  – Growth analysis data
  – Final yield, yield components
  – Weather data (essential)
  – Quality of data is key to good results

• Multiple environments (locations, years, dates)
  – Yield trials (Plant Breeders)
  – Multiple planting dates are very useful
  – Extreme (non-commercial) dates or locations
    • Especially helpful to predict phenology
  – Amount of data is key to good results
Limited Data?
One Site/Year vs. Multiple Sites/Yrs

• Limited data (one site/year):
  – Various combinations of cultivar coefficients might give the same answer
  – Model simulations may not be accurate for other environments
  – The coefficients may not be very stable across multiple environments (year and/or sites)

• Using data from multiple years &/or sites
  – Coefficients are more robust
  – Model will better predict yield and other traits

Need to be close to site of application
Estimating Coefficients: Trial and Error

- Initial values from “best guesses”
- Phenology coefficients
  - Data are observations of anthesis, maturity, and other stages
  - Simulate available experiments, varying genetic coefficients (e.g., P1, P2, and P5 in maize)
  - Criteria for selection
    - Visual closeness of simulated to observed data
- Growth, partitioning, and yield coefficients (fertility bias?)
  - Yield, yield component data (also growth analysis data?)
  - Simulate available experiments many times, varying genetic coefficients (e.g., G2, and G5 in maize)
  - Criteria for selection
    - Visual closeness to simulated to observed data
General Approach

• We assume:
  – The basic model is correct
  – We do not know physiological characteristics of a specific variety, cultivar or hybrid.

• The problem:
  – Find the best combination of genetic coefficients to minimize error (e.g., RMSE) between simulated and observed target variable(s).

• With DSSAT, use GBUILD, plot Evaluate.out, look at mean, RMSE, and d-stat for anthesis, maturity, etc.
An Orderly Approach

• Is your cultivar similar to any on “model” list?
  – Spring wheat vs. winter wheat.
  – Early, medium, vs. late maturity maize types.
  – Maturity group in soybean (latitude!!)

• First things first:
  – Life cycle (the stages come in order).
    • Set thermal unit, vernalization or daylength coefficients for
      anthesis first
    • Then set coefficients affecting maturity.
  – Coefficients affecting vegetative growth
  – Then yield components.
  – Then yield itself.

Veg & Reprod growth GC are less reliable: SLPF/SOC
Example for Maize

- **Target Variable** – Anthesis
- **Unknown genetic coefficients:**
  - P1: Thermal time from emergence to end juvenile phase
  - P2: Delay due to photoperiod
- **Observations of planting date, anthesis date for n experiments (n = # years times # sites/year)**
- Provide an initial guess of P1 and P2
- Simulate anthesis date for n trials
- Compute RMSE
  \[ \sum (\text{simulated} - \text{observed})^2/N \]
- Vary P1 and P2 within the range that is known to exist for maize, and for each pair, simulate n trials and compute RMSE
- Select the (P1, P2) pair that minimizes RMSE

Caution: Need multiple sowing dates and latitudes to provide variation in daylength.
Example of RMSE for P2=0.450
‘Fitting’ Coefficients: Software-Assisted

- Software (Gencalc2) is in DSSAT v4.5 and ready for use (*fussy, needs care & thinking*)
- Centered on a ‘Rules’ file with information on target traits, controlling coefficients, number of simulations and step size for each coefficient.
- Other software such as gradient search or Monte Carlo methods are very useful and available (but not in DSSAT)
- GLUE: General Likelihood Universal Estimation tool is now available in DSSAT v4.5.
`Fitting` Coefficients: Gencalc2 with Winter Wheat

*ERRORSUMMARY: GENCALC2

<table>
<thead>
<tr>
<th>@</th>
<th>RUN</th>
<th>ENUM</th>
<th>ERROR</th>
<th>EABS</th>
<th>ERMS</th>
<th>P1V</th>
<th>P1D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>8</td>
<td>-4.00</td>
<td>4.000</td>
<td>4.000</td>
<td>54</td>
<td>63.0</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>-2.00</td>
<td>2.000</td>
<td>2.000</td>
<td>54</td>
<td>66.5</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>-1.00</td>
<td>1.000</td>
<td>1.000</td>
<td>54</td>
<td>70.0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>54</td>
<td>73.5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>2.000</td>
<td>2.000</td>
<td>2.000</td>
<td>54</td>
<td>77.0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>-3.00</td>
<td>3.000</td>
<td>3.000</td>
<td>57</td>
<td>63.0</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>-2.00</td>
<td>2.000</td>
<td>2.000</td>
<td>57</td>
<td>66.5</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>57</td>
<td>70.0</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>57</td>
<td>73.5</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>8</td>
<td>2.000</td>
<td>2.000</td>
<td>2.000</td>
<td>57</td>
<td>77.0</td>
<td></td>
</tr>
</tbody>
</table>
General Suggestions

• Estimate phenology genetic coefficients first
  – Maize/Rice: P1& P2 for anthesis, P5 for maturity
  – Wheat: P1V, P1D for anthesis
• Estimate growth and yield parameters afterward (i.e., G2, G3). G3 for seed size.
  – Criteria: Biomass, grain yield, harvest index, biomass, seed/grain size, etc.
• Using times series data for dry weights, leaf area, etc.,
  – Greater confidence is possible for coefficients
General Comments

• Field observations can have substantial errors
  – Sampling
  – Reporting
  – Stresses that are difficult to measure
• Models are uncertain & may not adequately describe processes
• Techniques available that take into account uncertainties of both observations & model
• Breeders, agronomists often can provide a reality check
Solving for Genetic Coefficients from Cultivar Trial Data: Soybean Example

Mavromatis et al. papers (Crop Science)

Yield Trials: Only had flowering date, maturity date, & yield.

To apply models, traits are needed as soon as varieties are released. Used optimization, with crop model and variety trial data on yield and maturity collected over multiple sites and years (n >15)
<table>
<thead>
<tr>
<th>Cultivar</th>
<th>MG</th>
<th>CSDL</th>
<th>FLPM</th>
<th>FLSD</th>
<th>SDPM</th>
<th>SDPM/FLPM</th>
<th>Lfmax</th>
<th>THrsh</th>
</tr>
</thead>
<tbody>
<tr>
<td>A94-674017</td>
<td>2</td>
<td>13.47</td>
<td>44.10</td>
<td>12.00</td>
<td>32.10</td>
<td>0.73</td>
<td>1.010</td>
<td>76.5</td>
</tr>
<tr>
<td>IA2021</td>
<td>2</td>
<td>13.52</td>
<td>42.50</td>
<td>9.94</td>
<td>32.56</td>
<td>0.77</td>
<td>0.930</td>
<td>74.5</td>
</tr>
<tr>
<td>SD93-522</td>
<td>2</td>
<td>13.72</td>
<td>45.70</td>
<td>14.07</td>
<td>31.63</td>
<td>0.69</td>
<td>0.950</td>
<td>75.0</td>
</tr>
<tr>
<td>A94-774021</td>
<td>3</td>
<td>13.56</td>
<td>45.60</td>
<td>14.10</td>
<td>31.50</td>
<td>0.69</td>
<td>1.110</td>
<td>79.0</td>
</tr>
<tr>
<td>Charleston</td>
<td>3</td>
<td>13.56</td>
<td>47.20</td>
<td>13.77</td>
<td>33.43</td>
<td>0.71</td>
<td>1.030</td>
<td>77.0</td>
</tr>
<tr>
<td>U94-3412</td>
<td>3</td>
<td>13.56</td>
<td>45.60</td>
<td>14.10</td>
<td>31.50</td>
<td>0.69</td>
<td>1.070</td>
<td>78.0</td>
</tr>
<tr>
<td>KS4694</td>
<td>4</td>
<td>13.25</td>
<td>47.90</td>
<td>15.32</td>
<td>32.58</td>
<td>0.68</td>
<td>1.010</td>
<td>76.5</td>
</tr>
<tr>
<td>Stressland</td>
<td>4</td>
<td>13.30</td>
<td>45.50</td>
<td>13.79</td>
<td>31.71</td>
<td>0.70</td>
<td>1.090</td>
<td>78.5</td>
</tr>
<tr>
<td>HC90-196</td>
<td>4</td>
<td>13.35</td>
<td>45.50</td>
<td>11.39</td>
<td>34.11</td>
<td>0.75</td>
<td>1.030</td>
<td>77.0</td>
</tr>
<tr>
<td>HS93-4118</td>
<td>4</td>
<td>13.35</td>
<td>45.50</td>
<td>9.79</td>
<td>35.71</td>
<td>0.78</td>
<td>1.050</td>
<td>77.5</td>
</tr>
<tr>
<td>LS92-3660</td>
<td>4</td>
<td>13.35</td>
<td>45.50</td>
<td>13.79</td>
<td>31.71</td>
<td>0.70</td>
<td>0.950</td>
<td>75.0</td>
</tr>
<tr>
<td>LS92-4173</td>
<td>4</td>
<td>13.25</td>
<td>47.10</td>
<td>16.67</td>
<td>30.43</td>
<td>0.65</td>
<td>1.010</td>
<td>76.5</td>
</tr>
<tr>
<td>Md92-5850</td>
<td>4</td>
<td>13.35</td>
<td>45.50</td>
<td>9.79</td>
<td>35.71</td>
<td>0.78</td>
<td>0.970</td>
<td>75.5</td>
</tr>
<tr>
<td>Ripley</td>
<td>4</td>
<td>13.35</td>
<td>46.30</td>
<td>14.03</td>
<td>32.27</td>
<td>0.70</td>
<td>0.950</td>
<td>75.0</td>
</tr>
<tr>
<td>SS91-7138</td>
<td>4</td>
<td>13.25</td>
<td>45.50</td>
<td>12.99</td>
<td>32.51</td>
<td>0.71</td>
<td>1.050</td>
<td>77.5</td>
</tr>
</tbody>
</table>
Genetic Coefficients for 10 Cultivars Solved from Georgia Variety Trials

<table>
<thead>
<tr>
<th>Cultivar - rank</th>
<th>CSDL (h)</th>
<th>FLPM (PTD)</th>
<th>SDPM /FLPM</th>
<th>SFDUR (PTD)</th>
<th>LFMAX (mg m^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perrin (10)</td>
<td>12.24</td>
<td>46.36</td>
<td>0.717</td>
<td>23.7</td>
<td>1.00</td>
</tr>
<tr>
<td>Stonewall (6)</td>
<td>12.31</td>
<td>41.27</td>
<td>0.767</td>
<td>24.8</td>
<td>1.03</td>
</tr>
<tr>
<td>Hagood (5)</td>
<td>12.23</td>
<td>42.27</td>
<td>0.775</td>
<td>25.2</td>
<td>1.03</td>
</tr>
<tr>
<td>Cook (3)</td>
<td>12.22</td>
<td>41.52</td>
<td>0.778</td>
<td>25.2</td>
<td>1.10</td>
</tr>
<tr>
<td>Thomas (9)</td>
<td>12.20</td>
<td>41.63</td>
<td>0.801</td>
<td>25.7</td>
<td>0.96</td>
</tr>
<tr>
<td>Colquitt (7)</td>
<td>12.26</td>
<td>41.52</td>
<td>0.754</td>
<td>24.4</td>
<td>1.05</td>
</tr>
<tr>
<td>Brim (1)</td>
<td>12.47</td>
<td>41.58</td>
<td>0.740</td>
<td>24.2</td>
<td>1.11</td>
</tr>
<tr>
<td>Young (4)</td>
<td>12.53</td>
<td>44.35</td>
<td>0.788</td>
<td>26.0</td>
<td>0.99</td>
</tr>
<tr>
<td>Hutcheson (2)</td>
<td>12.78</td>
<td>44.98</td>
<td>0.784</td>
<td>25.9</td>
<td>1.11</td>
</tr>
<tr>
<td>DP105 (8)</td>
<td>12.74</td>
<td>46.33</td>
<td>0.743</td>
<td>24.6</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Comparison of critical daylength sensitivity (CSDL) solved on 10 varieties, separately solving on data from NC or data from Georgia. Good predictability of CSDL and maturity over regions.

\[ y = 0.9672x + 0.4163 \]

\[ R^2 = 0.9593 \]
Conclusion: Stability of genetic coefficients solved from variety trials and value for use in other regions

- The maturity date and genetic yield potential can be predicted in a new region, based on traits solved in a different region, despite traits not being exactly the same.

- **Most of the yield variation is attributed to weather and soil factors** rather than varietal traits.

- Phenology traits solve more accurately than do yield-influencing traits.