Interpreting Historical Agricultural Data

James Hansen

International Research Institute for Climate Prediction
Crop Reporting Data

An Ideal Data Set

**Aggregate value and its components**

- Production
  - Yield
  - Area harvested
- Price

**Segregated by season and rainfed vs. irrigated**

**Other desirable characteristics**

- Quality
- Duration
- Spatial resolution
What can historic data tell me?

**Exploration** (What variables seem to be related?)

**Hypothesis testing** (Are these relationships real and consistent?)

**Prediction** (What range of outcomes can I expect?)

**Process model calibration** (How can I improve prediction of agricultural response?) and evaluation (How well can I predict agricultural response?)
Non-climatic influences on agricultural output include changes in:
- Technology
- Land use
- Soil quality
- Input use and production intensity in response to markets

Trends mask climatic impacts

Detrending generally the first step
Accounting for Trends

What functional form?

Maize, SE USA

Grain yield, Mg/ha

Year

Accounting for Trends

What functional form?

Maize, Jamaica

Grain yield, kg/ha

Year

Smoothing functions are generally better than parametric (e.g., linear) functions.
Accounting for Trends

Spectral Smoothing Filter

- 3 year
- 7 year
- 15 year
- 30 year
- linear
Accounting for Trends

Variability About Fitted Trends

**winter tomato, FL**

![Graph showing fresh fruit yield and fruit yield anomaly for winter tomato, FL](image)
Accounting for Trends

Variability About Fitted Trends

Maize, SE USA

Grain yield, Mg/ha

Yield anomaly, Mg/ha

Year

Maize, SE USA

Year
Accounting for Trends

Variability About Fitted Trends

Yield anomaly, Mg/ha

Maize, SE USA

Year

Percent grain yield increase

Maize, SE USA

Year

SD = 0.28
0.63
SD = 0.123
0.140
Exploration

Graphs to visualize possible relationships among variables. E.g.,

- Crop calendar with mean monthly climate
- Time series
- Scatter plots
- Spin plots for 3 dimensions

Ideally driven by some mechanistic understanding of interactions between large-scale climatic influences, local climate, and agricultural response

Can suggest but not confirm relationships
Test whether results of exploratory analysis are real and consistent

Partial basis for deciding whether to promote application of forecast information

Appropriate methods depend on the nature and number of variables of interest
## Hypothesis Testing

<table>
<thead>
<tr>
<th>Independent</th>
<th>Dependent</th>
<th>Well-behaved</th>
<th>Ill-behaved</th>
</tr>
</thead>
<tbody>
<tr>
<td>cont. single or mult.</td>
<td>cont. single</td>
<td>correlation, regression</td>
<td>rank correlation, nonlinear regression</td>
</tr>
<tr>
<td>cont. mult.</td>
<td>cont. mult.</td>
<td>CCA</td>
<td>transform?</td>
</tr>
<tr>
<td>discr. single</td>
<td>cont. single</td>
<td>ANOVA</td>
<td>Kruskal-Wallis</td>
</tr>
<tr>
<td>discr. mult.</td>
<td>cont. mult.</td>
<td>factorial ANOVA, MANOVA</td>
<td>transform?</td>
</tr>
<tr>
<td>discr. single</td>
<td>discr. single</td>
<td>contingency table (e.g., $P^2$)</td>
<td></td>
</tr>
</tbody>
</table>
Presentation (Discrete Predictand)

Time Series Plot

Grain yield, Mg/ha

Maize, SE USA
Presentation (Discrete Predictand)

Cumulative Distribution Plot

Maize, SE USA

Grain yield, Mg/ha

Cumulative probability

La Nina
neutral
El Nino
Understanding ENSO’s Impacts

Crop impact of strong El Niño (82-83, 97-98) opposite to weak-to-moderate El Niño events

Subsequent ENSO phase

Preceding ENSO phase

<table>
<thead>
<tr>
<th>Crop</th>
<th>Percent yield increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheat</td>
<td></td>
</tr>
<tr>
<td>Soybean</td>
<td></td>
</tr>
<tr>
<td>Cotton</td>
<td></td>
</tr>
<tr>
<td>Peanut</td>
<td></td>
</tr>
<tr>
<td>Tobacco</td>
<td></td>
</tr>
</tbody>
</table>

La Niña | neutral | normal | strong

---- El Niño ----

neutral normal strong

---- El Niño ----

neutral normal strong

neutral normal strong

neutral normal strong
Presentation (Discrete Predictand)

Box Plots

Yield anomaly (Mg ha⁻¹)

ENSO phase

La Niña  neutral  El Niño

Tomato

Bell pepper

Sweet corn

Snap bean
Understanding ENSO’s Impacts

El Niño increases winter vegetable prices

Price anomaly ($ kg⁻¹)

ENSO phase

Bell pepper

Snap bean
Understanding ENSO’s Impacts

La Niña increases FL winter tomato value
The Multiplicity Problem

**Bonferroni Corrections**

Let

\[ E / \text{experimentwise (or family) error rate} \]
\[ " / \text{comparisonwise error rate} \]
\[ < n = \text{number of comparisons} \]

For \( n \) comparisons at given \( " \), upper limit of experimentwise error rate is

\[ E = 1 - (1 - " )^{(n-1)} \]

Comparisonwise error rate needed to assure given experimentwise error rate \( E \) is

\[ " = 1 - (1 - E)^{1/(n-1)} \]
Example, correlation of 6 crops with seasonal SSTs at 13 leads/lags

| Scenario                      | $E$ | $E = 0.05$ | $E = 0.10$ | $|r_{\text{min}}|$ |
|-------------------------------|-----|------------|------------|-----------------|
| 1 crop × 1 lag                | 0.05| 0.0500     | 0.344      |
| 6 crops, $E_{\text{crop}} = E_{\text{crop}}$ | 0.23| 0.0210     | 0.401      |
| 1 crop × 13 lags              | 0.46| 0.0087     | 0.477      |
| 6 crops × 13 lags             | 0.98| 0.0014     | 0.547      |
The Multiplicity Problem

Options

- Discipline – limit number of variables considered
- Statistical methods appropriate for testing experimentwise error
- Pre-filtering with, e.g., PCA
- Bonferroni correction