Intercomparison of statistical models for projecting winter oilseed rape yield in Europe under climate change

Behzad Sharif, David Makowski, Kurt Christian Kersebaum, Mirek Trnka, Kirsten Schelde, Jørgen Eivind Olesen
Outline

• Introduction
• Materials & Methods
• Results
  • Prediction power
  • Inference power
  • Uncertainty
• Conclusions
Some challenges in crop modelling

- Formualted and fitted to the same data
- Based on the past
- The true model is unknown
- Model uncertainties
- Ensemble models
- Pest & disease
# Process Based Models vs. Statistical models in projecting future yields

<table>
<thead>
<tr>
<th>Process based models</th>
<th>Statistical models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Include several modules</td>
<td>All-in-one</td>
</tr>
<tr>
<td>Dynamic</td>
<td>Static</td>
</tr>
<tr>
<td>Based on several valuable studies</td>
<td>Empirical and difficult to interpret without prior knowledge</td>
</tr>
<tr>
<td>Require calibration</td>
<td>Easier to use</td>
</tr>
<tr>
<td>Complicated</td>
<td>Easily understandable</td>
</tr>
<tr>
<td>Uncertainty analysis is difficult</td>
<td>Uncertainty analysis can be done easily</td>
</tr>
<tr>
<td>Pest &amp; disease correlation with climate variation is often absent</td>
<td>They can indirectly show some &quot;hidden&quot; correlations</td>
</tr>
</tbody>
</table>
Application of statistical methods in yield predictions: Previous studies

- Ordinary Least Squares regression
  - Some studies using quadratic terms/other regression techniques
- Limited to annual or seasonal averages (of temperature and precipitation)
- No systematic intercomparison of statistical techniques
- Less focus on uncertainty analysis
MATERIALS & METHODS
Data

• Climate data
  • Daily temperature, precipitation, radiation
    • Monthly (3*12 parameters) and fortnightly (3*26 parameters) averages over the daily climatic data

• Winter oilseed rape (yield and sowing date)
  • Denmark, Germany, Czech, (France, Belgium)
  • More than 1000 unique (site/year) observations
  • Covering more than 20 years of data up to 2013
Response function

\[
\text{Yield} = T_1 \times \text{TEMP}_1 + \ldots + T_n \times \text{TEMP}_n \\
+ P_1 \times \text{PREC}_1 + \ldots + P_n \times \text{PREC}_n \\
+ R_1 \times \text{RAD}_1 + \ldots + R_n \times \text{RAD}_n \\
+ YE \times \text{Year}
\]

Monthly resolution: 37 parameters
Forthnightly resolution: 79 parameters
Regression Techniques

- Ordinary Least Squars
- Stepwise regression
- PCR
- PLSR
- Shrinkage methods
  - Ridge
  - Elastic Nets (with alpha values of 0.25, 0.50 and 0.75)
  - Lasso

(R Packages: stats, glmnet, plsr)
Intercomparison

• Prediction
  • Hold-one-year-out for cross validation

• Inference
  • Features remaining in the final models
RESULTS

Prediction power
Root Mean Squared Error of Prediction - Denmark

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSEP (Kg/ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ridge</td>
<td>740</td>
</tr>
<tr>
<td>EL-0.25</td>
<td>760</td>
</tr>
<tr>
<td>EL-0.5</td>
<td>780</td>
</tr>
<tr>
<td>EL-0.75</td>
<td>800</td>
</tr>
<tr>
<td>LASSO</td>
<td>820</td>
</tr>
<tr>
<td>PLSR</td>
<td>840</td>
</tr>
<tr>
<td>PCR</td>
<td>860</td>
</tr>
<tr>
<td>STEPWISE</td>
<td>880</td>
</tr>
</tbody>
</table>

- **Fortnightly**
- **Monthly**
Root Mean Squared Error of Prediction – All countries
RESULTS

Inference power
# Estimated Coefficients - Denmark

<table>
<thead>
<tr>
<th>Start Date</th>
<th>01-08</th>
<th>15-08</th>
<th>29-08</th>
<th>12-09</th>
<th>26-09</th>
<th>10-10</th>
<th>07-11</th>
<th>21-11</th>
<th>05-12</th>
<th>19-12</th>
<th>02-01</th>
<th>16-01</th>
<th>30-01</th>
<th>13-02</th>
<th>27-02</th>
<th>12-03</th>
<th>26-03</th>
<th>09-04</th>
<th>23-04</th>
<th>07-05</th>
<th>1-05</th>
<th>04-06</th>
<th>18-06</th>
<th>02-07</th>
<th>16-07</th>
</tr>
</thead>
<tbody>
<tr>
<td>End Date</td>
<td>14-08</td>
<td>28-08</td>
<td>11-09</td>
<td>25-09</td>
<td>09-10</td>
<td>23-10</td>
<td>06-11</td>
<td>20-11</td>
<td>04-12</td>
<td>18-12</td>
<td>01-01</td>
<td>15-01</td>
<td>29-01</td>
<td>12-02</td>
<td>26-02</td>
<td>11-03</td>
<td>25-03</td>
<td>08-04</td>
<td>22-04</td>
<td>06-05</td>
<td>20-05</td>
<td>03-06</td>
<td>17-06</td>
<td>01-07</td>
<td>15-07</td>
</tr>
</tbody>
</table>

## Temperature

<table>
<thead>
<tr>
<th>Method</th>
<th>RIDGE</th>
<th>ELNET - Alpha =0.25</th>
<th>ELNET - Alpha =0.50</th>
<th>ELNET - Alpha =0.75</th>
<th>LASSO</th>
<th>PLSR</th>
<th>PCR</th>
<th>OLS</th>
<th>STEPWISE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start Date</td>
<td>08-07</td>
<td>15-08</td>
<td>29-08</td>
<td>12-09</td>
<td>26-09</td>
<td>10-10</td>
<td>07-11</td>
<td>21-11</td>
<td>05-12</td>
</tr>
<tr>
<td>End Date</td>
<td>14-08</td>
<td>28-08</td>
<td>11-09</td>
<td>25-09</td>
<td>09-10</td>
<td>23-10</td>
<td>06-11</td>
<td>20-11</td>
<td>04-12</td>
</tr>
</tbody>
</table>

## Radiation

<table>
<thead>
<tr>
<th>Method</th>
<th>RIDGE</th>
<th>ELNET - Alpha =0.25</th>
<th>ELNET - Alpha =0.50</th>
<th>ELNET - Alpha =0.75</th>
<th>LASSO</th>
<th>PLSR</th>
<th>PCR</th>
<th>OLS</th>
<th>STEPWISE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start Date</td>
<td>08-07</td>
<td>15-08</td>
<td>29-08</td>
<td>12-09</td>
<td>26-09</td>
<td>10-10</td>
<td>07-11</td>
<td>21-11</td>
<td>05-12</td>
</tr>
<tr>
<td>End Date</td>
<td>14-08</td>
<td>28-08</td>
<td>11-09</td>
<td>25-09</td>
<td>09-10</td>
<td>23-10</td>
<td>06-11</td>
<td>20-11</td>
<td>04-12</td>
</tr>
</tbody>
</table>

## Precipitation

<table>
<thead>
<tr>
<th>Method</th>
<th>RIDGE</th>
<th>ELNET - Alpha =0.25</th>
<th>ELNET - Alpha =0.50</th>
<th>ELNET - Alpha =0.75</th>
<th>LASSO</th>
<th>PLSR</th>
<th>PCR</th>
<th>OLS</th>
<th>STEPWISE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start Date</td>
<td>08-07</td>
<td>15-08</td>
<td>29-08</td>
<td>12-09</td>
<td>26-09</td>
<td>10-10</td>
<td>07-11</td>
<td>21-11</td>
<td>05-12</td>
</tr>
<tr>
<td>End Date</td>
<td>14-08</td>
<td>28-08</td>
<td>11-09</td>
<td>25-09</td>
<td>09-10</td>
<td>23-10</td>
<td>06-11</td>
<td>20-11</td>
<td>04-12</td>
</tr>
</tbody>
</table>
Number of Significant Features in the final model - Czech

- PLSR
- PCR
- RIDGE
- E_NETS
- LASSO

Legend:
- Month.
- Fortn.
RESULTS

Uncertainties
Variance Decomposition – Yield projection under climate change

\[ V(Y) = V_M[E_{\epsilon, \theta}(Y|M)] \]
\[ + E_M[V_\theta(E_\epsilon(Y|M, \theta))] \]
\[ + E_M[E_\theta(V_\epsilon(Y|M, \theta))] \]

Where
M: Model
\( \theta \): Set of parameters
\( \epsilon \): Residual errors
Variance decomposition - Czech

Baseline

2011-2030

2046-2065

2080-2099
Effect of model and parameter uncertainty percent of yield change predictions - Czech

Main Sample

"Best" model

"Good" models

bootstrapping
Conclusions

- State-of-the-art regression techniques could be useful, both in prediction and inference.
- Regression techniques can be useful in pointing out which climatic factors are influential for yield during which growth phases.
- Cross-validation of regression models across space (between countries) can provide a method for validating validity for use in climate change projections.
- Regression techniques offer a direct method for addressing parameter uncertainty.
THANK YOU!