Collaborative Infrastructure for Test-Driven Scientific Model Validation

Cyrus Omar + Jonathan Aldrich
carnegie mellon university

Richard Gerkin
arizona state university
Model Prediction

Empirical Observations

Observable Quantity

Cycle 23-24 Sunspot Number Prediction (February 2009)

NASA/MSFC/Hathaway
Science is Dynamic
Science is Dynamic
Science is Dynamic
A synthesis of solar cycle prediction techniques

David H. Hathaway, Robert M. Wilson, and Edwin J. Reichmann
NASA Marshall Space Flight Center, Huntsville, Alabama

Table 3. Precursor Prediction Method Errors (Prediction - Observed) for Cycles 19-22

<table>
<thead>
<tr>
<th>Prediction Method</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>RMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ohl’s method</td>
<td>-55.4</td>
<td>19.1</td>
<td>21.8</td>
<td>4.4</td>
<td>31.3</td>
</tr>
<tr>
<td>Feynman’s method</td>
<td>-42.8</td>
<td>9.6</td>
<td>26.9</td>
<td>3.6</td>
<td>25.8</td>
</tr>
<tr>
<td>Thompson’s method</td>
<td>-17.8</td>
<td>8.7</td>
<td>-26.5</td>
<td>-13.6</td>
<td>17.9</td>
</tr>
</tbody>
</table>
The state of the art?

“Other [models] were examined and found promising but were not considered due to our inability to test the techniques.”
PHILOSOPHICAL TRANSACTIONS:
GIVING SOME ACCOUNT
OF THE PRESENT Undertakings, Studies, and Labours
OF THE INGENIOUS
IN MANY CONSIDERABLE PARTS
OF THE WORLD.

Vol I.
For Anno 1665, and 1666.

In the SAVOY,
Printed by T. N. for John Martyn at the Bell, a little without Temple-Bar; and James Allestry in Duck-Lane,
Printers to the Royal Society.
scientific community ~ software development team
models ~ software components
papers ~ component documentation
implementations
interfaces
conformance tests
models ~ software components
model capabilities ~ interfaces
observable quantities ~ types
goodness-of-fit tests ~ unit tests
empirical observations ~ test parameters
summary tables ~ test result summaries

interface SunspotNumberPredictor
  predict : TimeSeries<SunspotNumber> -> TimeSeries<SunspotNumber>

OhlsMethod : SunspotNumberPredictor = ...  
FeynmansMethod : SunspotNumberPredictor = ...  
ThompsonsMethod : SunspotNumberPredictor = ...

SunspotTest : TimeSeries<SunspotNumber> -> Test<SunspotNumberPredictor, Metric> = ...

TestSuite([SunspotTest(cycle_data) for cycle_data in all_cycle_data]).judge([OhlsMethod, FeynmansMethod, ThompsonsMethod]).view()

<table>
<thead>
<tr>
<th>Prediction Method</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>RMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ohl's method</td>
<td>-55.4</td>
<td>19.1</td>
<td>21.8</td>
<td>4.4</td>
<td>31.3</td>
</tr>
<tr>
<td>Feynman's method</td>
<td>-42.8</td>
<td>9.6</td>
<td>26.9</td>
<td>3.6</td>
<td>25.8</td>
</tr>
<tr>
<td>Thompson's method</td>
<td>-17.8</td>
<td>8.7</td>
<td>-26.5</td>
<td>-13.6</td>
<td>17.9</td>
</tr>
</tbody>
</table>
interface SunspotNumberPredictor
  predict : TimeSeries<SunspotNumber> -> TimeSeries<SunspotNumber>

OhlsMethod : SunspotNumberPredictor = ...
FeynmansMethod : SunspotNumberPredictor = ...
ThompsonsMethod : SunspotNumberPredictor = ...

SunspotTest : TimeSeries<SunspotNumber> -> Test<SunspotNumberPredictor, Metric> = ...

TestSuite([SunspotTest(cycle_data) for cycle_data in all_cycle_data])
  .judge([OhlsMethod, FeynmansMethod, ThompsonsMethod]).view()

which programming language? a DSL?
Paper: SciUnit, a Python library
(widely adopted, good FFI support, simple syntax, IPython notebook)

collaborative workflow?
Paper: SciDash, a wiki pointing to IPython notebooks hosted on Github

adoption?
Paper: people who organize modeling competitions (e.g. a neural modeling competition we
are helping re-start)

Ongoing work: automatic test and model generation from existing repositories

Contact Me: cyrus@cmu.edu or @neurocy
Observable Quantity

Empirical Observations

Model Prediction
Cycle 23-24 Sunspot Number Prediction (February 2009)

Data Type

Test

Model Prediction

NASA/MSFC/Hathaway
A synthesis of solar cycle prediction techniques

David H. Hathaway, Robert M. Wilson, and Edwin J. Reichmann

NASA Marshall Space Flight Center, Huntsville, Alabama

Table 3. Precursor Prediction Method Errors (Prediction - Observed) for Cycles 19-22

<table>
<thead>
<tr>
<th>Prediction Method</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>RMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean cycle</td>
<td>-94.8</td>
<td>-9.1</td>
<td>-53.5</td>
<td>-48.6</td>
<td>59.8</td>
</tr>
<tr>
<td>Secular trend</td>
<td>-91.6</td>
<td>8.7</td>
<td>-36.2</td>
<td>-25.3</td>
<td>51.0</td>
</tr>
<tr>
<td>Gleissberg cycle</td>
<td>-80.4</td>
<td>18.5</td>
<td>-51.6</td>
<td>-51.1</td>
<td>55.0</td>
</tr>
<tr>
<td>Even-odd</td>
<td>-59.3</td>
<td>-22.3</td>
<td>-</td>
<td>-</td>
<td>44.8</td>
</tr>
<tr>
<td>Amplitude-period</td>
<td>-74.1</td>
<td>0.3</td>
<td>-61.2</td>
<td>-25.3</td>
<td>49.7</td>
</tr>
<tr>
<td>Maximum-minimum</td>
<td>-83.9</td>
<td>21.6</td>
<td>-22.9</td>
<td>-15.0</td>
<td>45.4</td>
</tr>
<tr>
<td>Ohl’s method</td>
<td>-55.4</td>
<td>19.1</td>
<td>21.8</td>
<td>4.4</td>
<td>31.3</td>
</tr>
<tr>
<td>Feynman’s method</td>
<td>-42.8</td>
<td>9.6</td>
<td>26.9</td>
<td>3.6</td>
<td>25.8</td>
</tr>
<tr>
<td>Thompson’s method</td>
<td>-17.8</td>
<td>8.7</td>
<td>-26.5</td>
<td>-13.6</td>
<td>17.9</td>
</tr>
</tbody>
</table>