On the reliability of seasonal forecasts

Antje Weisheimer
Weisheimer@atm.ox.ac.uk
ECMWF
University of Oxford

With thanks to Tim Palmer,
Laura Ferranti, Susanna Corti and Tim Stockdale
On a scale of 1 – 5, where 5 is very good, how skillful are seasonal forecasts today?

What does a “5” mean for probabilistic forecasts?

How close are we to achieving a “5”?

➤ Use reliability of non-climatological forecasts as a measure
Example: Forecasts for decision making
Agronomist to advice farmers on the type of crop to plant in the coming season

- crop yield \( C = C(X) \) with \( X \) ... meteorological variables
- seasonal forecast probability distribution \( \rho(X) \)
- climatological distribution \( \rho_C(X) \)
- expected crop yield: \( \langle C \rangle = \int_X C(X) \rho(X) dX \)
- climatological crop yield expectation: \( \langle C \rangle_c = \int_X C(X) \rho_C(X) dX \)
- if \( \rho(X) \approx \rho_C(X) \) \( \rightarrow \) climatological (reliable) information
- if \( \rho(X) \neq \rho_C(X) \) \( \rightarrow \) differences in expected crop yield \( \rightarrow \) decisions

It is essential that \( \rho(X) \) must be reliable
Reliability = correspondence between forecast probability and observed frequency of an event given the forecast.
Five proposed categories of reliability
Perfect reliability
Uncertainty ranges do include perfect reliability slope

category 5

Reliability diagram

BSS>0

observed frequency

forecast probability
Example from ECMWF’s System 4:

warm conditions in DJF over the Sahel
Significantly positive Brier Skill Score
Uncertainty ranges do not include perfect reliability slope
(includes over- and underconfident cases)
Example from ECMWF’s System 4:

wet conditions in JJA over Eastern Africa

![Diagram showing observed frequency vs forecast probability with shaded areas representing terciles.](image)
category 3

Significant positive reliability slope
Example from ECMWF’s System 4:

**dry conditions in DJF over West Africa**

![Graph showing observed frequency vs. forecast probability](image-url)
Non-significant positive reliability slope
Example from ECMWF’s System 4:

dry conditions in JJA over Southern Europe
Negative reliability slope

Reliability diagram

Category 1
Example from ECMWF’s System 4:

dry conditions in JJA over Northern Europe
cold conditions in DJF
warm conditions in DJF
cold conditions in JJA
warm conditions in JJA
dry conditions in DJF
wet conditions in DJF
dry conditions in JJA
wet conditions in JJA
Representing uncertainty in seasonal forecasts

Initial condition uncertainties $\rightarrow$ ensemble forecasting technique

- Design of the ensemble?
- Ensemble size?

**wet DJF North America**

- perfect reliability
- BSS>0

**dry JJA Europe**

- slope of reliability line
Representing uncertainty in seasonal forecasts

Initial condition uncertainties $\rightarrow$ ensemble forecasting technique
  o Design of the ensemble?
  o Ensemble size?

Model uncertainty
  o Multi-model ensembles MME
  o Perturbed physical parameters ensembles PPE
  o Stochastic physical parametrisations SPE
## Model uncertainty in monthly forecasts

<table>
<thead>
<tr>
<th>Temperature 2m</th>
<th>precipitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>May cold</td>
<td>Nov cold</td>
</tr>
<tr>
<td>cold</td>
<td>warm</td>
</tr>
<tr>
<td>MME</td>
<td></td>
</tr>
<tr>
<td>0.178</td>
<td><strong>0.195</strong></td>
</tr>
<tr>
<td>PPE</td>
<td></td>
</tr>
<tr>
<td>0.059</td>
<td>0.054</td>
</tr>
<tr>
<td>SPE</td>
<td></td>
</tr>
<tr>
<td>CTRL</td>
<td></td>
</tr>
</tbody>
</table>

Brier Skill Score for global land points and **first month** of the forecast.

The largest scores for each event are indicated in bold.

*Weisheimer et al., GRL (2011)*
# Model uncertainty in seasonal forecasts

<table>
<thead>
<tr>
<th></th>
<th>Temperature 2m</th>
<th>precipitation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>JJA</td>
<td>DJF</td>
</tr>
<tr>
<td></td>
<td>cold</td>
<td>warm</td>
</tr>
<tr>
<td>MME</td>
<td>0.084</td>
<td>0.082</td>
</tr>
<tr>
<td>PPE</td>
<td>0.004</td>
<td>0.046</td>
</tr>
<tr>
<td>SPE</td>
<td>0.059</td>
<td>0.054</td>
</tr>
<tr>
<td>CTRL</td>
<td>-0.024</td>
<td>-0.002</td>
</tr>
</tbody>
</table>

Brier Skill Score for global land points and **first season** of the forecast.

The largest scores for each event are indicated in bold.

*Weisheimer et al., GRL (2011)*
Warm conditions in DJF over the Amazon region

Impact of stochastic physics

stoch phys ON
stoch phys OFF

Impact of multi-model ensemble

System 4
EUROSIP
Wet conditions in DJF over Northern Europe

Impact of stochastic physics

Impact of multi-model ensemble

stoch phys ON
stoch phys OFF

System 4 EUROSIP
conditions in DJF over Eastern Africa

dry

warm

System 4
EUROSIP
Conclusions and outlook

What constitutes a “5”, to which a seasonal forecast system should aspire?

A perfectly reliable system, within the sampling uncertainty, that predicts probabilities different then climatology

\[ \rho(X) \neq \rho_c(X) \]

Reliability categories depend on variable, season and regions.

Need to represent physical processes accurately
- finer resolution of forecast models
- better representation of Earth System components
- improved model physics including stochastic approaches

Palmer and Weisheimer (2013)