Review JWGFVR 5th Verification workshop & SRNWP-V programme

Clive Wilson –Expert Team on diagnostics, validation & verification

With thanks to Marion Mittermaier, Rachel North, Matthew Trueman,Ric Crocker

34th EWGLAM/19th SRNWP meetings –Helsinki 8-11 October 2012
Contents

• WWRP/JWGFVR Cloud verification recommendations

• SRNWP-V programme
  • Progress & update
  • WebEx ET meeting – June 2012
  • Final reports
5th verification methods workshop
Melbourne, 5-7 December 2011

Marion Mittermaier

- 5 keynote presentations on 5 key topics

1. **Ensembles and probabilistic forecasts:** Verification of ensemble forecasts: a guided tour through a zoo of skill metrics – A. Weigel

2. **Seasonal and Climate:** Verification of seasonal forecasts – A. Watkins

3. **Aviation specific verification:** Recent developments in Aviation Verification at the UK Met Office – P. Gill

4. **User-oriented systems and topics:** Uncertainty Forecasts and the End-User – S. Jocelyn

5. **Tropical Cyclones and high-impact weather:** Verification of tropical cyclone forecasts – B. Brown
• Special verification issue of Meteorological Applications. ([Deadline was 15 September 2012.])

• Introductory paper to special issue by Ebert et al. on “Progress and challenges in forecast verification”

• Key areas where progress has been made include:
  - Improved reporting
  - Wider use of diagnostic verification
  - Development of new scores and techniques for difficult problems, and
  - Evaluation of forecasts for applications using meteorological information
Many interesting challenges remain:

• Improvements in methods to verify high-resolution ensembles
• Seamless predictions spanning multiple space and time scales
• Multivariate forecasts.

More work is needed to:

• make best use of new observations
• investigate links with data assimilation, and
• develop better intuitive “verification products” for end-users
RECOMMENDED METHODS
FOR EVALUATING
CLOUD AND RELATED PARAMETERS

March 2012

WWRP/WGNE Joint Working Group on Forecast Verification Research (JWGFVR)
Cloud verification - summary recommendations

• User-orientated:
  • total cloud cover and cloud base height
    • Low, medium, high desirable
    • Spatial bias – eg satellite cloud mask

• More generally use satellite cloud – but not analyses at short range – model contamination

• model-oriented:
  • Simulated cf observed radiances
  • profiles
Summary recommendations contd.

- Verify against:
  - Gridded observations and profiles
  - Remote-sensed satellite, lidar
  - Surface station observations
    - If auto and manual – do not mix
    - Automated cloud base for low thresholds (aviation)
- Model impact/sensitivity
  - Use cloud radar & lidar
- Compare against climatology, persistence forecasts
- Include 95%, median & IQ range for model assessments
Table 5 - A summary of metrics for evaluating categorical predictions

<table>
<thead>
<tr>
<th>Metric</th>
<th>TCA, Low, Medium, High</th>
<th>CBH</th>
<th>Cloud profile</th>
<th>LWC, MWC</th>
<th>Radiances, Brightness Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency Bias (FB)</td>
<td>***</td>
<td>***</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Symmetrical Extreme Dependency Score (SEDS) (Hogan et al., 2009)</td>
<td>***</td>
<td>***</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Odds or log-odds ratio (OR)</td>
<td>**</td>
<td>**</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Hanssen Kuipers Score (HK), Pierce Skill Score (PSS), Kuipers Skill Score (KSS)</td>
<td>**</td>
<td>**</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Probability Of Detection (POD), Hit Rate (HR)</td>
<td>***</td>
<td>***</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Probability Of False Detection (PODF), False Alarm Rate (F)</td>
<td>***</td>
<td>***</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Heidke Skill Score (HSS)</td>
<td>**</td>
<td>**</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Gilbert Skill Score (GSS), Equitable Threat Score (ETS)</td>
<td>**</td>
<td>**</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>False Alarm Ratio (FAR)</td>
<td>***</td>
<td>***</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Odds Ratio Skill Score (ORSS)</td>
<td>**</td>
<td>**</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Proportion Correct (PC)</td>
<td>**</td>
<td>**</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Threat Score (TS), Critical Success Index (CSI)</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>Gerrity Skill Score (GSS) for multi-category forecasts (see description in Annex B)</td>
<td>***</td>
<td>***</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

SRNWP-V Phase 2 Programme progress in red

• Continue & expand comparison
  • Longer more robust results – up-to-date, publication of results on EUMETNET Portal
  • higher resolution of future operational models
    • AROME and ARPEGE results processed since start of 2011
    • SEEPS scores calculated over common domain

• Additional products verified
  • Cloud amount/base
    • truth ? Auto/manual different biases
    • Distribution approach
  • Wind gust – validation underway, also distribution
Domains of 5 consortia reference models

Hirlam  UM  COSMO  ALADIN  Aladin-Lace
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean sea level pressure</td>
<td>mean bias and root mean square errors</td>
</tr>
<tr>
<td>2m temperature</td>
<td>Bias, rmse</td>
</tr>
<tr>
<td>2m relative humidity</td>
<td>Bias, rmse</td>
</tr>
<tr>
<td>10m winds</td>
<td>mean bias speed error and root mean square vector wind error</td>
</tr>
<tr>
<td>6 hourly total precipitation</td>
<td>equitable threat score and frequency bias for 0.5, 1.0 and 4.0 mm 6h⁻¹</td>
</tr>
<tr>
<td>Model</td>
<td>Label</td>
</tr>
<tr>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Hirlam reference run by FMI</td>
<td>UK-FI</td>
</tr>
<tr>
<td><strong>Aladin-France run by Meteo-France</strong></td>
<td>UK-FR</td>
</tr>
<tr>
<td>COSMO Europe run by DWD</td>
<td>UK-GE</td>
</tr>
<tr>
<td>The North Atlantic European configuration of the Unified Model run by the Met Office</td>
<td>UK-EU</td>
</tr>
<tr>
<td><strong>Aladin-Czech (LACE) run by CHMI</strong></td>
<td>UK-LC</td>
</tr>
<tr>
<td>ECMWF high resolution global model</td>
<td>EC-GM</td>
</tr>
</tbody>
</table>

Comparison over ALADIN-France domain unless otherwise stated
Sea level Pressure

3.5 year mean 01/2009-06/2012
Screen temperature
3.5 year mean
01/2009-06/2012
Relative humidity

NB screen temperature and humidity assimilated in UK-NAE
10m vector wind rmse

3.5 years
01/2009-06/2012

Regional models better than global especially shorter range
Models closer in 2012, Hirlam relatively poorer since 2011
ppn frequency bias >1mm/6h

Summer day bias, convection

Night negative bias
ppn ETS >1mm/6h

6hr Precip Accm (>= 1.0mm): Combined stations; Equitable Threat Score
Combined times: Land Obs

Cases: UK-EU, UK-FR, UK-GE, UK-FI, EC-GM

DA – radar/cloud

Equitable Threat Score

T+6

2009 2010 2011 2012

T+12

2009 2010 2011 2012

T+18

2009 2010 2011 2012

T+24

2009 2010 2011 2012

Equitable Threat Score
AROME v ARPEGE

AROME v ALADIN
Sea level Pressure
ALADIN-LACE domain
Mean 01/2009-06/2012
Screen temperature

ALADIN-LACE domain

Mean 01/2009-06/2012
10m vector wind rmse ALADIN-LACE domain
SEEPS = Stable Equitable Error in Probability Space
Rodwell et al, 2010, QJRMS 136

- Dry, light, heavy based on observed climatology (24h) at station – \( p_1, p_2, p_3 \)
- Contingency table probabilities based on these categories
- Scoring matrix – stable, equitable
  - SEEPS = 0 (perfect), \( = 1 \) (no skill, eg constant)
- Now applying to 6h accumulations in SRNWP-V
  - 6h climatology (courtesy Mark Rodwell)
Winter (DJF) 2009-2012 data

- Higher skill than summer
- EC best at longer range
- 3 groupings
  - EC/UM
  - Aladin/Hirlam
  - COSMO

Error bars 70% confidence intervals
Summer (JJA) 2009-2011 data
3 summers

- Lower skill (~-0.1) than winter
- 2 groupings
  - EC/UM/Aladin/Hirlam
  - COSMO
  - Aladin > EC
- Dip in skill at T+24 (evening-night)
- Lack of showers persisting into late evening

Error bars 70% confidence intervals
Timeseries of the 12-month running mean of the monthly average value forecast lead time T+12

- HIRLAM (T+12)
- ECMWF (T+12)
- ALADIN (T+12)
- 12km MetUM (T+12)
- COSMO (T+12)

Better EC earlier

All models show improving trend
Timeseries of the 12-month running mean of the monthly average value forecast lead time T+48
Extremal dependency scores

\[ SEDS = \log(q) - \log(H) \quad \text{log}(p) + \log(H) \]

\[ SEDI = \log(F) - \log(H) - \log(1-F) + \log(1-H) \quad \log(F) + \log(H) + \log(1-F) + \log(1-H) \]

Jan 2009 – June 2011

Matthew Trueman
Extremal dependency scores

\[ SEDI = \frac{\log(F) - \log(H) - \log(1-F) + \log(1-H)}{\log(F) + \log(H) + \log(1-F) + \log(1-H)} \]

- 99\textsuperscript{th} percentile threshold
- LAMS have better bias than ECMWF
- EC under-forecast extremes
- H, hit rate LAMS better
ND3: Spatial & scale selective verification of precipitation and cloud

- Verify against
  - Gridded analyses- ECMWF, Meteo-France, Met Office (**UK only**)  
    - Other national gridded sets?
  - High resolution radar (5 min, 1-2 km) -**UK**
  - OPERA radar composite **QC / gauge bias correction**

- Methods
  - Fractional skill (Roberts & Lean)
  - Upscaling
  - Intensity scale (Casati)
  - Structure, amplitude, location (SAL) (Wernli et al)
  - Contiguous rain areas (Ebert & McBride)
Fuzzy verification using SEVERI cloud mask as truth binary 0 or 1.
Simulating images to compare to SEVIRI

1. Spinning Enhanced Visible and Infrared Imager

Ch05 6.2 μm - upper water vapour sensitive
Ch06 7.3 μm
Ch09 10.8 μm - detects cloud and surface
SAL for clear area (negative cloud threshold). UKV model compared to cloud analysis

- Positive S,A – too little cloud
- Increase with forecast – even less and bigger holes
Other deliverables

• ND4 Inclusion of severe/high impact weather verification
  • Extreme dependency scores being applied to precipitation and wind forecasts from models

• Deliverable ND5:
  • Full documentation of the methods used in the intercomparison.
  • Newer spatial methods code to be portable.
Cumulative observed probability

$\begin{align*}
\text{Precipitation mm} & , \\
\text{Cumulative observed probability} & , \\
\text{heavy} & , \\
\text{light} & .
\end{align*}$

$p_2 = 2p_3$

$0.2$ dry $T_2$ $T_{\text{max}}$

EWGLAM 2012 © Crown copyright   Met Office
Thanks - Questions