Precipitation verification

Thanks to CMC, CPTEC, DWD, ECMWF, JMA, MF, NCEP, NRL, RHMC, UKMO
Outline

1) Status of WGNE QPF intercomparisons

2) Overview of the use of recommended methods for the verification of precipitation forecasts against high resolution limited area observations (JWGFVR, Nov 2013)
WGNE began verifying quantitative precipitation forecasts (QPFs) in the mid 1990s.

In 1995, NCEP and DWD began verifying QPFs from a number of global and regional operational NWP models against data from their national rain gauge networks.


These intercomparisons have evolved to take into account increased spatial resolution of NWP models and research advances on QPF verification methods.

WGNE QPF intercomparisons have been very useful over these years to evaluate QPF improvements of operational global NWP models.

A survey was proposed to:

i) summarize current characteristics of the WGNE QPF intercomparisons,
ii) collect suggestions for improving further these intercomparisons
WGNE QPF intercomparison survey

- Geographical domain of verification
- Observations (type, sample size, etc.)
- Observation processing (interpolation, quality control, etc.)
- Models evaluated (data characteristics)
- Model output processing
- Precipitation accumulation period (in hours)

- Precipitation thresholds
- Stratification (lead time, season, region, etc.)
- Operational QPF scores
- Confidence intervals
- Web site with WGNE QPF verification
- Contact person

Some examples:

<table>
<thead>
<tr>
<th>Forecast Centre</th>
<th>Models evaluated (data characteristics)</th>
<th>Observations (type, sample size, etc.)</th>
<th>Precipitation accumulation period (in hours)</th>
<th>QPF scores</th>
</tr>
</thead>
</table>
| NCEP            | NCEP, CMC, DWD, ECMWF, JMA, MF, UKMO   | 1) 24h/6h/3h 5km polar-stereographic grid radar+hourly gauge-based analysis, with climate calibration  
2) ~8,000 daily rain gauge over contiguous U.S. | 24h | FB, POD, FAR, POFD, TS, ETS, HK, HSS, OR, EDI, SEDS, SE |
| DWD             | NCEP, CMC, DWD, ECMWF, MF, UKMO        | Calibrated radar composite over Germany | 24h | ETS, FBI, FSS, BSS |
| MF              | NCEP, CMC, DWD, ECMWF, JMA, MF, UKMO   | French climatological rain-gauges network. ~4000 stations (1 obs/(12 km)^2) | 24h | FB, FAR, POD, POFD, HSS, CSI, ETS, EDS, SEDS |
| JMA             | NCEP, CMC, DWD, ECMWF, JMA, UKMO, BoM  | Japanese climatological rain-gauges network. ~1300 stations (1 obs/(17 km)^2) | 6, 12, 24h | FB, POD, POFD, TS, ETS, EDI |

Suggestions of improvements:
Improve spatial (0.25° or 0.2° or 0.1°) and temporal resolution (6h)
Verification against precipitation analysis
Move further to recommended scores
Fractions Skill Score
Scale: 9x9 GP

Problem with NCEP data!

5 mm/24h
Model intercomparison – deterministic forecast

24–h Precip, 20151201 to 20160229, ExTrop, Thr=5 mm, 12UTC runs

ETS

Lead time (days)

ECMWF (0.334)
UKMO (0.319)
NCEP (0.291)
JMA (0.283)
Model intercomparison – ensemble forecast

24-h Precipitation, +120 h, 12UTC runs, ExTrop (moving monthly avg)

- ECMWF (0.156)
- JMA (0.053)
- UKMO (0.027)
- NCEP (0.020)
Equitable Threat Score: FT0-24 2014/12-2015/02

Threshold [mm/24hr]
Equitable Threat Score: FT0-24 2015/06-2015/08

Equitable Threat Score vs. Threshold [mm/24hr] for different models:
- JMA
- ECMWF
- NCEP
- UKMO
- CMC
- DWD
- BoM
- MSM
Frequency Bias Index
Precipitation threshold 1 mm/day
Basis 0 UTC, accumulated rainfall 30–54 h, sample common

DirOP/COMPAS/COM 30/03/2016
ETS over ConUS, 1/2/3-day fcsts of Global Models

GFS, CMC, DWD, ECMWF, JMA, MF, UKMO

Apr-Sept 2015
Quarterly time series of Extremal Dependence Index all global models

25.4mm/day threshold

GFS, CMC, DWD, ECMWF, JMA, MF, UKMO

(verification for Metéo-France began in Mar 2011)
2) QPF recommendations

Reference note: Suggested methods for the verification of precipitation forecasts against high resolution limited area observations (JWGFVR, Nov 2013)

Primary temporal resolution (6h)
Thresholds (1, 2, 5, 10, 20, 50 mm per 6h)
Stratification (lead time, season, region, observed intensity threshold, ...)
Comparison against station observation or gridded observations
Aggregate verification scores should be accompanied by 95% confidence intervals

For deterministic model forecasts:
   Equitable threat score (ETS)
   Extremal dependency index (EDI)
   Fractions skill score (FSS) (where gridded observations are available)
   (Additional diagnostics: HR, FAR, FBI)

For probabilistic forecasts interpreted from ensembles, or by statistical post-processing
   Brier skill score BSS (and components)
   ROC area
   Continuous ranked probability skill score (CRPSS)
Survey on the use of recommended methods for the verification of NWP-based QPF against “high resolution limited area observations”

- Characteristics of high resolution precipitation observations used?
- Scores used for deterministic model
- Scores used for EPS
- QPF verification methods used for regional EPS
- Plans to move further towards suggested scores
- Any comments on the suggested methods?

Most centers are using climatological rain-gauges network and gridded precipitation analysis (combined raingauge-radar; combined raingauge-satellite).

Most centers have implemented recommended scores for deterministic model evaluation.

Few centers are computing recommended scores on EPS but many have plans to do so.

Lack of station climatology for BSS and CRPSS scores.

MF uses BSS_NO rather than FSS, which differ from the normalization. In BSS_NO, the persistence forecast is used for the reference.
Surface Observations – Daily Precip to 06Z

SYNOP

HDOBS

European Centre for Medium-Range Weather Forecasts
HDOBS 41r1-v-41r2 for France, DJF 2015/16

total precipitation
Equitable threat score value >5.0
France
Date: 201512 to 201602

![Graph showing total precipitation trends for France, DJF 2015/16 with HDOBS 41r2 and HDOBS 41r1 lines.](chart.png)
6 hours accumulated rainfall BSS_NO

Neighbourhood 50 km

Winter 2014-2015

Threshold 0.5 mm

AROME
ARPEGE
IFS

Threshold 5 mm

MF
6 hours accumulated rainfall BSS_NO

Neighbourhood 50 km
Summer 2015

Threshold 0.5 mm

Threshold 5 mm
Verification of QPF using SEEPS
Score with forecast lead time, April 2012 to February 2016

SEEPS skill score from UM
Global
6-hour accumulations
(6h to 48h)

Diurnal averages
Tropics: red & black
Global: blue & green

SEEPS skill score from UM
Global
24-hour precipitation accumulations (day 1 to 6)
Verification of QPF using SEEPS
Decomposition into constituent error sources

Diurnal Average 2012-2016

24-hour totals
10mm/6h @52km
Area of the study

349 lon points * 481 lat points with **0.00833** lat-lon increments.
1 grid size by **longitude** = \(111 \times 0.00833 = 930 \text{ m} \),
1 grid size by **latitude** = \(\cos(43^\circ 35'\times 930 \text{ m} = 0.72 \times 930 = \sim 670 \text{ m} \)
EDI of 1h precipitation, Sochi region,
Comparison with the station data (~23 stations)
as a function of threshold, COSMO-Ru1 and COSMO-Ru2

TOTAL PRECIPITATION - EDI - 00 Run

n contingency table interval from 329 to 541

3h lead time
EDI of 1h precipitation, Sochi region,
Comparison with the station data (~23 stations)
as a function of threshold, **COSMO-Ru1** and **COSMO-Ru2**
Conclusions and perspectives

Many contributions on QPF verification from centers to WGNE-31: ~140 slides

10 contributions on QPF survey

There is a clear move towards recommended methods for the verification of precipitation forecasts against high resolution limited area observations (JWGFVR, Nov 2013).

QPF verification of global models with high resolution national observation network is very useful:
- a lot of scores are produced (types, thresholds, period, etc.) and should be ideally available on Web site (password if necessary) like NCEP or MF
- some interest to increase forecast data resolution in time (at least 6h) and space (?)
- QPF intercomparison on EPS ?

Weaknesses:
- Inter-comparison of a limited number of models
- Inter-comparisons are done in several centers with similar but not identical methodologies. This does not provide a very comprehensive overview of QPF verification all over the world, like for instance for TC verification.
- Lack of station climatologies for computing BSS, CRPSS