# *Post-processing methods for short-range ensemble forecasts*

L.Descamps, C. Labadie

Météo-France DIRIC/PREVI

Météo-France CNRM/GMAP/RECYF

# ***Key points of this presentation***

- Comparison of various methods for statistical post-processing (SPP) of precipitation
  - What is the best technique ?
- The need for a reforecast data set
  - Can we do good job without reforecast data set ?
- Interest for ’rare’ events
  - Can we improve probabilistic prediction or ’rare’ events ?

# *Statistical post-processing of an operational E.P.S*

- Correcting deficiencies in estimating forecast errors
  - Remove biases
  - Increase reliability
  - Preserve resolution
  - based on learning of past forecast errors

# *Configuration of the study*

- Use of Météo-France operational system PEARP
- SPP of 24-h rainfall amount over France
- one-month period : June 2010
- Use of a 21-year reforecast data set
  - and also a sliding window of 45 days using the most recent available forecasts
- Use of SAFRAN reanalysis as reference

# *PEARP 2010 configuration*

- PEARP:Prévision d'Ensemble ARPEGE
- Initialization procedure
  - EDA + dry TE SVs
- Model characteristics
  - 35 4.5-Day forecasts run at T358c2.4 L65 resolution
  - using a set of 8 physical parametrization sets
- PEARP is runnig twice a day at 06 and 18 UTC
- Presentation of C. Labadie for details

# *Statistical post-processing of precipitation forecasts*

- A lot of methods proposed in the literature
  - Rank Histogram-based correction
  - Non-homogeneous Gaussian regression
  - CDF-based correction
  - Analog
  - Logistic Regression
  - Bayesian model Averaging
  - ...

# *Statistical post-processing of precipitation forecasts*

- In this study
- Try first some simple methods
  - simple Bias correction
  - CDF-based correction

# *Statistical post-processing of precipitation forecasts*

- In this study
- Try first some simple methods
  - simple Bias correction
  - CDF-based correction
- Try more sophisticated one
  - Rank-Analog
  - Logistic Regression
  - Bayesian model Averaging (ongoing work)

# ***Reforecast data set***

- 8-member ensemble reforecast
  - Each member uses a different set of parametrization
- Same resolution and forecast lead-time as PEARP
- Computed all four days from 01/05 to 28/07
  - 01/05, 05/05, ..., 24/07, 28/07
  - 23 days of a three-month period centered around June
- Computed over a 21-year period : 1989-2009
  - Initial condition from ERA-Int reanalysis
- Total of 8*21*23=3864 forecasts

# ***SAFRAN precipitation reanalysis***

- 24h Precipitation Amount - 06UTC-06UTC
- Gauged-based analysis and reanalysis system
- Make use of all available information in reanalysis mode
- 8-km Lambert II coordinates grid covering France (9892 points)
- Used in the hydrometeorological operational system of Météo-France

# *Implementation of the SPP methods*

- $0.25°$ gridded forecasts interpolated to the SAFRAN grid points
- A correction at each point of the SAFRAN grid
- 4 lead-times : 36, 60, 84 and 108h
- Each forecast is corrected depending on the set of parametrization it uses

# *Implementation of the SPP methods*

- Simple Bias correction
  - Biases computed for 4 classes : $P < 1mm$, $1mm < P < 5mm$, $5mm < P < 10mm$, $P > 10mm$

# *Implementation of the SPP methods*

- Simple Bias correction
- CDF-based correction
  - One $CDF_F$ computed for each set of parametrization and each lead time

# *Implementation of the SPP methods*

- Simple Bias correction
- CDF-based correction
- Rank-Analog
  - Use of 8 ’ensemble mean’ (mean of forecasts that use the same set of parametrization)
  - Searching area of 30 km
  - Use the dates with the smallest RMS rank difference

# *Implementation of the SPP methods*

- Simple Bias correction
- CDF-based correction
- Rank-Analog
- Logistic regression
  - A unique predictor : cube root of the mean forecasted amount
  - Enlarge the training sample size with 5 ’analog’ data from locations that have similar climatologies (see Hamill et al. mwr 2008)

# ***Results***

- Use of Brier Skill Score
- Use of Reliability and Resolution parts of BSS
- 5 thresholds : 0.2, 1, 5, 10 and 15mm
- 10000-member block bootstrap to quantify uncertainty in scores estimates

# ***Results***

- Results for BSS (the higher the better)

# ***Results***

- Results for BSS (the higher the better)
- Simple Bias correction always has lower skill score than ’Raw’ ensemble

# ***Results***

- Results for BSS (the higher the better)
- BSS for P>0.2mm

| Lead time (h)<br>Methods | 36 | 60 | 84 | 108 |
|---|---|---|---|---|
| RAW | 0.4 | 0.38 | 0.3 | 0.25 |
| LOGIT | 0.51 | 0.46 | 0.4 | 0.32 |
| ANALOG | 0.55 | 0.49 | 0.42 | 0.34 |
| CDF | 0.49 | 0.46 | 0.42 | 0.36 |

- Significant advantage against Raw ensemble for all methods

# ***Results***

- Results for BSS (the higher the better)
- Same results for P>1mm and P>5mm

# ***Results***

- Results for BSS (the higher the better)
- BSS for P>10mm

| Lead time (h)<br>Methods | 36 | 60 | 84 | 108 |
|---|---|---|---|---|
| RAW | 0.22 | 0.17 | 0.1 | 0.15 |
| LOGIT | 0.28 | 0.23 | 0.17 | 0.15 |
| ANALOG | 0.28 | 0.2 | 0.15 | 0.16 |
| CDF | 0.26 | 0.2 | 0.14 | 0.18 |

- Significant advantage against Raw ensemble for all methods except at 108h lead time

# ***Results***

- Results for BSS (the higher the better)
- Same results for BSS for P>15mm

# ***Results***

- Summary of the results for the Reliability part of the BSS
- The three methods have a significant advantage against Raw ensemble except for 10 and 15mm thresholds at 84 et 108-h lead time

# ***Results***

- Summary of the results for the Resolution part of the BSS
- The calibrated ensembles have better resolution than the Raw ensemble at early lead times for the lower thresholds

# ***Results***

- Is there a technique that has better results than the others ?
- It depends on thresholds and Lead-time

# ***Results***

- Is there a technique that has better results than the others ?
- It depends on thresholds and Lead-time
- Summary of the results
- Rank-Analog has higher scores at Early lead-times for 0.2 and 1mm thresholds
- Logistic-Regression approach has higher scores for higher thresholds

# ***Results***

- Can we do good job without a reforecast data set ?

# ***Results***

- Can we do good job without a reforecast data set ?
- Logistic Regression and CDF-based approaches using forecasts from the most recently available 45 days of forecasts as training
  - Summary of the results
  - Better scores for 5, 10 and 15mm thresholds when using the reforecast

# ***Results***

- Can we improve probabilistic prediction of infrequent events ?

# ***Results***

- Can we improve probabilistic prediction of infrequent events ?
- An exceptional high precipitation event over south-east of France on 15 June 2010
- More than 100 mm on Var region, more than 200 mm from St Tropez to Draguignan

# ***Results***

- Can we improve probabilistic prediction of infrequent events ?
- Using the 36h-lead time ’raw’ PEARP ensemble
- P(24h- Total Precipitation > 100mm)=0.02
- P(24h - Total Precipitation > 75mm)=0.08
- P(24h - Total Precipitation > 50mm)=0.27

# ***Results***

- Can we improve probabilistic prediction of infrequent events ?

- Using the 36h-lead time ’raw’ PEARP ensemble
- P(24h- Total Precipitation > 100mm)=0.02
- P(24h - Total Precipitation > 75mm)=0.08
- P(24h - Total Precipitation > 50mm)=0.27

- Using CDF correction
- P(24h- Total Precipitation > 100mm)=0.08
- P(24h - Total Precipitation > 75mm)=0.25
- P(24h - Total Precipitation > 50mm)=0.60

# ***Results***

- Can we improve probabilistic prediction of infrequent events ?

- Using the 36h-lead time ’raw’ PEARP ensemble
- P(24h- Total Precipitation > 100mm)=0.02
- P(24h - Total Precipitation > 75mm)=0.08
- P(24h - Total Precipitation > 50mm)=0.27

- Using Rank-Analog correction
- P(24h- Total Precipitation > 100mm)<0.001
- P(24h - Total Precipitation > 75mm)=0.005
- P(24h - Total Precipitation > 50mm)=0.04

# ***Results***

- Can we improve probabilistic prediction of infrequent events ?

# ***Results***

- Can we improve probabilistic prediction of infrequent events ?
- No Miracle ... but some questions

# ***Results***

- Can we improve probabilistic prediction of infrequent events ?
- No Miracle ... but some questions
- What can we do when one (or more) forecasted amount of the ensemble is larger than the largest value of the training sample (and the largest observed value) ?
- Using CDF correction or Rank-Analog method will automatically decrease the forecasted amount
- Can we do something else ?

# ***Conclusions and Questions***

- Probabilistic predictions can be greatly improve by using statistical post-processsing

# *Conclusions and Questions*

- Probabilistic predictions can be greatly improve by using statistical post-processsing
- No method is better than the others for all thresholds at all lead times

# *Conclusions and Questions*

- Probabilistic predictions can be greatly improve by using statistical post-processsing
- No method is better than the others for all thresholds at all lead times
- All methods have drawbacks
  - CDF correction : suppose that forecasted and observed rainfall amount are highly correlated
  - Rank-Analog technique : need a long reforecast data set / problem of ’smoothing’ when using the ensemble mean
  - Logistic regression : choice of predictor / hard to use for high thresholds

# *Conclusions and Questions*

- Probabilistic predictions can be greatly improve by using statistical post-processsing
- No method is better than the others for all thresholds at all lead times
- All methods have drawbacks
- Better scores for moderate and high thresholds with a reforecast data set as training period
  - Should we include the numerical cost of the reforecast in the global cost of EPS ?
  - How long should be the reforecast data set if we want to do good job for very high thresholds (40, 50 or 60mm) ?

Any questions ?