Impact of the statistical method, training dataset, and spatial scale of post-processing to adjust ensemble forecasts of the height of new snow

Jari-Pekka Nousu\textsuperscript{1,2}, Matthieu Lafaysse\textsuperscript{1}, Guillaume Evin\textsuperscript{3}, Matthieu Vernay\textsuperscript{1}, Joseph Bellier\textsuperscript{4,5}, Bruno Joly\textsuperscript{6}, Maxime Taillardat\textsuperscript{6,7}, Mickaël Zamo\textsuperscript{6,7}

\textsuperscript{1} Univ. Grenoble Alpes – Université de Toulouse – Météo-France – CNRS – CNRM, Centre d’Etudes de la Neige, Grenoble, France ; \textsuperscript{2} University of Oulu, Finland ; \textsuperscript{3} Univ. Grenoble Alpes – INRAE, UR ETNA, Grenoble, France ; \textsuperscript{4} NOAA Earth System Research Lab., Boulder, USA ; \textsuperscript{5} Univ. Grenoble Alpes, CNRS, IRD, Grenoble INP, IGE, Grenoble, France ; \textsuperscript{6} CNRM – Université de Toulouse – Météo-France – CNRS, GMAP, Toulouse, France, \textsuperscript{7} Météo-France, DirOP/COMPAS, Toulouse, France
Context

- Forecasting the height of new snow:
  - Safety and economic concerns

- Meteo-France **automatic forecasts** currently available (website and smartphone apps):

\[
24\text{h height of new snow} = \sum_{24\text{h}} \text{Hourly NWP precipitation} \times \text{Density} = f(\text{temperature}) \text{ (empirical law)}
\]

- Compaction during snowfall
- Possible melting
- Rain on snow

→ Severe biases
SAFRAN:
• Spatial aggregation of ARPEGE on massifs (~1000 km²)
• Adjust meteorological variables at various elevations

Crocus:
• Falling snow density = f(temperature, wind speed)
• Explicit mechanical compaction
• Melting (energy balance)
• Compaction due to liquid water (rain on snow)

Vionnet et al., 2012

Durand et al., 1998

Alternative: Physical modelling SAFRAN-Crocus
Ensemble forecasts PEARP-S2M

PEARP

SAFRAN

Crocus

35 members

PEARP

SAFRAN

Crocus

4-day forecast, Mercantour region, 2100 m elevation

Vernay et al. 2015

Experimental from 2014
Operational: October 2019

Snow depth (cm)

26/02 12h 27/02 00h 27/02 12h 28/02 00h 28/02 12h 29/02 00h 29/02 12h 01/03 00h
Raw ensemble forecasts PEARP-S2M

Evaluation over all massifs Winter 2017-2018

Rank diagram

Quantile-Quantile plot

Underdispersion!

Bias!
State of the art

- Physical ensemble modelling of the snowpack improves the forecast of the height of new snow compared to:
  - Direct NWP outputs (Champavier et al., 2018)
  - Deterministic systems (Vernay et al., 2015)

- Ensemble Model Output Statistics (EMOS) are useful to forecast the height of new snow from direct ensemble NWP outputs (precipitation and temperature) (Stauffer et al., 2018; Scheuerer and Hamill, 2019)

- Quantile Regression Forests (QRF) can incorporate more predictors and have added value for precipitation forecasts (Taillardat et al., 2019)

Questions

- Can Ensemble Model Output Statistics (EMOS) improve the forecasts from physical modelling?
  - What is the best training dataset?
  - What is the spatial validity of the post-processing?
- Can Quantile Regression Forests (QRF) improve the skill compared to EMOS?
Statistical post-processing: method

- In Nousu et al., NPG, 2019, we apply the EMOS method used by Scheuerer and Hamill (2015; 2018) for precipitation forecasts:
  - We assume that the conditional distribution of the forecast HN to the raw ensemble forecasts follow a Censored Shifted Gamma (CSG) defined by 3 parameters: **Mean** $\mu$; **Variance** $\sigma^2$; **Shift** $\delta$.

  ![Density Histogram with PDFs](image)

  - Regression model between CSG parameters and synthetic properties of the raw ensemble (mean, dispersion, probability of 0 cm)
Statistical post-processing: calibration

**Predictand:**
Network of local observations of the 24h height of new snow

**2 Predictor datasets:** Ensemble forecasts PEARP-S2M

<table>
<thead>
<tr>
<th>Period</th>
<th>Members</th>
<th>Initial conditions</th>
<th>Resolution and physics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reforecast</strong></td>
<td>1994-2016</td>
<td>10</td>
<td>Unperturbed</td>
</tr>
<tr>
<td><strong>Real-time forecasts</strong></td>
<td>2014-2017</td>
<td>35</td>
<td>Perturbed</td>
</tr>
</tbody>
</table>

**Spatial scale of the calibration:**
- Massif scale
- Station scale

**Evaluations:**
- From real-time forecasts, winter 2017-2018
EMOS: results

Nousu et al., NPG, 2019

**Raw forecasts**

- Remove bias and underdispersion
- Improvement of CRPS on most stations.
- Larger improvement at short lead times

**Corrected forecasts**

Training: reforecasts 1994-2016
Evaluation: real-time forecasts 2017-2018
Sensitivity to training dataset
Nousu et al., NPG, 2019

Training: reforecasts 1994-2016
Evaluation: real-time forecasts 2017-2018

Training: real-time forecasts 2014-2017
Evaluation: real-time forecasts 2017-2018

Severe events less reliable
Low events more reliable

CRPSS (Reference: raw forecasts)

Pros and cons on both sides

Better skill at the longest lead time
Sensitivity to spatial scale
Nousu et al., NPG, 2019

Training: local scale
Evaluation: local scale

Training: massif scale
Evaluation: local scale

Similar skill for all criteria
Limitation of EMOS:
- When all raw members expect 0 cm of snow but some rainfall, EMOS always forecast 0 cm (it does not account for potential errors in the rain-snow limit elevation).

QRF has been tested with a large set of variables as predictors:
- It is shown that rainfall amount and temperature are useful predictors to be associated with the simulated new snow depth, especially at the longest lead times.

---

**Added value of Quantile Regression Forests (QRF)**

Evin et al., in prep.

---

### Weight in QRF

**24 h**
- Height of 24h new snow
- Solid precip. amount
- Other meteorological predictors

**96 h**
- Height of 24h new snow
- Solid precip. amount
- Liquid precip. amount
- Air temperature and humidity
Added value of Quantile Regression Forests (QRF)  

- The **statistical properties** of the post-processed are **satisfactory** in both cases (flat rank histograms for both EMOS and QRF)

- A significant improvement of CRPS is obtained with QRF in theoretical experiments based on the 22-year reforecast dataset (22* [21-year training, 1-year validation] )
  → Better predictive power
Illustrations on specific cases (24h lead time forecasts):

- Raw forecasts = 0
- EMOS = 0
- Corrected with QRF

Observation:

- EMOS and QRF both correct the underdispersion and bias of raw forecasts
- On dry days, QRF provides a lower spread than EMOS
Conclusions

- Raw ensemble forecasts + snowpack modelling provide predictive but **biased and underdispersive** forecasts not well suited for **automated products**.

- **Ensemble Model Output Statistics (EMOS)** improve the forecasts from physical modelling.
  - What is the **best training dataset**?
    - *Long reforecasts* improve the **reliability** of the post-processed forecasts for the severe and unusual events
    - *But they should be more homogeneous* with the operational system (initial perturbations)
  - What is the spatial validity of the post-processing?
    - *Spatial consistence of biases* allows to apply corrections at the massif scale (1000 km²)

- **Quantile Regression Forecasts (QRF)**
  - Better predictive skill in theoretical experiments thanks to other predictors
  - Further work required to test the robustness when transferred to real time forecasts
References

More details for the EMOS results in our main reference:


Other references


