ENSEMBLE PREDICTION OF AVALANCHE HAZARD

Matthieu Vernay[1], Matthieu Lafaysse[1*], Laurent Mérindol[1], Gérald Giraud[1] and Samuel Morin[1]

[1] Météo-France - CNRS, CNRM-GAME UMR3589, Centre d'Études de la Neige (CEN), Grenoble, France

ABSTRACT: The prediction of avalanche hazard results from an analysis of current snow conditions, the upcoming meteorological conditions and their combined impact on the future state of the snowpack. The SAFRAN – SURFEX/ISBA-Crocus – MEPRA (S2M) chain of numerical models is used by French avalanche forecasters to estimate present and future avalanche hazard over areas assumed to be meteorologically homogeneous (massifs), primarily as a function of altitude. The input provided until now to S2M in terms of meteorological forecast comes from the deterministic numerical weather prediction model ARPEGE with a lead-time of 2 days. Here we introduce an application to avalanche hazard forecast of ensemble meteorological prediction using the output of an ensemble of 35 ARPEGE forecasts, which are used to feed S2M and thus provide an ensemble of 35 different predicted snowpack conditions. A posteriori ensemble forecasts were generated and evaluated in the French Alps for the winter 2013-2014 with 4 days lead time, initialized each day at 6 UTC. Forecasts over the Pyrenees during the exceptional winter and spring 2012-2013 are also described. Results indicate that accounting for the uncertainty in meteorological forecast improves significantly the skill and the usefulness of the modeling chain, regardless of the prediction lead time. The predictability of snowpack conditions using the ensemble forecast technique remains good at a 4 day lead time. This will allow, in the future, building probabilistic estimates of simulated avalanche hazard level in support of operational avalanche hazard forecasting activities.

KEYWORDS: ensemble forecast, avalanche hazard, numerical modeling, predictability, probabilistic forecast

1. INTRODUCTION

The prediction of regional-scale avalanche hazard (i.e., for example the production of avalanche bulletins) results from an analysis of current snow conditions, the upcoming meteorological conditions and their combined impact on the future state of the snowpack. In France, the SAFRAN – SURFEX/ISBA-Crocus – MEPRA (S2M) chain of numerical models (Durand et al, 1999; Lafaysse et al, 2013) is used to provide an objective assessment of the past and future snow conditions including mechanical stability. It explicitly accounts for altitude, slope and aspect within geographical areas assumed to be meteorologically homogeneous (massifs). French avalanche forecasters use S2M outputs for current and predicted snow conditions in combination with weather forecasts and field information from dedicated ground observation networks. Until now, the meteorological inputs to SAFRAN used for operational S2M forecast have been primarily based on the output of the deterministic numerical weather prediction (NWP) model ARPEGE operated by Météo-France with a lead-time of 2 days, which in practice allows to provide information relevant only for the day after bulletin issuance. The prediction system does not allow to account for errors originating from the meteorological forecast itself, nor does it take into account intra-massif variability of snowpack properties within a given altitude, slope and aspect topographical class.

Ensemble prediction is increasingly used for meteorology and hydrology applications (e.g. Thirel et al, 2008, 2010; Voisin et al, 2011), such as large scale medium-range meteorological forecast and short-term forecast of potentially extreme events such as flash-floods, but to the best of our knowledge has not been applied to avalanche hazard forecasting hitherto. However, the large sensitivity of the snowpack state to meteorological conditions makes its prediction challenging especially in mountain regions, thereby providing motivation to develop ensemble-based prediction techniques applied to avalanche hazard warning. Here we introduce an ensemble prediction system bringing together tools and data operated by Météo-France. A comprehensive evaluation of the performance of the system is assessed through comparison with the current deterministic prediction

* *Corresponding author address:*
Matthieu Lafaysse, Météo-France, Centre d'Etudes de la Neige, 1441 rue de la Piscine, 38400 Saint Martin d'Hères, France ;
tel: +33 476637928
email: matthieu.lafaysse@meteo.fr

system and the analysis of snow conditions, in terms of fresh (less than one day old) snow thickness and a dedicated regional scale natural avalanche hazard index (NHI).

2. MATERIAL AND METHODS

2.1 SAFRAN – SURFEX/ISBA-Crocus – MEPRA model chain

The generation of consistent meteorological variables used to perform numerical simulations of the snowpack is carried out by the meteorological downscaling and surface analysis tool SAFRAN (Durand et al, 1999). The main original feature of SAFRAN is that it operates at the geographical scale of meteorologically homogeneous mountain ranges (referred to as « massifs ») within which meteorological conditions are assumed to depend only on altitude and aspect. Based on a robust data assimilation scheme combining large-scale output from NWP models, ground-based and radiosonde observations and remotely-sensed cloudiness, SAFRAN provides hourly meteorological conditions for each massif for 300m-spaced altitude bands. Variables dealt with by SAFRAN include not only precipitation (rainfall and snowfall rate) and air temperature but also relative humidity, wind speed, incoming longwave and shortwave radiation. SAFRAN also operates in forecast mode, solely using NWP output. In this case, SAFRAN is a downscaling tool from the NWP model grid to the massifs/altitude bands geometry. SAFRAN outputs feed the detailed snowpack model SURFEX/ISBA-Crocus (Vionnet et al, 2012), which computes the evolutions of the physical properties of the snowpack and the underlying ground as a function of altitude, slope and aspect within each massif. The mechanical stability of the snowpack simulated by Crocus is estimated by the model MEPRA (Giraud, 1992) which diagnoses the avalanche hazard level for each simulated snowpack profile and provides an integrated natural avalanche hazard index at the massif scale (NHI) ranging from 1 to 8 (8 corresponds to highest hazard level). This index takes into account the mechanical stability at different altitudes from 1500 to 3000m for 40° slopes and 8 aspects.

2.2 PEARP ensemble prediction

SAFRAN (thus S2M) is currently fed by the deterministic large scale NWP model ARPEGE operated by Météo-France. Here we use the 35 members of the ensemble prediction system based on ARPEGE (PEARP; Nicolau, 2002). This system is based on a 6-member ensemble assimilation (Berre et al, 2007) combined with the singular vectors perturbations method (e.g. Molteni et al, 1996) to provide 35 initial states to the NWP model. The 35 members are randomly run with 10 different physical parameterizations of sub-grid processes (Descamps et al., submitted). The PEARP horizontal resolution is larger than ARPEGE deterministic configuration (about 15 km vs. 10 km over France, respectively) and covers 4 .5 days lead time (108 h). Two different types of output of the PEARP were used in this study. The first one includes all the vertical levels available from ARPEGE (named PEARP-FULL) but without operational real-time availability and a data storage limited to 6 months. The second one, referred to as PEARP-BDAP, corresponds to real-time available outputs (from the BDAP operational database) which could be used as such for a real-time forecasting system and which does not suffer from a time restricted data storage, but it contains fewer vertical levels. Here we use PEARP-FULL data to estimate the influence of the number of vertical levels available on the forecast quality and assess whether PEARP-BDAP data could still be used operationally for our purpose.

2.3 Evaluation

The performance of the new PEARP-BDAP - S2M model chain was assessed and compared to results of the deterministic and the PEARP-FULL model chains with 4 days lead time. In both cases the reference used was the analysis S2M chain in order to compare similar variables, because the spatial scale of simulated variables do not have observation equivalents. Quantitative evaluations were made using 2760 forecasts corresponding to the 23 massifs of the Alps and 120 analysis dates spanning the period from 1 November 2013 to 1 March 2014 (one 4-day forecast per day). A more qualitative study was carried out using the months of January, February and June 2013 in the Pyrenees, where challenging situations occurred (major avalanches in winter, snowmelt floods in June). The variables that were evaluated here are the thickness of fresh snow fallen in the last 24h at 1800 m altitude (H24) and the NHI.

2.4 Scores

The scores used for these evaluations are both deterministic and probabilistic: the Root Mean Square Error (RMSE) is a classical deterministic score, which requires restricting the analysis to the ensemble average. It is therefore useful to also compute the ensemble dispersion, which represents the standard deviation of the members relatively to their average:

$$D = \frac{1}{N}\sum_{k=1}^{N}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2}$$

where x_i is the forecast of member i for day k, m the ensemble mean, n the number of members and N the number of days. The dispersion is expected to be of the same magnitude as the RMSE if the ensemble represents correctly the forecast uncertainty.

The most common probabilistic score is the Brier score (Brier, 1950) which describes the ensemble forecast system performance in terms of a given threshold exceedance. It is defined by:

$$BS = \frac{1}{N}\sum_{k=1}^{N}(y_k - o_k)^2$$

where y_k and o_k are respectively the forecasted probability of the event for the day k, and the corresponding binary observation (o_k=0 or 1)

The Brier Score ranges from 0 to 1, where 0 corresponds to a perfect score. It can been decomposed in three terms (Murphy, 1973): the *reliability* is the ability to predict good probabilities, the *resolution* is the ability to separate probability classes and the *uncertainty* is independent of the system.

The Brier Skill Score (BSS) offers the possibility to compare the Brier score of an ensemble forecast system with the Brier score of a reference system:

$$BSS = 1 - \frac{BS}{BS_{ref}}$$

It was used to compare the PEARP-BDAP chain with the deterministic and the PEARP-FULL ones. In the case of the deterministic forecast, we assumed that the model only predict probability values of 0 and 1.

As the evaluation samples size is far smaller than recommendations from Candille and Talagrand (2005) to evaluate the performance of an ensemble forecast system, the robustness of the computed scores was tested using a Bootstrap method (Efron, 1979).

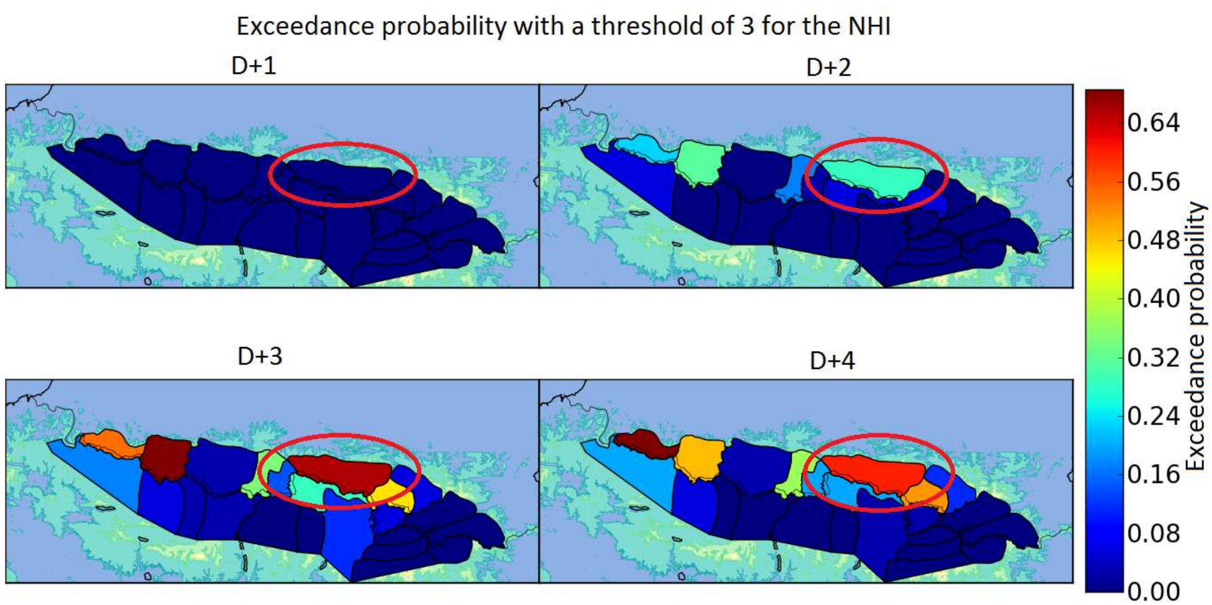

Fig. 1: Map of exceednce probability with a threshold of 3 for the massif-level NHI over the Pyrenees. Forecast of PEARP-BDAP-S2M system from the 01/13/2014 at 06 UTC for D+1 (01/14/2014 at 06 UTC), D+2 (01/15/2014 at 06 UTC), D+3 (01/16/2014 at 06 UTC), and D+4 (01/17/2014 at 06 UTC). The circled massif is the Couserans massif (see below)

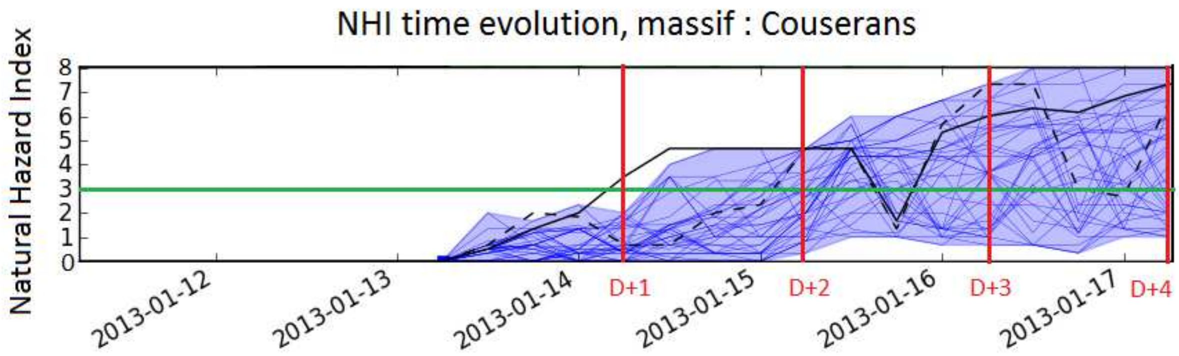

Fig. 2: NHI time evolution forecast for the Couserans massif from the 01/13/2014 at 6 UTC to 01/17/2014 at 06 UTC. Blue lines: 35 members of PEARP-BDAP-S2M; black solid line: S2M analysis; black dash line: ARPEGE-S2M deterministic forecast.

3. RESULTS AND DISCUSSION

3.1 *Example*

An example of results of the PEARP-BDAP-S2M chain is shown on Figure 1 consisting in a map of exceedance probabilities with a threshold of 3 for the NHI over the Pyrenees. In this particular case, high probabilities of important avalanche events are forecast for many French massifs, especially at a 3 days lead time. During that day, numerous big avalanches indeed occurred, resulting in roads break and important damages in many areas of the Pyrenees. The corresponding time evolution of the 35 members is illustrated on Figure 2 for the Couserans massif (located in the red circle in Figure 1). The number of forecast members over the threshold 3 (green line) illustrates the computed probabilities for the different prediction lead times. The dispersion between members increases with lead time, as expected. The deterministic forecast (black dash line) sometimes differs significantly from the ensemble mean.

3.2 *Evaluation of the PEARP-BDAP-S2M chain over French Alps*

The general evaluation of the PEARP-BDAP - S2M chain over the French Alps is summarized in Table 1. Scores concerning the quality of the forecasts of the PEARP-BDAP - S2M chain (Brier score, reliability, resolution, uncertainty, RMSE and dispersion) show that this model chain features promising forecast skills at all lead times. The low Brier score indicates a good forecast of threshold exceedance probabilities. However, the dispersion is lower than the RMSE which shows an under-dispersion of the system. The deterministic scores such as RMSE are slightly better for the ensemble mean than for the deterministic forecast. From a probabilistic point of view, the ensemble forecast is far better with BSS ranging from 0.15 to 0.25. These very satisfactory results are also obtained for the H24 (not shown).

Score	*D+1*	*D+2*	*D+3*	*D+4*
Brier Score	0.07	0.07	0.08	0.09
Reliability	0.01	0.01	0.01	0.01
Resolution	0.06	0.06	0.04	0.03
Uncertainty	0.12	0.12	0.12	0.12
Brier Skill Score	0.21	0.25	0.24	0.24
Dispersion (cm)	1.9	2.2	2.6	3.3
RMSE (cm)	4.0	4.2	4.9	5.1
Deterministic RMSE	4.1	4.3	4.9	5.3

Tab. 1: Summary of the different scores computed for the forecast of NHI (with a threshold of NHI = 1 for the Brier score

To better illustrate the quality of the forecast probabilities, all forecasts were gathered in 8 classes of forecast probabilities and for each class, the observed occurrence frequency of the event is computed. Reliability diagrams of Figure 3 present the 8 couples (forecast probability; observed frequency) for different NHI thresholds and lead times. The diagonal corresponds to a perfect ensemble forecast system. The probabilities appear slightly underestimated, especially for probabilities below 0.5 but in general they are very reliable, even at D+4. Another very good behavior is that when the extreme probabilities are forecast (0 member or conversely all members predict the events), the events were almost never (respectively always) observed.

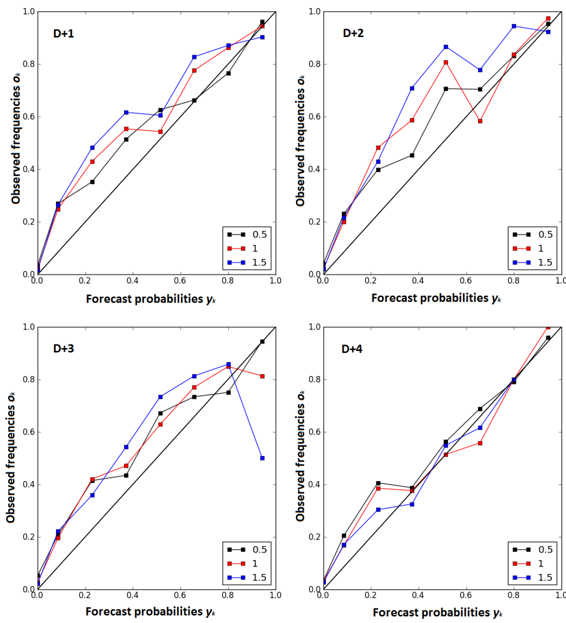

Fig. 3: Reliability diagrams of the forecasted probabilities of NHI with different exceedance thresholds (black: 0.5; red: 1; blue: 1.5) and for the 4 prediction lead times (D+1, D+2, D+3, D+4).

The results are not homogeneous in space as illustrated by the map of Brier scores (Figure 4). The spatial patterns of skills depend on scores and lead time. The bootstrap method showed that the assessment of these scores at the massif scale is less robust than at the scale of the whole French Alps (typical uncertainty of 0.005 against 0.001) but the differences between massifs are significant (no-overlapping confidence intervals even for contiguous massifs, as illustrated on Figures 4). In spite of these differences, there are very few massifs where the ensemble chain does not present better skills than the deterministic one (data not shown).

3.3 <u>Sensitivity to the number of available vertical levels</u>

All scores were also computed for the PEARP-FULL-S2M system (with more atmospheric vertical levels). The difference between the Brier Score of both systems is +0.003 at D+1, -0.0006 at D+2, -0.0007 at D+3 and +0.0004 at D+4. The sign of the difference depends on lead time, and the values of the difference are very low (10 to 100 times smaller than the values of the score). When looking at the differences obtained for each massif (data not shown), the medians of the bootstrap samples scores are very close to 0 and the differences between massifs are much lower than the uncertainty on the scores assessment linked to the sampling.

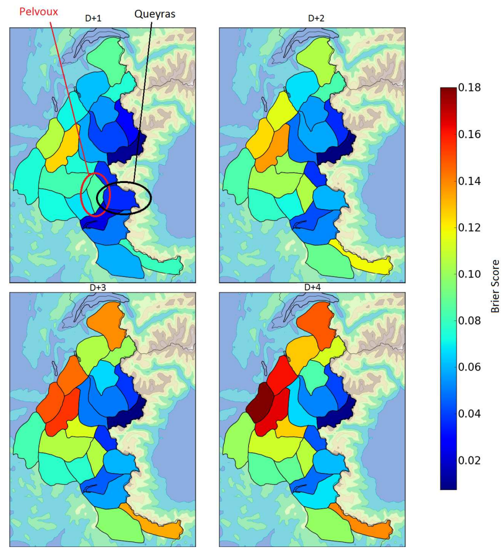

Fig. 4: Brier score by massif for NHI forecasts with a threshold of 1 for the 4 prediction lead times (D+1, D+2, D+3, D+4).

Therefore, the comparison of outputs using PEARP-BDAP vs. PEARP-FULL leads to conclude that differences in terms of forecast quality are insignificant.

3.4 <u>Case studies over Pyrenees</u>

The period between 16 to 18 January 2013 was exceptional in the Pyrenees with very high fresh snowfall quantities (60 cm to 1m) and with severe natural avalanche activity. Figure 5 illustrates the forecasts of the PEARP-BDAP-S2M system for this exceptional event in terms of snow depth and NHI over 2 massifs (Aspe-Ossau in the Western part and Couserans in the Eastern part). For the Aspe-Ossau, the event is anticipated by the ensemble forecast system from January 13th, but underestimated: the most pessimistic member always simulates snow depth 20 cm below S2M analysis and only few members forecast NHI as high as the analyzed ones. Forecasts are more satisfactory for the Couserans massif as the analyzed snow depth is included into the ensemble dispersion. It is also very interesting to note that in some cases (especially at the end of this period), the deterministic model forecast a significant decrease of NHI while the analyzed NHI remained very high. In the ensemble forecast, the dispersion is very high but some members clearly suggest that the NHI would be able to keep its highest values. In this case, the probabilistic forecast provides therefore much more relevant information than the deterministic one. The forecast would have been obviously uncertain based on the probabilistic chain, but it would have been completely wrong based on the deterministic chain.

On 17 and 18 June 2013, high precipitation amount (between 60 and 160 mm) were observed over the Western and Central parts of the Pyrenees in the context of a snow cover exceptionally important for the season above 2000 m and with meteorological conditions boosting snowmelt (high temperatures, strong winds, high solar radiation, cloudy nights preventing snow refreezing). As a result, exceptional floods occurred. The damages were larger than expected solely based on precipitation amount, because snowmelt significantly contributed to the saturation of catchments. Figure 6 shows the time evolution of S2M analyzed snow water equivalent (black line) over the Luchonnais massif at 2400 m. The snowpack lost 250 mm water equivalent within one week. This intense snowmelt is very well anticipated by the ensemble forecast, even at the farthest lead times. The dispersion between members is much lower than in previous plots, illustrating a much better predictability than during winter snowfalls.

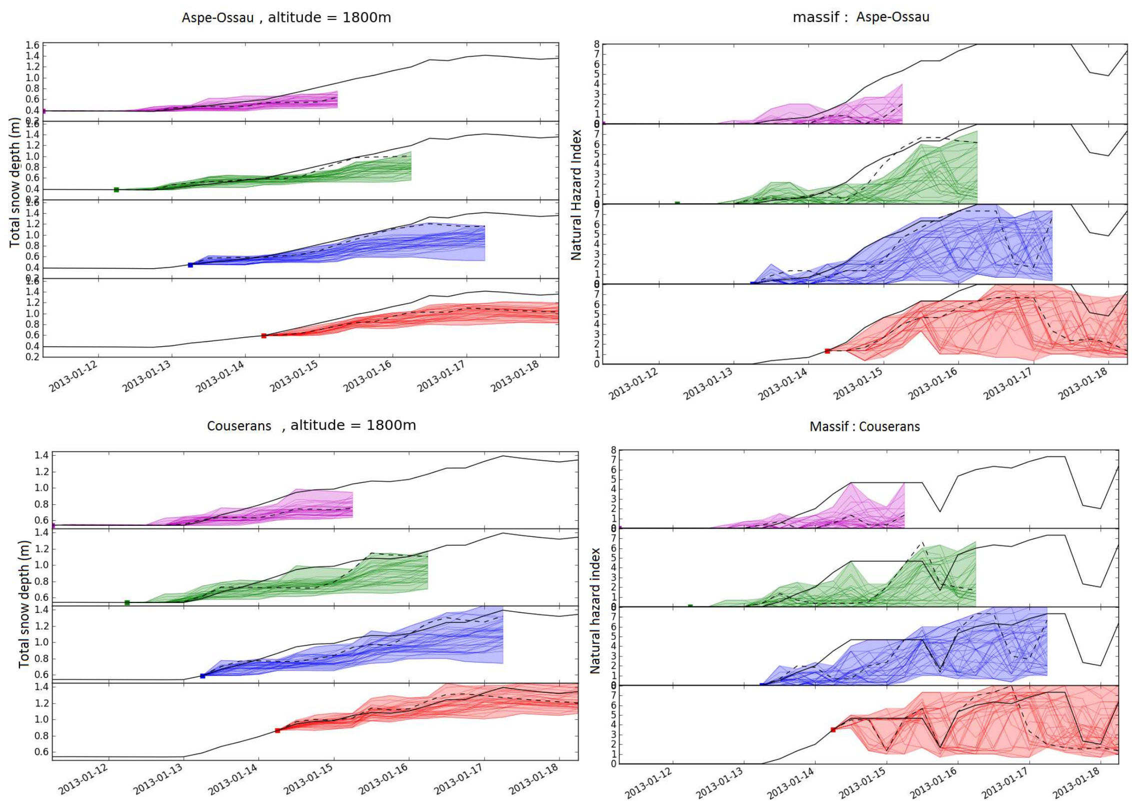

Fig. 5: Total snow depth (left) and NHI (right) time evolution for the Aspe-Ossau (top) and Couserans (bottom) massifs. The 4 forecasts (4 lines) start from the 01/11/2014 at 6 UTC to 01/14/2014 at 06 UTC and cover the 4 following days respectively. Colored lines: 35 members of PEARP-BDAP-S2M; black solid line: S2M analysis; black dash line: ARPEGE-S2M deterministic forecast.

4. CONCLUSIONS AND OUTLOOK

This study has allowed implementing for the first time, to the best of our knowledge, an ensemble prediction system dedicated to forecast avalanche hazard. The implementation brings together recent developments in terms of meteorological ensemble prediction at Météo-France (PEARP) together with the new model chain used to support avalanche hazard activities (SAFRAN - SURFEX/ISBA-Crocus - MEPRA). Results obtained in various mountain ranges and for various types of meteorological conditions including events with exceptionally high avalanche levels in the Pyrenees indicate that the ensemble forecasting system PEARP-S2M is able to contribute to the improvement of the avalanche hazard warning activities through its probabilistic methodology and its increase of the prediction lead time. Even if they do not provide the entire range of atmospheric levels output from PEARP, data stored in the operational database BDAP were found to be sufficient to drive PEARP-S2M. A few specific short-term improvements are still needed, such as the development of a dedicated precipitation unbiasing dataset applicable to PEARP-S2M outputs. Nevertheless, the technical environment and the development approach chosen make it possible to implement within a few month this modeling chain in real time allowing its operational use by avalanche forecasters, at least experimentally.

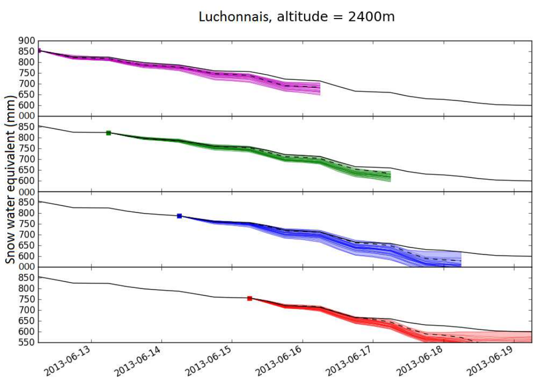

Fig. 6: Total snow water equivalent time evolution for the Luchonnais massif. The 4 forecasts (4 lines) start from the 06/12/2014 at 6 UTC to 06/15/2014 at 06 UTC and cover the 4 following days respectively. Colored lines: 35 members of PEARP-BDAP-S2M; black solid line: S2M analysis; black dash line: ARPEGE-S2M deterministic forecast.

As illustrated from one example, we believe that other applications could take benefit from this approach, such as hydrological prediction in mountainous catchments, meteorological forecast in mountainous regions and additional services yet to develop targeting the broader mountain community (i.e. ski resort managers including snowmaking).

ACKNOWLEDGEMENTS

We thank Pierrick Cébron, Carole Labadie and Pascal Raynaud from the CNRM/GMAP/RECYF team for providing PEARP-FULL data and for their explanations about the PEARP system. We also thank Fabienne Rousset-Regimbeau (DCLIM/AVH), Pascal Laveau (ENM) and Cécile Coléou (CEN) for useful advices and discussions.

REFERENCES

Berre L, Pannekoucke O, Desroziers G, Stefanescu SE, Chapnik B, Raynaud L. (2007). A variational assimilation ensemble and the spatial filtering of its error covariances: increase of sample size by local spatial averaging. In: ECMWF Workshop on Flow-dependent aspects of data assimilation. Reading, pp. 151-168.

McClung, D. M. and P. A. Schaerer (2006): *The Avalanche Handbook*. 3rd ed The Mountaineers, 347 pp.

Brier, G. W. (1950). Verification of forecasts expressed in termes of probability. *Mon. Weather Rev*., 78:1–3.

Candille G, Talagrand O (2005) Evaluation of probabilistic prediction systems for a scalar variable *Q J R Meteorol Soc* 131:2131–2150.

Descamps, L., C. Labadie, A. Joly, E. Bazile, P. Arbogast, P. Cébron, PEARP, the Météo-France short-range ensemble prediction system, Quarterly Journal of the Royal Meteorological Society, in revision.

Durand, Y., Giraud, G., Brun, E., Mérindol, L. et Martin, E. (1999). A computer-based system simulating snowpack structures as a tool for regional avalanche forecasting. *J. Glaciol.*,45(151):469–484.

Efron, B. (1979). 1977 Rietz Lecture - Bootstrap methods, another look at the jackknife. *Ann. Stat*., 7(1):1–26.

Giraud, G. (1992). MEPRA: an expert system for avalanche risk forecasting. In Proceedings of the International snow science workshop, 4-8 oct 1992, Breckenridge, Colorado, USA, 97–106.

Lafaysse, M., Morin, S., Coléou, C., Vernay, M., Serça, D., Besson, F., Willemet, J.-M., Giraud, G. et Durand, Y. (2013). Toward a new chain of models for avalanche hazard forecasting in French mountain ranges, including low altitude mountains. In Proceedings of the International snow science workshop, 7-11 oct 2013, Grenoble, France, 162–166.

Molteni F, Buizza R, Palmer TN, Petroliagis T (1996) The ECMWF ensemble prediction system: methodology and validation*. Q J R Meteorol Soc* 122:73–119

Murphy, A. H. (1973). A new vector partition of the probability score. *J. Appl. Meteorol*.,12:595–600.

Nicolau, J., (2002) Short-range ensemble forecasting at Météo-France—A preliminary study. Proc. Tech. Conf. on Data Processing and Forecasting Systems, Cairns, QLD, Australia, WMO/Commission on Basic Systems.]

Thirel G., Regimbeau F., Martin E., Habets F. (2008) On the impact of short-range meteorological forecasts for ensemble streamflow predictions *J Hydrometeorol* 9(6):1301-1317

Thirel G., Regimbeau F., Martin E., Noilhan J., Habets F. (2010) Short - and medium - range hydrological ensemble forecasts over France *Atmos Sci Lett* 11 (2):72-77

Vionnet, V., Brun, E., Morin, S., Boone, A., Martin, E., Faroux, S., Moigne, P. L. et Willemet, J.-M. (2012). The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2. *Geosci. Model. Dev*., 5:773–791.

Voisin, N., Pappenberger, F., Lettenmaier, D. P., Buizza, R. et Schaake, J. C. (2011). Application of a Medium-Range Global Hydrologic Probabilistic Forecast Scheme to the Ohio River Basin*. Wea. Forecasting*, 26(4):425–446.