COMPARISON OF VARIOUS FORECAST PRODUCTS OF HEIGHT OF NEW SNOW IN 24 HOURS ON FRENCH SKI RESORTS AT DIFFERENT LEAD TIMES

Champavier Robin[1], Lafaysse Matthieu[1], Vernay Matthieu[1], and Coléou Cécile[1]

[1] *Univ. Grenoble Alpes, Université de Toulouse, Météo-France, CNRS, CNRM, Centre d'Études de la Neige, Grenoble, France*

ABSTRACT: This work compares the skill of various forecasting products of Height of new Snow (HS) available in French mountainous areas. The solid precipitation of AROME and ARPEGE Numerical Weather Prediction (NWP) models at the grid cell scale are converted in HS with an empirical density law used in Meteo-France operational automatic products or with a fixed density. We also consider massif-scale forecasts including a physical simulation of the snowpack on the ground and adjusted at the elevation of the stations. Finally, we consider supervised forecasts at a fixed elevation. The skill of the products is assessed at 10 ski resorts in Northern French Alps. The skill highly depends on the snowfall amount with significant and systematic biases of some products for the highest HS values. This work demonstrates that an optimal automatic forecast of HS requires (1) a spatial aggregation of the NWP model outputs and (2) a statistical post-processing of the forecasts to remove model biases.

KEYWORDS: Forecasting, Height of new snow, French ski resorts, Skill score, Models

1. INTRODUCTION

Forecasts of Height of new Snow (HS, Fierz et al, 2009) are crucial in mountainous areas for various economic activities (tourism attractiveness, snow management in ski resorts, roads managements, avalanche hazard survey, etc.). Realistic forecasting of this variable is still a challenge for several reasons. First, it is difficult to describe the high variability of HS as a function of elevation in mountainous areas, even at the best spatial resolution available in Numerical Weather Prediction (NWP) models. Then, several processes are not or not well represented in NWP models such as the density of falling snow, mechanical compaction during the deposition and variations of the rain-snow limit elevation during some storm events.

Currently in France, there are two types of HS forecasts. First, the forecasters provide daily an assessment of the expected HS for the next 2 days at a 24 hour time step at the massif scale (about 1000 km²) for the specific elevation of 1800 m a.s.l. These estimates are provided on the Meteo-France website and in the avalanche hazard estimation bulletins. The main limitation is the fixed and unique altitude. Automatic prediction offers a better spatial coverage. They can be extracted from the Numerical Weather Prediction models of Meteo-France: AROME (high spatial resolution model, Seity et al, 2011) and ARPEGE (global circulation model, Courtier et al, 1994). Despite the continuous improvement of their horizontal resolution, the surface elevation remains smoothed in these models. Therefore, the snow forecasts at the point scale are affected by potential discrepancies with the elevation of the model grid cell. Furthermore, these models only provide snow precipitation estimates and do not represent accurately the processes affecting the snow on the ground. Using a snowpack numerical modeling system can solve this issue. In France, the SAFRAN – SURFEX/ISBA-Crocus – MEPRA (S2M) chain (Durand et al, 1999; Lafaysse et al, 2013) includes two main components. The SAFRAN system (Durand et al, 1998) adapts ARPEGE forecasts to a geometry best suited to mountainous areas including more specifically a detailed elevation resolution. The Crocus snowpack model (Vionnet et al, 2012) contains all physical equations describing the evolution of the snowpack on the ground. It can be seen as a physical tool to adapt SAFRAN meteorological outputs to snow forecasts. The S2M chain provide meteorological predictions adapted at the specific elevation of each ski resort and account for the topographic specificities of each point (aspect, shadows, etc). The goal of this study is to compare the skill of these different products. Section 2 describes the dataset and the evaluation metrics chosen for the study. Section 3 presents the results. The main learnings are discussed in Section 4.

2. MATERIAL AND METHODS

This study focuses on the comparison between these models outputs and observations from 10 French ski resorts selected for their number of available observations and to cover a large part of the mountainous area of French Northern Alps as well as a large range of elevation (Figure 1). Their elevation range between 1400 m and 2600 m a.s.l. is shown on Figure 1. Table 1 shows the difference between resort elevation and NWP models grid cells elevation. It shows that there is a significant difference between some resorts elevation and models grid cells elevation. These differences can induce some systematic bias. The SAFRAN – Crocus chain is not affected by this issue because

* *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

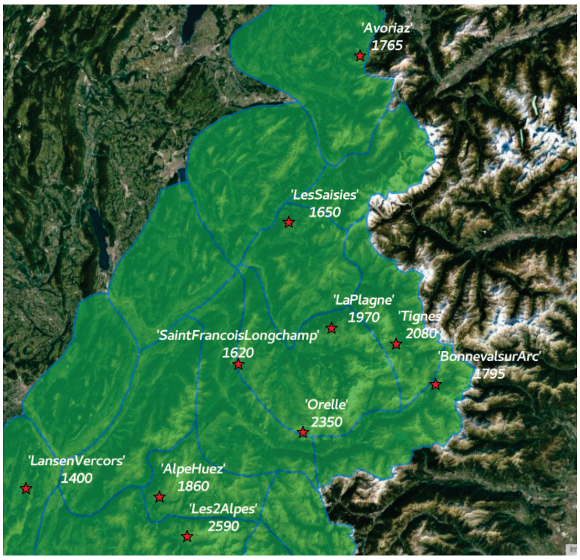

Figure 1 : Resorts used in this study with altitude in meters over the SAFRAN mountainous area.

Resort	Elevation	AROME	ARPEGE
La Plagne	1970	2198	2132
Les Saisies	1650	1675	1746
Tignes	2080	1705	1962
Lans En Vercors	1400	1357	1265
Orelle	2350	2600	2683
Saint François Lonchamp	1620	1816	2087
Les 2 Alpes	2590	2594	2609
Avoriaz	1765	1865	1994
Bonneval sur Arc	1795	2093	3015
Alpe d'Huez	1860	1910	2250

Table 1 : Difference between resorts and models points grid elevation in meters

simulations are applied at the specific elevation of each resort.

2.1 Height of new snow assessment

2.1.1 NWP models

NWP models do not provide a value of HS on the ground (in centimeters), but a solid precipitation flux (in kg.m^{-2}h^{-1}). A common way to convert this variable into HS is to assume a fixed density and to neglect compaction, melting and other processes happening to the snow on the ground. A density of 100 kg.m^{-3} is commonly used by Meteo-France forecasters to assess HS. In our study, this value is applied to all NWP models (AROME, ARPEGE and SAFRAN).

Meteo France has also developed an equation named snow potential (PN), which outputs a value of HS from the solid precipitation flux based on a density which is a function of air temperature. The critical air temperature is -2°C and the snow density D in kg.m^{-3} is included between 50 kg.m^{-3} and 300 kg.m^{-3} following the next equations :

$$D=\begin{cases} max(50;85+5\,x\,T)\,if\,T\leqslant -2\,°C \\ min(300;160{,}371\,x\exp(0{,}38\,x\,T))\,if\,T>-2\,°C \end{cases}$$

This equation comes from the forecasters but has never been extensively evaluated on mountainous areas. In this study, it has been calculated for AROME and ARPEGE at a hourly time step and then cumulated at a daily time step. They are named respectively ARO_PN and ARP_PN in the following text.

2.1.2 Crocus outputs

Currently, only Crocus output provide directly HS on the ground with an empirical parameterization of falling snow density (Vionnet et al., 2012; Pahaut, 1975) from air temperature and wind speed which is applied on SAFRAN solid precipitation output. Then, the simulation accounts for compaction during and after the snowfall, and potential melting in the case of a warming with or without rainfall. As the age of the snow layers is an explicit model diagnostic, it is possible to extract the 24h HS which is directly comparable to observations.

2.2 Datasets

All observations are given at 06h UTC in the morning. Here, the evaluation of 24h HS is performed for the four pasts winters (2015-2018) between the 1st of December and the 30th of April and at a 30h lead time for all the automatic products previously described. It means that the period of new snow accumulation is included between 06h UTC present day and 06h UTC next day. NWP models AROME and ARPEGE are used with the run initialized at 00h UTC. The corresponding ARPEGE run is used to force the S2M chain.

2.3 Skill score

The statistical distribution of daily precipitation is far from Gaussian (Ye et al, 2018). The higher snowfall rates are the least frequent but their impact can be very important. This is why we have decided to use skill scores conditioned to a threshold to explicitly discriminate the skill of the high HS values. We computed the following classical scores: frequency bias (BIAS), Heidke skill score (HSS), probability of detection (POD), false alarm rate (FAR). All these scores are defined from the contingency table, represented on Figure 2 as follows:

$$BIAS=\frac{a+b}{a+c} \qquad POD=\frac{a}{a+c} \qquad FAR=\frac{b}{a+b}$$

$$HSS=\frac{ad-bc}{((a+c)*(c+d)+(a+b)*(b+d))/2}$$

		Observation: yes	Observation: no	Marginal of Fcst
Forecast	yes	a: hits	b: false alarms	$a+b$: yes fcsts
Forecast	no	c: misses	d: correct rejection	$c+d$: no fcsts
Marginal of Obs		$a+c$: yes obsvd	$b+d$: no obsvd	N: Total forecasts

Figure 2 : Contingency table

The Heidke Skill Score measures the improvement of the forecast compared to a random forecast.

Each score is computed separately for each ski resort. Then, an average of each score among the 10 resorts is also computed to summarize the general behaviour.

2.4 *Statistical calibration*

Statistical calibration methods are commonly applied to meteorological forecasts to remove systematic model biases and improve the final forecast products (Wilks, 2011). Therefore, as a first step, we evaluate here the skill of a linear regression between observations and forecasts. We also computed the previous skill scores for the forecast product with the best linear correlation (Crocus, section 3.3). This is done independently for each ski resort. This calibration is called 'AS_CRO' in the following text.

2.5 *Comparison with forecasters predictions*

Forecasters predictions are given at 1800 meters a.s.l. for all mountainous areas. Due to the lack of a systematic storage of these predictions in an easily accessible database, it was only possible to compare supervised and automatic forecasts during the previous 2017-2018 winter at La Plagne and Les Saisies ski resorts.

3. RESULTS

3.1 *Averaged skill score*

Figure 3 shows the averaged results over the 10 resorts for each forecast product. First, a positive frequency bias (i.e. overestimation) appears for the forecast products based on ARPEGE with the low thresholds (especially 2 cm). AROME is unbiased on that criteria. When the threshold increases, the bias of the products with a 100 kg.m^{-3} density remains stable whereas the positive bias of the snow potential highly increases for both ARPEGE and AROME. Conversely the bias of SAFRAN-Crocus physical simulations becomes significantly negative especially above 20 cm. The simple statistical adaptation by linear regression (AS_CRO) is not able to completely remove this bias.

About probability of detection, AROME gets the worst score for the lowest threshold. On the other hand, Crocus statistical adaptation gets the best score until 5 centimeters. For threshold higher than 5 centimeters the detection rates decrease for all models.

The false alarm rate tends to increase with the threshold with rates included between 20% and 40% for the low thresholds and exceed 50% above 20 cm for the snow potential. However, Crocus gets the best score in average because it remains almost constant regardless of the threshold.

As for POD, the Heidke Skill Scores are also higher on low thresholds than above 20 cm. Crocus exhibits the best score for a threshold of 2 cm but the poorest above 20 cm. For NWP models, the snow potential exhibits a slight improvement compared to a density of 100 kg.m^{-3} for thresholds between 1 and 5 cm.

3.2 *Spatial variability*

The averaged scores of section 2.1 are expected to depict the mean behaviour through French Northern Alps but they can dissimulate strong disagreements between stations as the spatial variability of snowfall is known to be high. To show an example of variability, Figure 4 represents the same scores as Figure 3 for 2 forecast products but without averaging over 10 resorts. Crocus is shown in red and AROME with 100 kg.m^{-3} density in green and each line corresponds to one resort.

Crocus gets a lower variability of the frequency bias between resorts but this variability is higher for other scores and increases with the threshold. The large variability of AROME frequency bias skill between 0.5 and 2.1 over ski resorts means that its high horizontal resolution make the skill of the forecasts more heterogeneous.

Elevation differences between AROME grid cells and ski resorts only partly explain the variability of observed skill score. Indeed, the highest frequency bias and false alarm rate for the high threshold represents Saint François Longchamp ski resort with very strong overestimations of snowfall by AROME, while Crocus gets a lower underestimate score. This AROME trend may be due to model grid cell effects, which may overrate orographic effects on this specific area.

La Plagne resort is represented by the full line with black dots on Figure 4 for AROME and Crocus. The corresponding results are more detailed in next section.

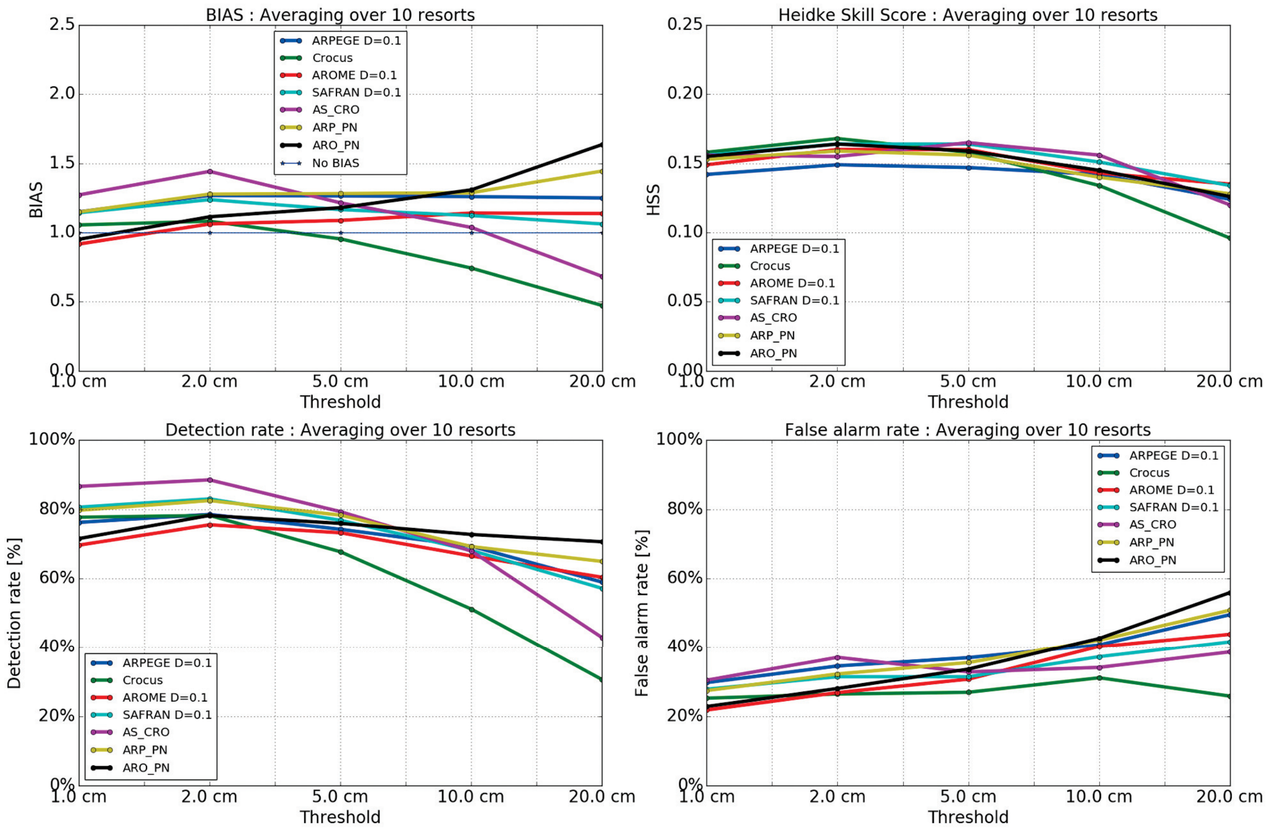

Figure 3 : Skill scores of 24h HS for all products, 10-resorts average at 30h lead time as a function of HS threshold (ARP=ARPEGE ; ARO=AROME ; AS=Statistical calibration ; PN=Snow Potential ; CRO=Crocus)

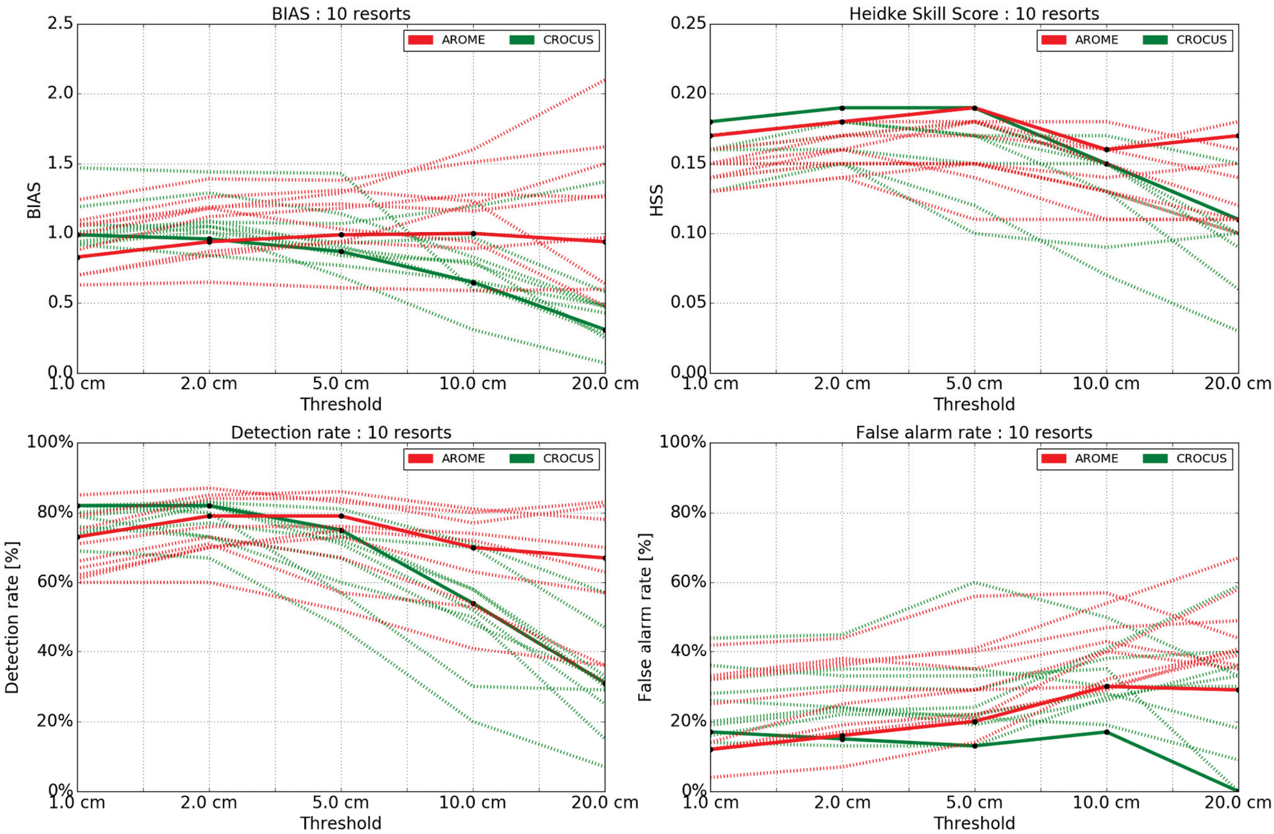

Figure 4: Skill score of 24h HS for Crocus and AROME models over 10 ski resorts (10 lines) at 30-Hours lead time as a function of HS threshold. La Plagne resort is represented with full line and black dots.

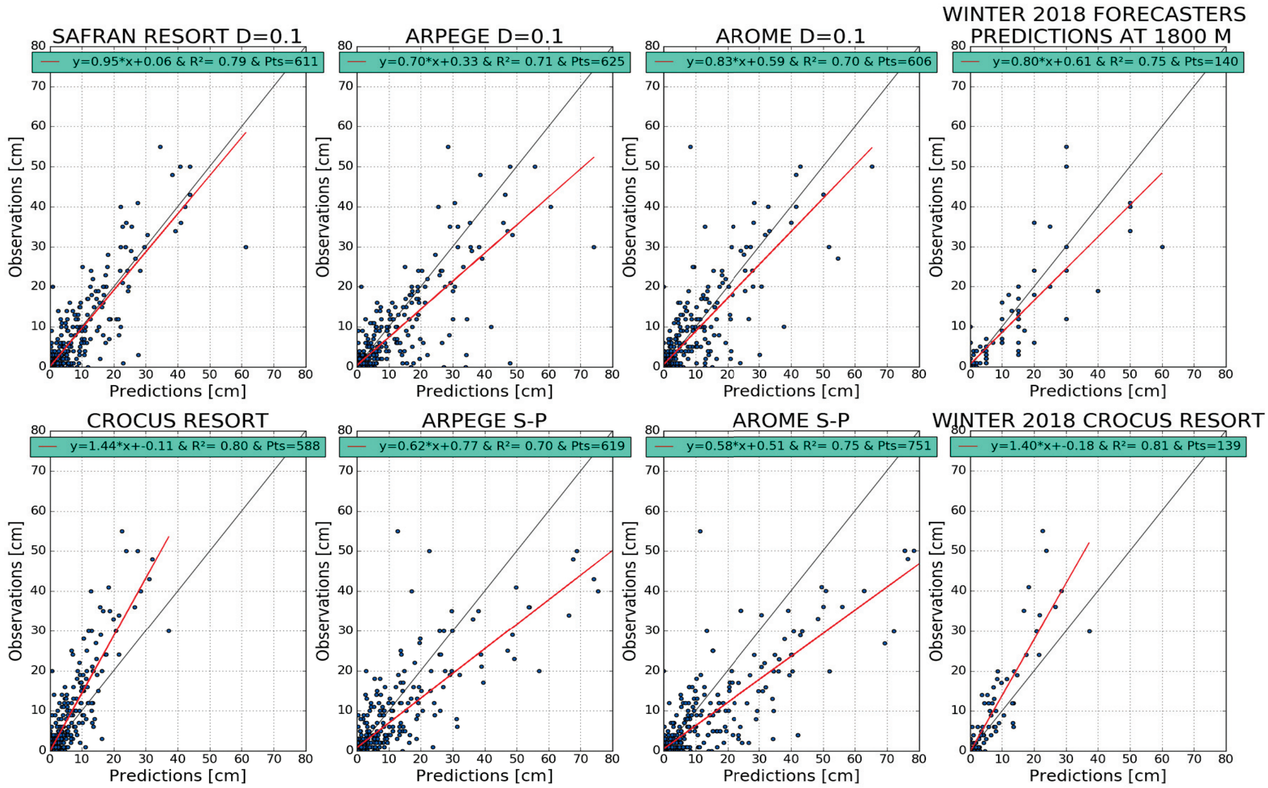

Figure 5: HS scatter plot between SAFRAN, ARPEGE, AROME, CROCUS, ARPEGE SNOW POTENTIAL, AROME SNOW SNOW POTENTIAL models and La Plagne observations at 30-Hours lead time. On the right column, supervised forecast at 1800 meters and Crocus over winter 2018.

3.3 Scatter plot example

If the previous scores can summarize the forecast skill over a long time period and to compare the skill between products and resorts, it is also useful to compare forecasts and observations at the event scale to get a better idea of the magnitude of the errors. Figure 5 shows a scatter plot example for each forecast products at the resort of La Plagne at 1970 meters a.s.l.. Some results previously described can also be seen here such as the negative bias of Crocus and the strong positive bias of the snow potential for the highest snowfalls. Note that the 150 to 200 m elevation difference between the NWP models grid cells and the station may partly explain the bias. For this resort the less biased product is SAFRAN with a 100 kg.m^{-3} snow density. Crocus outputs exhibit a higher linear correlation with observations (0.81) than ARPEGE and AROME outputs regardless the density assumption. It means that it explains better the variance of HS. This behaviour is also obtained in average for the 10 ski resorts (not shown).

The last figures on the right of Figure 5 show a comparison between the forecasters predictions at 1800 meters and Crocus output for La Plagne during the winter 2018 in order to compare these products over the same period. This figure shows that Crocus gets a better coefficient of determination, but with an important negative bias. Forecasters predictions show a positive bias although it is given at 1800 meters, 170 meters lower than the resort elevation. A negative bias would be expected with such a positive elevation difference. This positive bias is consistent with the biases of the products the most commonly used by the forecasters (ARPEGE and AROME precipitation amounts with assumptions on density).

4. DISCUSSION, CONCLUSION, OUTLOOK

4.1 Learnings of this study

Several forecast products of 24h HS were compared for 10 ski resorts of Northern French Alps. They exhibit a high variability of skill depending on the considered height threshold and on the ski resort.

The improvement expected by the high resolution of the AROME system turned to increase the spatial variability of the skill and it results in some resorts with very strong biases. The fact that SAFRAN outputs are better than ARPEGE outputs demonstrate that a spatial aggregation of NWP models is still required to improve the reliability of the forecasts at the mountainous area local scale.

The improvement expected by the physical representation of snow on the ground in the Crocus snowpack model turned to significantly underestimate new snowfall amount at all ski resorts and direct output from SAFRAN with a density of 100 kg.m^{-3} are less biased and with similar correlations and skill scores. This disappointing result encourages to better identify if this problem is due to incorrect physics in the model (falling snow density, compaction) or to compensation errors with precipitation amount. This variable was not evaluated here because the measurements of solid precipitation are highly uncertain (Kochendorfer et al, 2017). New evaluations of snowfall density and of compaction laws are therefore required on sites where measured precipitation are available (Krinner et al, 2018). Various formulations of these processes are already implemented in the Crocus model and can be easily tested (Lafaysse et al, 2017).

Finally, the very simple linear adjustment between SAFRAN-Crocus forecasts and observations demonstrate that statistical calibrations can be useful to significantly reduce model biases. More sophisticated methods are available in the literature (Gneiting et al., 2014) and could be applied both on deterministic forecasts or on ensemble forecasts of HS. Our first comparisons with raw forecasts suggest that such unbiased automatic forecasts should be able to reach or outperform the skill of supervised forecasts but with a more extended spatial coverage (at all altitudes).

4.2 <u>Required extensions</u>

The skill of forecasts usually decrease with lead time (Vernay et al, 2015). This is also the case for the products of this study but the results were not presented here. The main behaviours (main biases, threshold and spatial dependencies) are also valid at longer forecast lead times. A spatial extension of the evaluations over all French massifs will be required for operational purposes. A higher number of ski resorts would also be useful for more robust evaluations and to investigate possible dependencies of the skill with elevation.

Finally ensemble forecasts are also available for all the automatic products considered in this study: PE-AROME for AROME (Bouttier et al, 2016), PEARP for ARPEGE (Descamps et al, 2014) and PEARP-S2M for the S2M chain (Vernay et al, 2015). If such ensemble systems are expected to help accounting for the uncertainty of the forecast products, preliminary results demonstrate that significant biases of the ensemble versions of the systems compared to the deterministic ones lead to deteriorate the skill scores. These biases are mainly due to the different horizontal resolutions of the ensemble NWP models compared to their deterministic configuration, and this issue has to been investigated for a better quantification of the uncertainty of HS forecast.

ACKNOWLEDGMENTS

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 730203.

REFERENCES

Bouttier, F & Raynaud, Laure & Nuissier, Olivier & Ménétrier, Benjamin. 2015. Sensitivity of the AROME ensemble to initial and surface perturbations during HyMeX. Quart. J. Roy. Meteorol. Soc. 142. 10.1002/qj.2622.

Courtier P., J.-N. Thépaut et A. Hollingsworth, 1994 : A strategy for operational implementation of 4D-VAR, using an incremental approach. Quart. J. Roy. Meteor. Soc., 120, 1367-1387.

Descamps, L & Labadie, C & Joly, A & Bazile, Eric & Arbogast, Philippe & Cébron, P. 2014. PEARP, the Météo-France short-range ensemble prediction system. Quart. J. Roy. Meteorol. Soc. 141. 10.1002/qj.2469.

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.

Durand, Y., Giraud, G., Mérindol, L. 1998. Short-term numerical avalanche forecast used operationally at Meteo-France over the Alps and Pyrenees. International Glaciological Socitety. Anals of Glaciology, 26.

Fierz, C., Armstrong, R.L., Durand, Y., Etchevers, P., Greene, E., McClung, D.M., Nishimura, K., Satyawali, P.K., Sokratov, S.A., 2009. The international classification for seasonal snow on the ground. IHP-VII Technical Documents in Hydrology n 83, IACS Contribution n 1.

Gneiting, Tilmann & Katzfuss, Matthias. 2014. Probabilistic Forecasting. Annual Review of Statistics and Its Application. 1. 125-151. 10.1146/annurev-statistics-062713-085831.

Kochendorfer, J. et al..: The quantification and correction of wind- induced precipitation measurement errors, Hydrol. Eart Syst. Sci., 21, 1973–1989, doi:10.5194/hess-21-1973-2017, 2017.

Krinner, G. et al.: ESM-SnowMIP: Assessing models and quantifying snow-related climate feedbacks, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-153, in review, 2018.

Lafaysse, M., Cluzet, B., Dumont, M., Lejeune, Y., Vionnet, V., and Morin, S.: A multiphysical ensemble system of numerical snow modelling, The Cryosphere, 11, 1173-1198, https://doi.org/10.5194/tc-11-1173-2017, 2017.

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 haz- ard forecasting in French mountain ranges, including low altitude mountains. In Proceedings of ISSW, 162–166, 2013.

Pahaut, E.: La métamorphose des cristaux de neige (Snow crystal metamorphosis), Monographies de la Météorologie Nationale, Vol. 96, Météo France, 1975.

Seity, Y., P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson, 2011: The AROME-France Convective-Scale Operational Model, Mon. Wea. Rev., 139, 976–991, https://doi.org/10.1175/2010MWR3425.1

Vernay, M., Lafaysse, M., Merindol, L., Giraud, G., and Morin, S.: Ensemble Forecasting of snowpack conditions and avalanche hazard, Cold. Reg. Sci. Technol., 120, 251–262, doi:10.1016/j.coldregions.2015.04.010, 2015

Vionnet, V., Brun, E., Morin, S., Boone, A., Faroux, S., Le Moigne, P., Martin, E., and Willemet, J.-M.: The detailed snowpackscheme Crocus and its implementation in SURFEX v7.2, Geosci. Model Dev., 5, 773–791, doi:10.5194/gmd-5-773-2012, 2012.

Wilks S. D., 2011. Statistical Methods in the Atmospheric Sciences, Volume 100 3rd Edition. Academic Press 2011

Ye, L., Hanson, L. S., Ding, P., Wang, D., and Vogel, R. M.: The Probability Distribution of Daily Precipitation at the Point and Catchment Scales in the United States, Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-85, in review, 2018.