WEATHER TYPES PREDICTION AT MEDIUM-RANGE FROM ENSEMBLE FORECASTS

Gabriel Jouan[1,3], Anne Cuzol[2], Valerie Monbet[3], Goulven Monnier[1]

***Abstract*—Uncertainties assessment performed by numerical weather ensemble forecast system is active research in the statistical weather community. The univariate correction challenge of ensemble forecast known as univariate calibration is a well-known problem with linear models. The recent application of a non-linear approach from the machine learning domain offers a new way to perform calibration of the ensemble forecast. The multivariate component of weather forecast with medium forecasting range represents a high economic value for decision making but an actual difficult border to cross in calibration. One way to study the multivariate calibration and keep a high economic value is the weather types classification. For example, to plan maintenance activities outside or an important manifestation, good weather is preferred and can be studied in the ensemble forecast at medium-range. Good weather can be seen as a weather type created by the interaction of the wind speed and the precipitation. In this article, the multivariate calibration will be managed by a weather types classification problem using the non-parametric approach from the machine learning domain and ensemble forecast at medium-range.**

I. MOTIVATION

Nowadays, meteorological institutes provide ensemble forecasts like, for instance, the European Center for Medium-range Weather Forecast (ECMWF) ensemble. However such ensemble forecasts of surface weather parameters are known to be under-dispersed and often biased [1],[2] [3],[4].

To improve the accuracy of such forecasts, statistical post-processing has been studied these last years. One of the most common approach is to calibrate the ensemble, for each variable separately, by a regression model which helps to predict observations of the variable given a description of the ensemble as input. For example, the state-of-art method, referred to as Ensemble Model Output Statistics (EMOS;[5]), is based on an heteroscedastic linear regression. More recently, non-parametric algorithms have been proposed [6], [7], [8]. Multivariate calibration were also developed to reproduce dependencies between variables [9].

In this article, we focus on multivariate forecasting for horizon higher than 3 days. Large range forecasting is valued for maintenance operations in many fields. Here, we focus on wind intensity and precipitations. However, the problem of large range forecasting is difficult because the weather is not well predictable. To simplify the problem of multivariate calibration, we chose to deal with weather types and the quantitative variables were transformed into qualitative ones. Qualitative information like ”Good weather” or ”Rainy weather” is sufficient for many applications. Our goal becomes to predict weather types using an ensemble forecast.

vm: a reprendre The machine learning algorithm of random forest (RF; [10]) will be applied to the problem of classification using the ensemble forecast of wind speed and rainfall at a fixed spatial point and forecasting medium-range. To assess the performance of classification of the RF algorithm, a Multinomial Lasso regression (MLR) will be performed [11]. The recent univariate calibration approaches cited before have shown a high capacity to produce reliable ensemble forecast. Consequently, we have adopted a methodology of univariate calibration and classification. The forecast ensembles post-processed by the univariate calibration methods will be reordered by an empirical copula method to assure the same dependencies has the multivariate observation. The multivariate ensemble obtained would be classified. This method would allow us to compare the performance of the random forest applied on the classification of weather types with the recent univariate calibration methods.

The paper is organised as follows. In Section II, part A, the considered data are introduced and the weather type are defined. Then, the classification algorithms used for the post-processing are described.

Corresponding author: G Jouan, gabriel.jouan@scalian.com [1]Scalian Alyotech Rennes; [3]Univ Rennes, IRMAR; [2]Mathematics laboratory of Bretagne Atlantique, University of Bretagne sud

In Section IIIA., the performances of the proposed methods are compared for the forecasting of wind and rainfall at the ranges 5 days and 10 days for a chosen location in North-West part of France. The paper ends with some concluding remarks in Section III.B.

II. METHODOLOGY FOR THE PREDICTION OF WEATHER TYPES

Our aim is to calibrate multivariate large range forecast ensembles for one location. Since qualitative information is sufficient for some applications, we propose to tackle the problem of weather type prediction, each weather type being defined from several variables as described in Section A hereafter. To solve this classification problem, two approaches are considered. The first one is based on a direct application of machine learning classification algorithms (Section B). The second one consists in applying the weather type definition to the output of classical calibration methods (Section C). J'ai change le plan : classification directe d'abord puis utilisation des techniques de calibration standard apres. Peut-łtre qu'il faut revenir au plan precedent?

A. Data description and definition of the weather types

Ensemble forecast data of the ECMWF ([12]) are collected from the Thorpex interactive grand global ensemble archive (TIGGE; [2],[1]). The ECMWF ensemble system is composed of 50 exchangeable ensemble members generated from random perturbations in the initial conditions and stochastic physics parametrization ([13],[14]). The TIGGE archive includes a minimum of 10 ensemble forecasting system on a time-period from 2006 (for older ensemble forecast system) to actually.

Collected data are composed of observations and ensemble forecasts of precipitations (Precip, mm) and wind speed (WS, m.s^{-1}) at 5 days and 10 days of forecasting range, two runs (6 am and 6 pm) for the French city of Rennes. The data set is splitted in two subsets. The first one (years 2008 to 2013) is used to learn the models and the second one (years 2014 to 2018) is used to compare the algorithms. As mentioned earlier, the continuous data are transformed to define weather types. $K = 4$ balanced weather types are chosen. For an observation vector $y = (y^{Precip}, y^{WS})^\top$, and the set of thresholds $W = \{0.02, 2.8, 4.0\}$, the ϕ thresholding function is defined as:

$$\phi(y, W) = \begin{cases} 1 & y^{Precip} < 0.02, y^{WS} < 2.8 \\ 2 & y^{Precip} \geq 0.02, y^{WS} < 4.0 \\ 3 & y^{Precip} \geq 0.02, y^{WS} \geq 4.0 \\ 4 & y^{Precip} < 0.02, y^{WS} \geq 2.8 \end{cases} \tag{1}$$

The four weather types are referred to as *good* if $k = 1$, *rainy* if $k = 2$, *rainy and windy* if $k = 3$ and *windy* of $k = 4$.

The ensemble members can not be used directly as inputs of the machine learning algorithm because they are not exchangeable [15]. Then, following [6], the considered features are some statistics of the ensembles of the precipitation and wind speed, namely means, standard deviations, kurtosis, skewness, first and ninth deciles, interquartile range and precipitations propabilities. It is standard to also add the control and the high resolution members to the features set. The month and the hour, considered as factor inputs, allows to taken into account the daily and yearly cycles existing in the data. Finally, since we consider (observed) weather types as output, we decide to also add in the inputs the corresponding weather types computed for the ensembles by applying the thresholding function ϕ.

To estimate raw ensemble classification performance, each raw ensemble member is classified with the same methods described in the step 5 of section II.B. vm: je ne comprends pas cette phrase... est ce que a ne devrait pas łtre dans la partie suivante voire seulement dans la partie validation?

B. Direct classification

The problem is now to predict a discrete variable with 4 classes given continuous and discrete inputs. Two classical Machine Learning algorithms are considered: random forest and lasso logistic regression.

Random forest classifier (RFC) was proposed by [16] and [10]. It combines elementary classification trees, learned on subsamples of the data, generated randomly, to estimate the probability of each weather type. The principle of each tree is to infer a partition of the input space by a greedy algorithm. Each part is called a leave. At each step of the algorithm, the current leave is split into two parts if it improves a Gini impurity. At the end, the probabilities of the weather types estimated the mean of the responses of all the trees.

Penalized multinomial Lasso regression (MLR) is described for instance in [11]. Each multinomial regression model predicts the probability of a weather type against the other ones so that $K - 1$ models are fitted. The Lasso penalty helps to select the most discriminant variables. The idea is to penalize the log likelihood by a the sum of the absolute values of the coefficients of the regression. It has the consequence to shrink to zero the coefficients of the not useful inputs and to lead to more a robust prediction tool.

C. *Classification from multivariate calibration*

The direct machine learning classification algorithms proposed in this paper have to be compared to other machine learning solutions proposed in the ensemble forecast calibration literature, in particular [17] which also used random forests. In [17] and reference therein, the authors perform univariate calibration with quantile regression forests [18].

Here, we propose to apply a quantile regression forest for each variable, wind speed and precipitation, separately. Then the two independent calibrated ensembles are combined by a Schaake Shuffle (SS) algorithm [19] to reproduce the dependent structure existing between observed wind speed and precipitations. In SS, the marginal post processing are combined to reproduce the empirical copula estimated from past observations. In recent improvement, referred to as SimSchaake; [20], [9] proposed to combine SS algorithm with analog approaches. It allows to select past observations from meteorological configurations close to the current one and it reduces the bias in the estimation of the dependence structure. After applying SimSchaake, the thresholding function ϕ is used to derive the calibrated weather type ensemble.

Est-il utile de garder les differentes etapes decrites ci-dessous? The classification obtained from a multivariate calibration approach for a new ensemble forecast can be defined by the following steps:

1) $\forall p \in \{1,..,P\}$, fit p random forest [18] to perform univariate calibration on each p weather variable with the T learning set of $Y = \{(y_1^1, ..., y_T^1), ..., (y_1^P, ..., y_T^P)\}$ observation and $X = \{(x_1^{(1,1)}, ..., x_M^{(1,1)}), ..., (x_1^{(P,T)}, ..., x_M^{(P,T)})\}$ of the ensemble forecast with M members for a chosen spatial location and forecasting range.

2) For a new $T+1$ ensemble forecast, generate $\tilde{x} = \{(\tilde{x}_1^1, ..., \tilde{x}_N^1), ..., (\tilde{x}_1^P, ..., \tilde{x}_N^P)\}$ a set of the post-processed ensemble composed of N members using the trees prediction with the random forest applied on the $T+1$ ensemble forecast.

3) As mentioned in the article [20], compute the similarity between the $T+1$ ensemble forecast and each $t \in \{1, ..., T\}$ ensemble of X. Select N observation of Y corresponding of the N elements minimising the similarity calculated above. After computing the ordering statistics $y_{(1)}^p \leq ... \leq y_{(N)}^p$, $\forall p \in \{1,..,P\}$, store the permutation induced $h_n^p = \text{rank}(y_n^p)$, $\forall n \in \{1,..,N\}$ in a permutation set H.

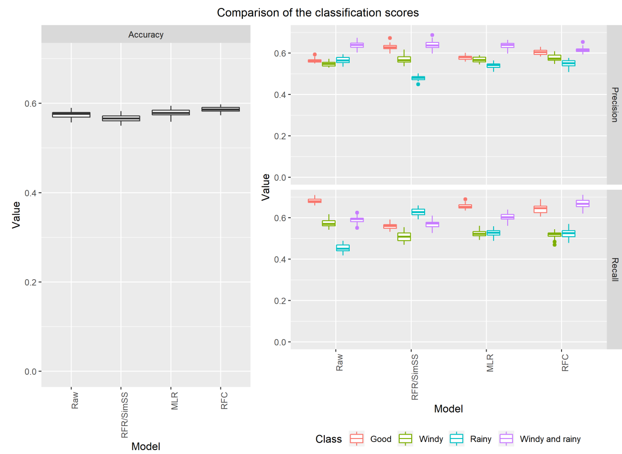

Fig. 1. Classification scores of models on Rennes data set with ensemble forecast at 5 days (5d) estimated by cross validation. left- Accuracy score by model, top-right- Precision score by model and classes, bottom-right- Recall score by model and classes. RFR/SimSS: classification obtained from a multivariate calibration, MLR: Multinomial lasso regression , RFC: Random forest classifier.

4) Apply H on $\tilde{x}$ to produce the set $\hat{x}$ containing the $\tilde{x}$ elements reordered: $\hat{x}_n^p = \tilde{x}_{h_n^p}^p$.

5) $\forall n \in \{1,..,N\}$ member of $\hat{x}$, use the vector $\hat{x}_n = (\hat{x}_n^1, ..., \hat{x}_n^P)^\top$ to predict $\hat{z} = \max_{k \in \{1,...,K\}} \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}_{\{\phi(\hat{x}_n, W) = k\}}$.

ϕ is a thresholding function defined by the user to construct the weather types and explicit in the A.case study part II. The complete method will be called RFR/SimSS in the evaluation of part III.

III. EVALUATION

The performances of the classification algorithms are compared using classical scores like the accuracy, the precision, the recall and the F_1 measure [21].

In the Figure 1, RFC algorithm has the best accuracy mean, while RFR/SimSS method gets a lower accuracy mean. The MLR algorithm has an equivalent accuracy mean with the raw ensemble.

The assessment of the recall and precision scores of the raw ensemble prediction show a high recall, lower precision for good Weather and high precision and lower recall for Rainfall. These scores are suggesting that the raw ensemble got a tendency to overestimate the good weather while the rainfall is underestimated.

The same evaluation of these scores for the MRL and RFC algorithms presents an increase of the recall and a decrease of the precision for the rainfall and rainfall with wind. The increase of recall implies a better detection of these weather types. But, the decrease of precision shows an augmentation of the false positive prediction error. These results suggest an overestimating of these classes for MLR and RFC models. The recall and precision of RFR/SimSS method reflect the same and stronger result with the Windy weather types.

While the good weather and windy type show for MLR and RFC models an increase of precision and a decrease of the recall for all methods. The modification of recall and precision implies an underestimating of these weather types. Moreover, RFR/SimSS model shows the same results stronger and with other weather types added: Rainfall with wind.

Compare to the 5 days ensemble forecast, in the Figure 2, the 10 days ensemble displays a global decrease on all classification scores due to the increase of the uncertainties of the numerical weather prediction system. In the accuracy score, the same results as above are noticed, the RFC got the highest accuracy. Nevertheless, RFR/SimSS has a higher difference with the models and raw accuracy score than with the 5 days ensemble forecast.

The recall of RFR/SimSS is very low for good weather and Rainfall with wind classes and high for the rainfall class. The RFC and MLR models are more accurate to predict the good weather and rainfall with wind than RFR/SimSS and raw ensemble. The RFR/SimSS highly overfits the rainfall class and MLR and RFC overfit the good weather and rainfall with wind classes.

After an analysis on the importance of features of the model with an ensemble forecast at 5 days and 10 days (not shown here), a set of few features has shown a particular interest. The median and mean of the 10 days forecast ensemble of the WS are the main features of the RFC model and improve the classification score of the good weather and rainfall with wind but hugely decrease the score of the rainfall and windy classes. Or, these features got less impact on the calibration random forest. Also, for the MLR model, these features got less impact for the rainfall and windy classes with their β coefficients associated are set close to zero. The difference between the performance of the windy class prediction for RFC and MLR is explained by the uses of variance, IQR,

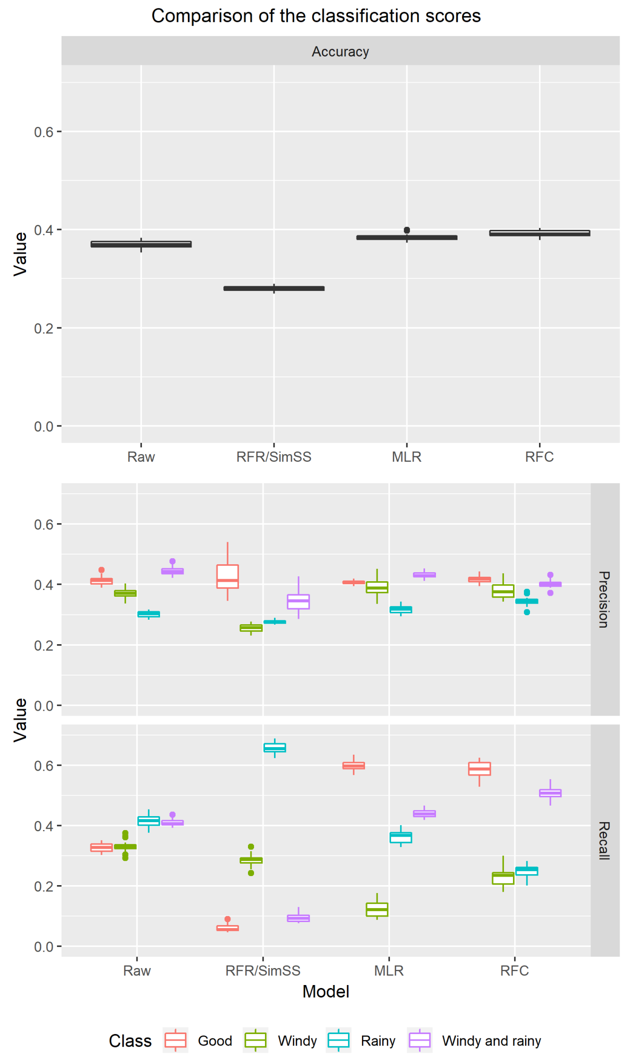

Fig. 2. Classification scores of models on Rennes data set with ensemble forecast at 10 days (10d) of forecasting horizon estimated by cross validation.

skewness of the Precip ensemble as features for the RFC model.

The temporal factor hour for 10 days ensemble forecast is mainly used by direct classification models and not by calibration random forest. In the direct classification models, this factor highly interferes with the good weather and rainfall classes.

IV. Concluding remarks

Overall classification scores shown the random forest classifier obtains the best results for 5 days and 10 days ensemble forecast at Rennes. The classification obtained from a multivariate calibration revealed an interesting high capacity of detection of the rainfall.

Where direct classification models follow raw ensemble performances and are better for good weather and rainfall with wind.

Between forecast ensemble at 5 days and 10 days forecasting range, few models performances differences are noticed. The increasing of weather types overfitting of the models with the increasing of the global weather types prediction error represents one of the main objectives to correct in future work.

The method of classification obtained from a multivariate calibration uses the nonlinear approach of the random forest to calibrate. In the work of Taillardat, this model has been compared to linear approaches for short-range. A comparison of linear approach on the wind speed and cumulative rainfall ensemble at 5 days and 10 days is needed. The recent EMOS models of [22] for calibration of wind speed and [23] for calibration of precipitation will be applied.

Other ensemble with a shorter medium-range (3 days) will be tested and compared to assess the classification results obtained at 5 days and 10 days. Also, the weather types prediction problem needs to be investigated on other spatial locations.

ACKNOWLEDGEMENTS

This research was supported by funding from the compagny Scalian and the mathematical research institute of Rennes 1.

REFERENCES

[1] P. Bougeault, Z. Toth, C. Bishop, B. Brown, D. Burridge, D. H. Chen, B. Ebert, M. Fuentes, T. M. Hamill, K. Mylne, *et al.*, "The thorpex interactive grand global ensemble," *Bulletin of the American Meteorological Society*, vol. 91, no. 8, pp. 1059–1072, 2010.

[2] Y.-Y. Park, R. Buizza, and M. Leutbecher, "Tigge: Preliminary results on comparing and combining ensembles," *Quarterly Journal of the Royal Meteorological Society*, vol. 134, no. 637, pp. 2029–2050, 2008.

[3] T. M. Hamill and S. J. Colucci, "Verification of eta–rsm short-range ensemble forecasts," *Monthly Weather Review*, vol. 125, no. 6, pp. 1312–1327, 1997.

[4] T. M. Hamill and J. S. Whitaker, "Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application," *Monthly Weather Review*, vol. 134, no. 11, pp. 3209–3229, 2006.

[5] T. Gneiting, A. E. Raftery, A. H. Westveld III, and T. Goldman, "Calibrated probabilistic forecasting using ensemble model output statistics and minimum crps estimation," *Monthly Weather Review*, vol. 133, no. 5, pp. 1098–1118, 2005.

[6] M. Taillardat, O. Mestre, M. Zamo, and P. Naveau, "Calibrated ensemble forecasts using quantile regression forests and ensemble model output statistics," *Monthly Weather Review*, vol. 144, no. 6, pp. 2375–2393, 2016.

[7] S. Scher and G. Messori, "Predicting weather forecast uncertainty with machine learning," *Quarterly Journal of the Royal Meteorological Society*, vol. 144, no. 717, pp. 2830–2841, 2018.

[8] J. B. Bremnes, "Constrained quantile regression splines for ensemble postprocessing," *Monthly Weather Review*, vol. 147, no. 5, pp. 1769–1780, 2019.

[9] R. Schefzik and A. Möller, "Ensemble postprocessing methods incorporating dependence structures," in *Statistical Postprocessing of Ensemble Forecasts*, pp. 91–125, Elsevier, 2018.

[10] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[11] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of statistical software*, vol. 33, no. 1, p. 1, 2010.

[12] E. Directorate, "Describing ecmwfs forecasts and forecasting system," *EcMWF Newsletter*, vol. 133, pp. 11–13, 2012.

[13] M. Leutbecher and T. N. Palmer, "Ensemble forecasting," *Journal of Computational Physics*, vol. 227, no. 7, pp. 3515–3539, 2008.

[14] F. Molteni, R. Buizza, T. N. Palmer, and T. Petroliagis, "The ecmwf ensemble prediction system: Methodology and validation," *Quarterly journal of the royal meteorological society*, vol. 122, no. 529, pp. 73–119, 1996.

[15] M. Courbariaux, *Contributions statistiques aux prévisions hydrométéorologiques par méthodes densemble*. PhD thesis, Université Paris-Saclay, 2017.

[16] L. Breiman, J. Friedman, R. Olshen, and C. Stone, "Classification and regression trees. wadsworth int," *Group*, vol. 37, no. 15, pp. 237–251, 1984.

[17] M. Taillardat, A.-L. Fougères, P. Naveau, and O. Mestre, "Forest-based and semi-parametric methods for the postprocessing of rainfall ensemble forecasting," *Weather and Forecasting*, no. 2019, 2019.

[18] N. Meinshausen, "Quantile regression forests," *Journal of Machine Learning Research*, vol. 7, no. Jun, pp. 983–999, 2006.

[19] M. Clark, S. Gangopadhyay, L. Hay, B. Rajagopalan, and R. Wilby, "The schaake shuffle: A method for reconstructing space–time variability in forecasted precipitation and temperature fields," *Journal of Hydrometeorology*, vol. 5, no. 1, pp. 243–262, 2004.

[20] R. Schefzik, "A similarity-based implementation of the schaake shuffle," *Monthly Weather Review*, vol. 144, no. 5, pp. 1909–1921, 2016.

[21] C. Rijsbergen, "v.(1979)," *Information retrieval*, vol. 2, 1979.

[22] S. Baran and S. Lerch, "Mixture emos model for calibrating ensemble forecasts of wind speed," *Environmetrics*, vol. 27, no. 2, pp. 116–130, 2016.

[23] M. Scheuerer and T. M. Hamill, "Statistical postprocessing of ensemble precipitation forecasts by fitting censored, shifted gamma distributions," *Monthly Weather Review*, vol. 143, no. 11, pp. 4578–4596, 2015.