

Weather types prediction at medium-range from ensemble forecasts

Gabriel Jouan, Anne Cuzol, Valérie Monbet, Goulven Monnier

▶ **To cite this version:**

Gabriel Jouan, Anne Cuzol, Valérie Monbet, Goulven Monnier. Weather types prediction at medium-range from ensemble forecasts. 9th International workshop on Climate Informatics, Oct 2019, Paris, France. hal-02425230v2

HAL Id: hal-02425230
https://hal.science/hal-02425230v2

Submitted on 2 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

9th International Workshop on Climate Informatics
October 02-04, 2019
Hosted by École Normale Supérieure, Paris, France

WEATHER TYPES PREDICTION AT MEDIUM-RANGE FROM ENSEMBLE FORECASTS

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

***Abstract*—Medium-range weather forecasts can be of high economic value in many fields: agriculture, renewable energy production, maintenance operations planning. Such forecasts can be based on ensembles derived from weather models, and the postprocessing of such ensembles is an active research problem in the statistical weather community. In this work, we try to face the problem of long forecasting horizons, and focus on the multivariate case where different meteorological variables interact. The prediction problem is simplified and defined as the prediction of a weather type, which is a categorical variable defined by the interaction of the meteorological variables. We use machine learning techniques to predict this weather type from the multivariate ensemble forecasts. The algorithms are applied to a 5 to 10 days weather forecasting in the north-west of France, based on wind and precipitation data from the ECMWF ensemble system.**

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 postprocessing has been studied these last years. One of the most common approach to calibrate the ensemble for one given variable is based on a regression model which helps to predict observations of the variable given a description of the ensemble as input. For example, the state-of-the-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 techniques have also been developed in order to reproduce dependencies between variables [9].

Corresponding author: G Jouan, gabriel.jouan@scalian.com
[1]Scalian Alyotech Rennes; [2]Univ. Bretagne-Sud, LMBA; [3]Univ. Rennes, IRMAR

In this article, we focus on multivariate forecasting for horizon higher than 3 days. Such medium to long-range forecasts are of high value, for instance for maintenance operations in many fields, but this problem is known to be difficult. In this work, the goal will be to predict weather types from ensemble forecasts, instead of performing a calibration of the whole multivariate distribution of the meteorological variables of interest. The weather type is a categorical variable, described for instance by ”good”, ”windy”, ”rainy”. Such qualitative information is sufficient for many applications.

A natural approach to predict such weather types is to apply direct classification algorithms. A state-of-the-art non-linear method is the random forest classifier (RFC) [10] based on the aggregation of tree classifiers [11]. To compare random forest results on multiple weather types classification, a linear approach can be used: the multiple logistic regression, also called multinomial lasso regression (MLR) [12].

Such direct classification algorithms applied to the ensemble forecasts jointly perform a correction and a classification, which can be difficult. An other way to solve the weather type prediction problem is to perform a multivariate calibration, followed by a transformation of the output into weather types. The multivariate calibration consists in applying independent univariate calibrations for each variable, followed by a reordering method [13]. This approach will be used as comparison.

The paper is organized as follows. In Section II, we describe the classification algorithms used for the prediction of weather types. In Section III part A, the considered data are introduced and the weather types are defined. In part B, the performances of the proposed methods are compared for a forecasting range of 5 days and 10 days, for a chosen location in the north-west of France. The paper ends with some concluding remarks in Section IV.

II. METHODOLOGY FOR THE PREDICTION OF WEATHER TYPES

Our aim is to calibrate multivariate medium-range forecast ensembles for one location. Since qualitative information is sufficient for some applications, we propose to tackle the problem of weather types prediction, each weather type being defined from several meteorological variables. For instance, the weather types can be defined as "good", "windy", "rainy", "windy and rainy". Our contribution will be to propose classification algorithms where the inputs are given by a multivariate ensemble and the output is the weather type.

To solve the classification problem, two approaches will be considered. The first one is based on a direct application of machine learning classification algorithms (Section A). The second one consists in applying the weather type definition to the output of classical calibration methods (Section B).

The ensemble members can not be used directly as inputs of the machine learning algorithm because they are exchangeable [14]. This means that ensemble members are invariant under permutation, and consequently can not be used as predictors. Then, following [6], [15], 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 precipitation probabilities. 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, allow to take 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 to the inputs the corresponding weather types computed from the raw (uncalibrated) ensembles.

A. *Direct classification*

Two classical machine learning algorithms are considered: random forest and penalized multinomial regression. The random forest is known to be more flexible and multinomial regression more robust.

Random forest classifier (RFC) was proposed by [11] and [10]. It combines elementary classification trees, learned on random samples generated from the data, 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 leaf. At each step of the algorithm, the current leaf is splitted into two parts if it improves the Gini impurity. At the end, the probabilities of weather types are obtained from the mean of all trees.

Penalized multinomial regression (MLR) is described for instance in [12]. Each multinomial regression model predicts the probability of one of the K weather types against the others so that $K-1$ models are fitted. We expect that all input variables are not of the same importance to help to discriminate the weather types. So, a Lasso penalty is introduced in the estimation task which 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 useless inputs and to lead to a more robust prediction tool.

B. *Classification from multivariate calibration*

The direct machine learning classification algorithms proposed in this paper has to be compared to other machine learning solutions proposed in the ensemble forecast calibration literature, in particular [6] which also used random forests. In [6] and reference therein, the authors perform univariate calibration with quantile regression forests [16].

Here, we propose to apply a quantile regression forest separately for the calibration of each meteorological variable (wind speed and precipitation for instance). Then the two independent calibrated ensembles are combined by a Schaake Shuffle (SS) algorithm [13] to reproduce the dependent structure existing between variables. In this reordering algorithm, the marginal postprocessings are combined to reproduce the empirical copula estimated from past observations. One recent improvement, referred to as SimSchaake ([17], [9]), proposes to combine the 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 this reordering procedure, the multivariate output is transformed into the predefined weather types.

Methods have been applied using the R software with the "randomForest" [18], "glmnet" [19] and "quantregforest" [20] packages for the RFC, MLR and quantile regression forests algorithms.

III. APPLICATION

The classification algorithms are now applied to data from the north-west of France (city of Rennes), described in Section A. The performances of the proposed methods are then compared in Section B.

A. Data description and weather types definition

Ensemble forecast data of the ECMWF [21] are collected from the Thorpex interactive grand global ensemble archive (TIGGE) [2], [1]. The TIGGE archive includes a minimum of 10 ensemble forecasting systems on a time-period from 2008 to 2018.The ECMWF ensemble system is composed of 50 exchangeable ensemble members generated from the Ensemble of Data Assimilations (EDA) based perturbations with singular vectors in the initial conditions and stochastic physics models [22], [23].

Collected data are composed of observations and ensemble forecasts of precipitation (Precip, mm) and wind speed (WS, m.s^{-1}) at forecasting range 5 days and 10 days, two runs (6 am and 6 pm) for the French city of Rennes.

As mentioned earlier, the continuous variables are transformed to define weather types. $K = 4$ balanced weather types are chosen. The data contains approximately the same number of observations in each weather type avoiding an unbalanced classification problem. For an observation vector $y = (y^{Precip}, y^{WS})^\top$, and the set of thresholds $\{0.02, 2.8, 4\}$, the ϕ thresholding function is defined as:

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

The four weather types are referred to as "good" if $\phi(y) = 1$, "windy" if $\phi(y) = 2$, "rainy" if $\phi(y) = 3$ and "rainy and windy" if $\phi(y) = 4$. Other thresholds could be chosen depending of the application in mind.

B. Results

The classification algorithms performances are evaluated and compared using classical scores like the accuracy, the precision and the recall [24]. All the scores are computed by cross-validation. For that, the data set is randomly splitted into 2 subsets. The validation subset contains 912 days, randomly extracted from the period 2014-2018. The learning subset is composed of all remaining days over the period 2008-2018. This is repeated 30 times in order to approximate the distribution of the scores.

A k-fold cross-validation has been performed ($k = 10$), and the penalization hyperparameter of the MLR model has been fixed to 0.03.

The reference result that we seek to improve is the forecast obtained from the uncalibrated multivariate ensemble transformed into weather types following rules (1). This forecast will be referred to as "Raw" in the sequel.

Figure 1 shows the accuracy score on the left panel. It is an overall criteria which is close to 1 if the weather types are correctly classified. The accuracy is close to 0.6 for all the methods, but the direct classification algorithms (MLR and RFC) slightly improve both the "Raw" result and the classification based on multivariate calibration.

The right panel of Figure 1 shows the precision and the recall which allow to analyze results per weather type. Precision (also called positive predictive value) is the fraction of relevant instances among the retrieved instances, while recall (also known as sensitivity) is the fraction of relevant instances that have been retrieved over the total amount of relevant instances. We want to achieve a good compromise between precision and recall for each weather type.

We can see on Figure 1 that the "Raw" forecast at 5 days leads to the highest recall for the "good" weather type, but with the lowest precision. On the other hand, it leads to the lowest recall for the "rainy" class, while having the highest precision.

Compared to the "Raw" result, the RFR/SimSS method improves the recall of the "rainy", but deteriorates the precision. On the other hand, the direct classifications (MLR and RFC) show an improvement of recall for the "rainy" and "windy and rainy" classes, while maintaining a precision close to the "Raw" result for these two weather types. Note that this gain in recall is higher for RFC, but with a slight decrease in precision compared to MLR.

It can be observed that all methods lead to a comparable precision for the "windy" class, but none of the methods is able to improve the recall obtained by the "Raw" ensembles for this weather type.

Figure 2 shows the forecast results for a horizon of 10 days. A global decrease of all classification scores can be observed, due to the increase of the uncertainties of the numerical weather prediction system. However, the RFC method is still leading to the highest accuracy for this long-range forecasting problem. One interesting point is that the classification obtained from the multivariate calibration (RFR/SimSS) is not able to improve the "Raw" result obtained from uncalibrated ensembles.

The results of the precision and recall for 10 days ensemble forecast display higher variations between the weather types than the 5 days ensemble forecast.

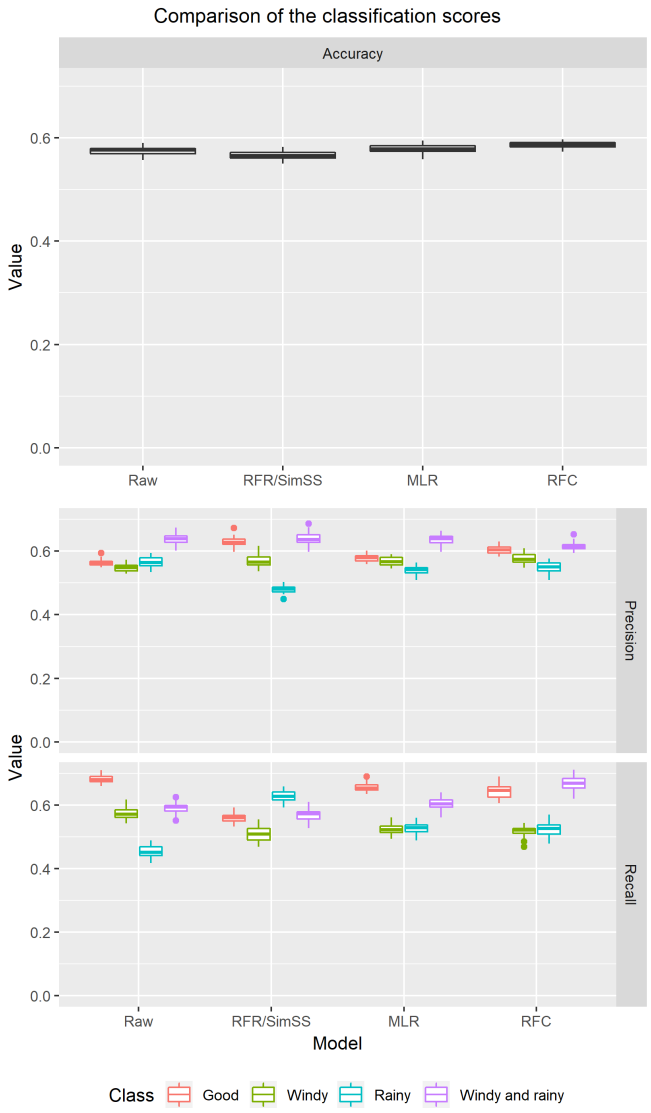


Fig. 1. Classification scores for a horizon of 5 days. Raw: Forecast from uncalibrated multivariate ensemble; RFR/SimSS: Classification obtained from a multivariate calibration; MLR: Multinomial lasso regression; RFC: Random forest classifier. Left: Accuracy scores; Top-right: Precision scores for each weather type; Bottom-right: Recall scores for each weather type.

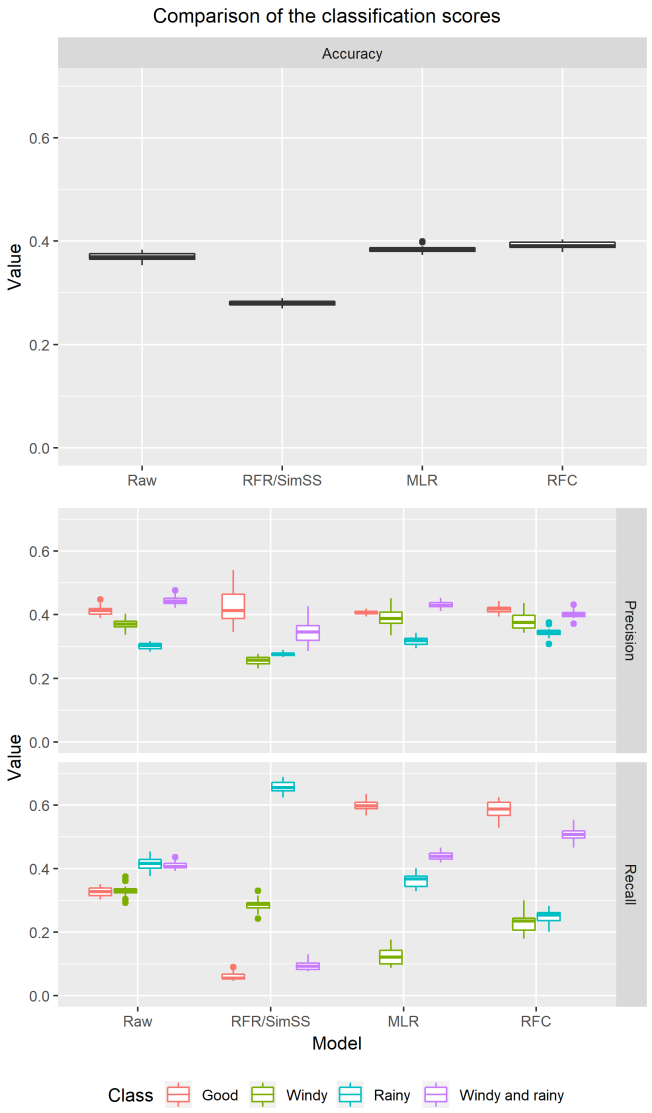


Fig. 2. Classification scores for a horizon of 10 days. Raw: Forecast from uncalibrated multivariate ensemble; RFR/SimSS: Classification obtained from a multivariate calibration; MLR: Multinomial lasso regression; RFC: Random forest classifier. Left: Accuracy scores; Top-right: Precision scores for each weather type; Bottom-right: Recall scores for each weather type.

In the right panel of the Figure 2, RFR/SimSS model is highly overestimating the ”rainy” type and shows a poor probability of detection of ”good” and ”windy and rainy” types. Direct classification algorithms (MLR and RFC) lead to the best recall for ”good” and ”windy and rainy”, while maintaining a comparable level of precision. However, there is a decrease of performance over the ”Raw” result for the ”windy” and ”rainy” weather types.

IV. CONCLUDING REMARKS

Compared to the reference result obtained from the uncalibrated multivariate ensemble of forecasts, direct classification models lead to a small improvement in prediction accuracy. This is not the case for the classification obtained after a multivariate calibration. However, the study of precision and recall scores show that this improvement is not observed for all weather types. For instance, direct classification models increase the probability of detection of ”rainy” and ”windy and rainy” weather types for a horizon of 5 days.

For a longer horizon (10 days), these models lead to better detections for "good" and "rainy and windy" weather types. A study of variables importance in the classification models (not shown in this paper) can help understanding the differences in performance between weather types.

In [6] and [15], the quantile regression forest has been compared to linear approaches for short-range ensemble forecast. A comparison of linear approaches on the wind speed and cumulative rainfall with ensemble forecast at 5 days and 10 days is needed. The recent EMOS models of [25] for calibration of wind speed and [26] for calibration of precipitation will be applied.

Other ensembles 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 Scalian group and IRMAR.

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

[12] 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.

[13] 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.

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

[15] 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.

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

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

[18] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.

[19] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for cox's proportional hazards model via coordinate descent," *Journal of Statistical Software*, vol. 39, no. 5, pp. 1–13, 2011.

[20] N. Meinshausen, *quantregForest: Quantile Regression Forests*, 2017. R package version 1.3-7.

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

[22] R. Buizza, M. Leutbecher, and L. Isaksen, "Potential use of an ensemble of analyses in the ecmwf ensemble prediction system," *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, vol. 134, no. 637, pp. 2051–2066, 2008.

[23] R. Buizza, "Weather prediction in a world of uncertainties: should ensembles simulate the effect of model approximations?," in *ECMWF/WWRP Workshop: Model Uncertainty*, (ECMWF, Reading), 05/2016 2016.

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

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

[26] 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.