GEOFIZIKA VOL. 40 2023

DOI: https://doi.org/10.15233/gfz.2023.40.5
Original scientific paper

Application of the Random Forest method on the observation dataset for visibility nowcasting

David Sládek

Department of Military Geography and Meteorology, University of Defence, Brno, Czech Republic

Received 16 November 2022, in final form 15 May 2023

Accurate visibility forecasting is essential for safe aircraft operations. This study examines how various configurations of the Random Forest model can enhance visibility predictions. Preprocessing techniques are employed, including correlation analysis to identify fundamental relationships in weather observations. Time-series data is transformed into a regular Data Frame to facilitate analysis. This study proposes a classification framework for organizing visibility data and phenomena, which is then used to develop a visibility forecast using the Random Forest method. The study also presents procedures for hyperparameter tuning, feature selection, data balancing, and accuracy evaluation for this dataset. The main outcomes are the Random Forest model parameters for a three-hour visibility forecast, along with an analysis of errors in low visibility forecasts. Additionally, models for one-hour forecasts and visibility forecasting under precipitation are also examined. The resulting models demonstrate a deterministic forecast accuracy of approximately 78%, with a false alarm rate of around 6%, providing a comprehensive overview of the capabilities of the Random Forest model for visibility forecasting. As anticipated, the model demonstrated limitations in accurately simulating fast radiative cooling or abrupt decreases in visibility caused by precipitation. Specifically, in relation to precipitation, the model achieved an accuracy of 79%, yet exhibited a false alarm rate of 19%. Additionally, this method sets a foundation for enhancing prediction accuracy through the inclusion of supplementary forecast data, while its implementation on real-world datasets expands the reach of machine learning techniques to the members of the meteorological community.

Keywords: aviation meteorology, visibility forecasting, nowcasting, landing forecast (trend), machine learning, random forest, feature selection, hyperparameters tuning

1. Introduction

Visibility forecasting is one of the fundamentals of aviation meteorology. It is included in forecasts, observations and warning information. Very often, fore-

casters cannot rely on numerical models and have to use their own forecasting procedure. Thus, visibility provides great space for testing different forecasting approaches and developing new tools for forecasting improvement.

The visibility forecast is issued by meteorologists in accordance with ICAO Annex 3. It is included in both longer-term TAF forecasts (valid for 9, 24 or 30-hour period) and shorter-term forecasts, *e.g.* landing 'TREND' forecasts (valid for 2 hours from the time of the report which forms part of the landing forecast) (ICAO, 2010). It is the possibility of improving these short-term forecasts that motivates this research.

Examination of the quality of aviation landing forecasts in recent years has shown that these forecasts are not at the forefront of forecasters' minds (Sládek, 2021). The work reveals that the quality of TREND forecasts corresponds to the persistent forecasting. While its accuracy is high, the skill of the forecast is minimal. Thus, a product that could alert forecasters to an impending change would undoubtedly find use at any forecast station. Here, several options can be considered for such a short (nowcasting) interval.

In the area of visibility prediction, several studies have been published that have taken different approaches to prediction. A study from the area of eastern Canada focused on more accurate parameterization within numerical models (Gultepe et al., 2006). This work relies on very advanced measurement and observation methods that may not be standard worldwide.

In general, the use of complex statistical methods that are combined into Machine Learning models is now considered a very effective way of improving forecast accuracy (Patriarca et al., 2022). Although more for capturing local patterns that can be characterized using meteorological data, rather than as a full replacement for numerical models (Schultz et al., 2021).

An approach using deep learning methods was chosen by the author team to predict visibility at Urumqi international airport based on commonly observed variables (Zhu et al., 2017). The work concludes that the method is quite successful, but especially the prediction of the time of deterioration and improvement of the situation is very difficult. Similar problems were also encountered in another neural network application, for tested intervals of 3, but more notably for intervals of 6 or 8 hours (Deng et al., 2019).

International research team focused on Rio de Janeiro Airport (Cordeiro et al., 2021) proposed ML methods to predict very short-term changes at the airport using more frequent measurements within the airport, radar measurements and other inter-hourly data collected. Their paper focuses on forecasts of ceiling and visibility shorter than one hour and confirms the usefulness of ML methods as a complement to NWP.

If we consider the contribution of each work, it is undeniable that they achieved certain results within their domain. However, it is important for aviation to have a good control over the progress of machine learning and to know

the reason why errors occur in the results. Thus, it seems reasonable that we should rather focus on supervised methods in research.

Among the well-known statistical models, the authors Salman and Kanigoro used the ARIMA model to predict visibility. In their conclusions, while they admit that work can continue in this direction by testing different variants, they rather see an opportunity in further developments in the field of neural networks (Salman and Kanigoro, 2021).

From a forecasting perspective, the use of decision tree-based methods also makes sense, as a decision tree is something that weather forecasters often use, consciously or unconsciously, when making forecasts. A very compelling argument for the use of the decision tree method is, for example, the work of F. Wantuch from Hungarian Meteorological Service, who presents that his proposed method is already in operational use (Wantuch, 2001).

One suitable method using the principle of decision trees is the Random Forest method. The method conveniently uses groups of decision trees and returns the most likely possible solution. The advantages of this method can also be used, for example, to predict ground characteristics such as soil temperature or greenhouse water content (Tsai et al., 2020). From a purely meteorological perspective, it can be used to forecast complex processes and influential phenomena such as flood forecasting (Schumacher et al., 2021) or extensive rainfall (Hill and Schumacher, 2021). However, both of the aforementioned works have also applied this method to NWP outputs.

Since the above papers demonstrate that there are suitable methods that can create room for improved prediction, it is reasonable to ask how will the machine learning methods perform on only observational data? How will they be successful if they only "know" the conditions at a given location?

The answer to this question may provide some new insights:

1. a new perspective on the statistical treatment of visibility observations at weather stations,

2. the basic accuracy of the machine learning model, which should improve as more data is added,

3. the identification of complex situations where changes occur that are statistically unlikely, and where human error in forecasting may also occur,

4. identifying the limits of the chosen method on the dataset used.

As mentioned, a method that is very close to human decision making even in operational practice is the decision tree-based method. It is one of the reasons why Random Forest was chosen as the main one. The goal of this paper is not only to establish a solid foundation for forecasting using machine learning, but also to show a different approach to it. Priority will be given to the prediction of dangerous phenomena over the accuracy of the whole model for insignificant values. The paper also aims to convey insight into the random forest method to

forecasters just on the prediction of this crucial element for them. After all, it is forecasters who will – without exaggeration – depend on ML products in the near future.

2. Data and methods

As a main dataset, SYNOP reports (WMO, 2019) with hourly frequency were used. These reports follow the same format worldwide as defined by World Meteorological Organization (WMO). It contains of numerous valuable information, such as visibility, wind speed, temperature, dew point, pressure, precipitation sum, observed phenomena, etc. Advantage of using SYNOP reports is its standard quality, as methodology of observation does not significantly differ through time or station by station. Therefore, method can be used worldwide with a minor calibration, mainly in data preprocessing (units change, data cleaning, etc.).

The machine learning model will be implemented and tested on data from synoptic observations from the station at Brno-Tuřany Airport (WMO: 11723, ICAO: LKTB) from January 1998 to October 2019 (Fig. 1).

As with all research, at least a basic exploratory analysis will need to be done. However, it cannot be assumed that the basic meteorological elements would be significantly different from the normal daily or annual climatic pattern, or that their distribution would be significantly different from the climatic pattern. The distribution of meteorological phenomena that are probably most important for visibility prediction (and visibility itself) is depicted in the Figure 2.

A Spearman correlation matrix was calculated to confirm the assumptions about the factors influencing visibility. The influence of these factors was assumed and derived from experience and physical laws, *e.g.* relative humidity, weather phenomena, precipitation, wind speed, cloud height, etc. Although not all of these factors need to directly affect visibility, an empirical link is assumed

Figure 1. Schematic figure of the station position.

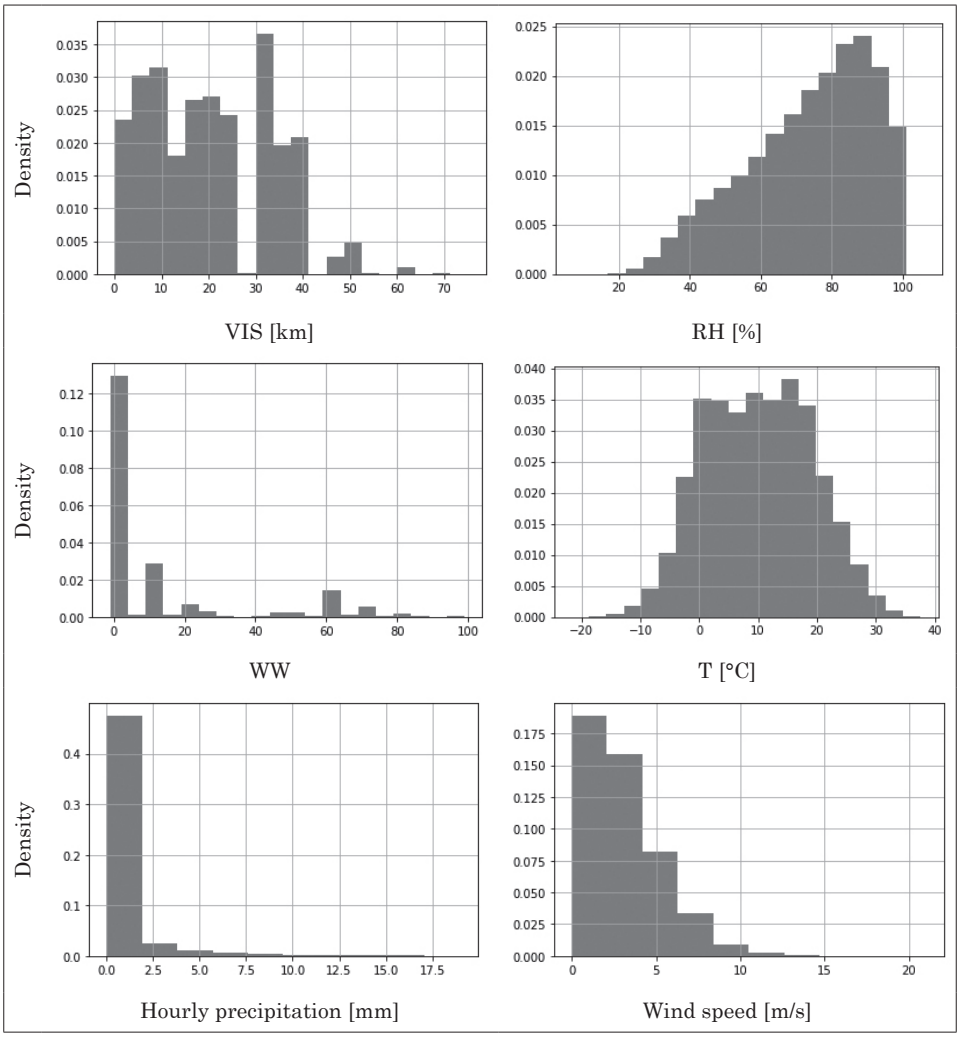


Figure 2. Histograms of visibility (*VIS*), relative humidity (*RH*), present weather (*WW*), temperature (*T*), hourly precipitation sums and wind speed from observations from Brno-Tuřany station (1998–2019).

between some of them. Some of the inappropriate parameters have been removed from the matrix depicted by Tab. 1.

Spearman correlation showed a relatively high dependence between visibility and temperature, probably due to thermodynamic principles and radiative cooling of air and condensation of water vapour. The correlation with relative humidity is as expected, as well as with wind speed. On the contrary, a higher value of correlation might be expected to precipitation. A higher correlation coef-

Table 1 Spearman correlation matrix for relevant features (from left to right: T – temperature, Td – dew point temperature, RH – relative humidity, vis – visibility, dd – wind direction, ff – wind speed, n – cloud cover, ww – phenomena, r1 – precipitation in recent hour, sr6 – precipitation in recent 6 hours, u, v wind components, WWL – duration of current phenomena)

	T	*Td*	*RH*	*vis*	*dd*	*ff*	*n*	*ww*	*r1*	*sr6*	*u*	*v*	*WWL*
T	1.00	0.89	–0.54	0.50	0.01	0.09	–0.27	–0.34	0.01	0.01	–0.04	–0.36	1.00
Td	0.89	1.00	–0.14	0.26	0.01	–0.06	–0.13	–0.15	0.00	0.01	–0.08	–0.21	0.89
RH	–0.54	–0.14	1.00	–0.67	0.01	–0.36	0.40	0.54	–0.02	0.00	–0.10	0.46	–0.54
vis	0.50	0.26	–0.67	1.00	0.13	0.24	–0.35	–0.58	0.01	–0.22	0.07	–0.59	0.50
dd	0.01	0.01	0.01	0.13	1.00	–0.02	0.09	0.03	0.00	–0.73	0.08	–0.06	0.01
ff	0.09	–0.06	–0.36	0.24	–0.02	1.00	0.07	–0.05	0.02	0.02	0.23	–0.10	0.09
n	–0.27	–0.13	0.40	–0.35	0.09	0.07	1.00	0.53	0.03	–0.12	–0.07	0.35	–0.27
ww	–0.34	–0.15	0.54	–0.58	0.03	–0.05	0.53	1.00	0.06	–0.05	–0.02	0.66	–0.34
r1	0.01	0.00	–0.02	0.01	0.00	0.02	0.03	0.06	1.00	–0.02	–0.01	0.00	0.01
sr6	0.01	0.01	0.00	–0.22	–0.73	0.02	–0.12	–0.05	–0.02	1.00	–0.12	0.08	0.01
u	–0.04	–0.08	–0.10	0.07	0.08	0.23	–0.07	–0.02	–0.01	–0.12	1.00	–0.05	–0.04
v	–0.36	–0.21	0.46	–0.59	–0.06	–0.10	0.35	0.66	0.00	0.08	–0.05	1.00	–0.36
WWL	1.00	0.89	–0.54	0.50	0.01	0.09	–0.27	–0.34	0.01	0.01	–0.04	–0.36	1.00

ficient with duration of ongoing meteorological events (*WWL*) is very promising. For the correctness, it should be noted that the correlation of non-quantitative values (*ww*) is not standard, arguably not statistically correct. However, this correlation has shown that the code itself can correlate with visibility and provide a suggestion for classifying phenomena. For completeness, p-values were also calculated to test statistical significance (at significance level of 0.05) and for none of the visibility variables did the probability value exceed 1%. Therefore, none of the predictors will be dropped in this phase.

2.1. Data preparation

First, the data had to be cleaned, or missing or wildcard values (*e.g.* 'X', '/') had to be replaced with NaN (short for "Not a Number" in python). As the following table (Tab. 2) indicates, the data is very consistent and does not contain many records with missing values

Table 2. Basic counts and characteristic of chosen database series of overall count – 189589 observations (VIS - visibility, T - temperature, RH - relative humidity, dd – wind direction, ff – wind speed, r1 – hourly precipitation sum, n – cloud coverage, h – cloud height).

	VIS	*T*	*RH*	*dd*	*ff*	*r1*	*n*	*h*
NaN	9	9	9	13	10	46	23	23
Range	0/75	–21/37	14/100	0/360	0/21	0/60	0/8 (9)	0/21000
Units	km	°C	%	°	m/s	mm/h	octas	m

The dataset contains many other columns corresponding to the data contained in the SYNOP report (WMO, 2019). There the counts of the redundant data were higher. Columns that are only included in certain terms (*e.g.*, minimum or maximum daily temperature, 24-hour precipitation, etc.) have also been removed.

2.1.1. Wind handling

While the value for wind speed is unambiguous, there is one problematic value for direction, particularly the variable wind direction. It was necessary to translate the "variable" value in the SYNOP code "99" to some understandable value. Variable wind direction only accounts for 4.4% of the wind direction data. It is mostly associated with very low speeds (Tab. 3) therefore, using value of direction from preceding term would not cause significant bias.

Table 3. Occurrence counts of wind speeds by variable wind direction in examined dataset.

Wind speed [m/s]	Count	Relative
1	5570	68.7 %
2	2503	30.9 %
3	28	0.3 %
4	5	0.06 %

It should be noted, however, that some papers and online tutorials often may not address this problem. This is because they use average daily data which does not contain the variable value. Of course, this completely changes the research question, problem, or continuity of the data. For a correct assessment of wind direction and wind speed, the u and v wind components were used according to the formulas (1) and (2):

$$u = w_s \times \sin\varphi, \tag{1}$$

$$v = w_s \times \cos\varphi, \tag{2}$$

where w_s is wind speed and φ stands for wind direction.

The advantage of using u and v wind components should be the separate quantification of the influence of zonal and meridional flow. For example, during significant advection, using u and v components could yield better results. It is also much easier to process for ML methods, because in the original format, for example, the values 350 and 10 are very far apart, although they indicate similar directions.

2.1.2. Creating shifted values

Another issue addressed during the preparation was how to deal with the time series data and prepare it correctly for the algorithm. In the case of applications such as neural networks, we have tools for time series, or a suitable type,

called long-short term memory (LSTM) networks (Brownlee, 2017). In conventional machine learning methods, especially supervised learning, we will certainly want to see what the importance of each observation term for forecasting is, etc.

For each term, duplicate predictors set values were created from the terms three and four hours ago (Fig. 3). Thus, the values themselves were not transformed or subtracted in any way, but only copied from the term before the given time interval.

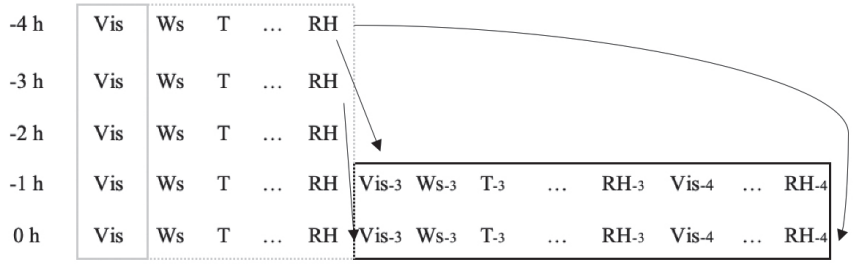


Figure 3. The process of assigning predictors (black frame) to a predicted value (gray frame), values in the grey dotted frame do not enter modelling process directly.

Theoretically, the problem could also be approached by classifying the trend of presumably important variables (*e.g.* relative humidity, temperature, wind speed…) and thus expressing the trend of the past terms well. However, such an approach could lose valuable information and the classification itself could degrade our results and we would hardly know the reason for the lower success rate.

The validity of such an approach can be anticipated with the help of statistics, *i.e.* Pearson's correlation coefficient of forecasting features and forecasted parameter (Tab. 4). This could suggest how significant the values from previous measurements will be.

The low correlation of precipitation, which did not reach more than 0.05, may be rather surprising. Moreover, it was experimentally tested by hourly rainfall, on a limited data set, with stratification by hourly rainfall, but even so a very low correlation was obtained.

Table 4. Pearson's correlation for visibility in time of forecasting and 3 and 4 hours in advance.

	T	*Td*	*RH*	*vis*	*dd*	*ff*	*n*	*ww*	*w1*	*w2*	*sr6*	*u*	*v*
vis	0.48	0.24	–0.63	1.00	0.03	0.22	–0.26	–0.39	–0.31	–0.40	–0.02	–0.24	0.05
visshift3	0.43	0.24	–0.53	0.84	0.00	0.15	–0.27	–0.30	–0.27	–0.34	–0.03	–0.19	0.07
visshift4	0.41	0.24	–0.49	0.79	0.00	0.12	–0.27	–0.28	–0.26	–0.32	–0.03	–0.18	0.07

2.1.3. Time handling (cyclic)

Since, especially with fog, the time of day or month of the year when the fog is formed is quite important, the time must be handled rationally. It seems sensible to create a cyclic dataset that helps to determine the hour of the day and month of the year so that December and January are not seen as the most distant numbers, but relatively close ones. Similarly, the 23rd hour of the day and the 1st hour of the day could be read wrong by the method. For these purposes, the goniometric functions sine and cosine applied to the number of seconds since the beginning of the day and hour of year work well (Fig. 4).

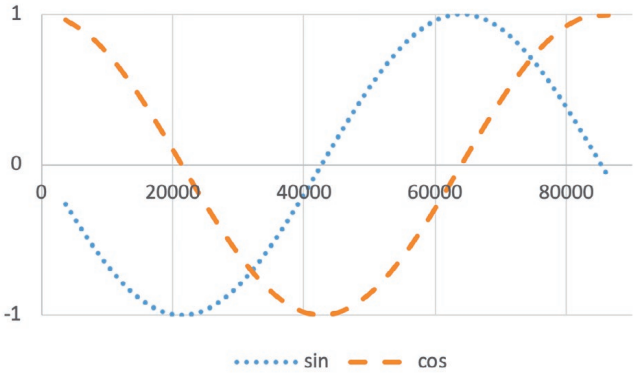


Figure 4. Cyclical time (sin and cos function applied to a second of the day or hour of the year).

The graph corresponds very suitably to the part of the day in terms of incoming solar radiation and daytime temperature, so this visualization shows the correct treatment of the time characteristics.

2.1.4. Visibility labelling

For testing purposes, two visibility category labels were chosen. One is subjective, chosen to see the effect of method parameters on arbitrarily chosen categories. The other refers to the criteria for issuing SPECI reports and classifying change groups according to ICAO Annex 3 (International civil aviation organization, 2010), which is a fundamental regulation in aviation meteorology.

The criteria set out in the aeronautical meteorology regulations, which specify under what conditions an exceptional observation report must be issued, can be considered formally correct. These criteria are shown for clarity in Tab. 5.

For the purpose of method tuning, subjective partitioning of the dataset according to visibility, or data labelling, was used. Values were chosen that are important in meteorology, *e.g.* instrument meteorological conditions (IMC),

Table 5. Visibility categories based on ICAO Annex 3 and their occurrence frequency in dataset.

Visibility	Category	Count in dataset
< 0.8 km	0	3815
0.8 – 1.5 km	1	2529
1.5 – 3.0 km	2	7119
3.0 – 5.0 km	3	13140
> 5.0 km	4	162940

Table 6. Description of test categories and their counts in the dataset.

Visibility	Label	Description	Count in dataset
< 1 km	0	Fog	4221
1 – 2 km	1	Severe mist	4218
2 – 3km	2	Moderate mist	5024
3 – 4 km	3	Moderate mist	6059
4 – 5 km	4	IMC threshold	7081
5 – 6 km	5	Reduced visibility within VMC conditions	5795
6 – 10 km	6	Mist by SYNOP report	25140
> 10 km	7	Visibility not reduced	132005

visual meteorological conditions (VMC) or fog (FG) or mist (BR) threshold. All these categories are listed with their counts in Tab. 6.

As can be seen from the table, categories were created in intervals of one kilometre, with only three categories above five kilometres. First, category 5 (5–6 km), to show how the model handles values between high visibility and mist in the resulting matrix. Next one, category 6 (6,015,010 km), where there are, for example, the colour code boundaries, the mist boundaries in the SYNOP report or boundaries of visibility observation (code 9999) in the METAR report (WMO, 2019). The last category contains visibilities above 10 km.

It is clear that in both cases of categorization we are dealing with an imbalanced dataset (counts of categories are very different). Therefore, since there is an imbalance in the target variable, it will be correct to focus on balancing techniques as well (Brownlee, 2019).

2.1.5. Phenomena classification and labelling

Different meteorological phenomena have different effects on visibility. To evaluate their influence within the RF method, it will undoubtedly be necessary to group them into categories. Without grouping, RF would assume that, for example, codes 52 and 54 are completely different, even though they indicate intermittent drizzle in both cases. For the purposes of this research, a classification of thirteen categories will be used (Tab. 7), which emphasizes primarily the characteristics of precipitation and its likely effect on visibility.

Table 7. Subjective categorization and labelling of phenomena used for RF modelling.

Label	Synop code range (WW)	Description
1	0–12	No precipitation
2	13–19	Precipitation in vicinity
3	20–29	Precipitation or fog in preceding hour, not during observation
4	30–39	Dust storm, sandstorm, drifting or blowing snow
5	40–49	Fog or freezing fog present
6	50–59	Drizzle
7	60–69	Rain
8	70–79	Solid precipitation
9	80–82	Rain showers
10	83–84	Sleet showers
11	85–90	Showers with prevailing solid precipitation
12	91–94	Thunderstorm within preceding hour
13	95–99	Thunderstorms

This categorization of phenomena could undoubtedly be improved. It would certainly be possible to experiment with different groups of phenomena and observe their influence on the modelling results. The greatest contribution can be expected with precipitation, *i.e.* the designation of its state.

Involvement of past weather code represented in SYNOP report by W1 and W2 was questionable. Their usefulness would probably be outweighed by unnecessary duplication of information. Indeed, in the RF method, highly correlated values or duplicate information reduce their impact on the modelling. Thus, here the effect of the current WW weather ensemble would be suppressed (Breiman, 2001), (Chase et al., 2022).

In the experimental phase, also one-hot encoding of phenomena was tested. Besides removing rare events, assigning low importance, one-hot encoding did not bring any significant improvement. Therefore, it was not used in final research.

2.2. Methods

The Random Forest method was chosen as the most suitable for visibility modelling. Firstly, it corresponds well with the procedure of the forecaster, who compares all possible values of the predictors and determines whether they are suitable for fog occurrence, or visibility improvement, etc.

The second motivation is data-based. Decision trees, and hence RF, can handle categorical and continuous variables well, which is useful in such a diverse dataset.

And the third reason was the testing that was done on a limited subset of data (3 years). Models of several traditionally used methods in their baseline calibrations were created and their success rates were visualized (comparison of

testing set of 3 years dataset in Appendix 2 – Methods comparison). According to the results, the random forest performed best in modelling.

The Random Forest consists of a random set of decision trees that can be compared to the forecaster's thought processes. After testing all sequences (trees), the result that came out the most often (majority vote) from all trees is selected (Fig. 5).

The way the nodes are split is the same as for ordinary decision trees. In this case, the default Gini index was chosen based on testing, since none of the options (gini, entropy, log_loss) recorded very different results. The Gini index for each node determines the probability of misclassification when randomly selected from the set. Thus, for example, if there are only values of one class in a node, the probability of misclassification will be zero (Pedregosa et al., 2011).

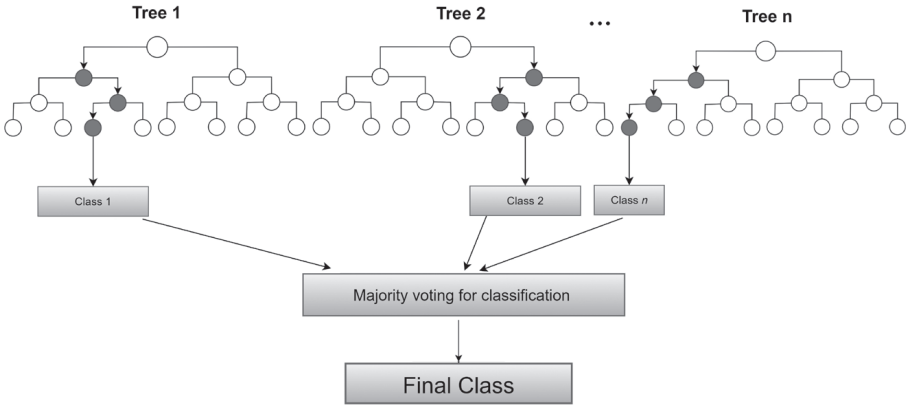


Figure 5. Schematic model of Random Forest method (Sruthi, 2021).

As creators of a specific RF model, we can define its main properties, the hyperparameters. These are, for example, the number of trees in a forest (number of estimators), the maximum depth (number of tree levels to split), the *min sample split* specifying the minimum number of observations at a point to be split, etc. (Pedregosa et al., 2011).

Since we want to know the ideal values of the hyperparameters, it is possible to create a loop that creates a model with new values and returns its accuracy at each iteration. The most accurate combination of hyperparameters would be selected. However, this procedure would be very computationally and time consuming. Thus, it will only be used to tentatively determine the hyperparameters for control of the automatic algorithm.

The other subactivity will be dataset balancing. If the dataset is imbalanced, *i.e.*, values in one category are far more numerous than in another, then balanc-

ing needs to be addressed. This is done because with a large imbalance, the algorithm would be trained on a high number of high visibility examples and a very small number of fog examples. It would then have a hard time identifying fog instances when it knew the minimum number of them. There are different ways to deal with imbalanced dataset, using advanced combined methods, *e.g.*, threshold changes, cost sensitive learning, or data sampling (Brownlee, 2021). From the mentioned methods, sampling is very easy to apply, and it would not affect idea of the research.

The data sampling is either done manually during data preprocessing, where some data from the most populated category is manually removed, or data from less populated categories is duplicated. The second option is to use the already built-in BalancedRandomForest model, where the sampling strategy can be adjusted. This particular option will be used because it takes a minimum of lines of code and requires much less testing of the appropriate sampling ratio (BalancedRandomForestClassifier, 2012).

2.2.1. Hyperparameter tuning

To obtain best set of hyperparameters, there were created three separate data subsets. This was done because in this stage, we could not surely say whether we would use balanced, imbalanced dataset or *e.g.*, separate model for precipitation. Therefore, procedure was applied to three scenarios, where all predictors were used:

1. Random forest without balancing dataset,
2. Balanced random forest,
3. Observations with precipitation only.

First, the roughly calibrated model was processed in a loop and examined to obtain a possible range of parameters and to detect any serious errors in the automatic hyperparameter search. The loop iterated through the list of possible values, evaluated the accuracy, and finally only the values with the highest accuracy were retained.

After obtaining outcomes from manually programmed loop, RandomizedSearchCV method was applied in estimated grid of parameters (sklearn.model_selection.RandomizedSearchCV, 2022). This method uses random selection of parameters in a specified grid (sklearn.model_selection.GridSearchCV, 2022), tests models with these parameters, performs cross-validation, determines accuracy of the model and returns the most accurate combination. As a scoring metrics, negative mean squared error was chosen.

This procedure simulated three scenarios (the above mentioned imbalanced, balanced, collision). Three sets of hyperparameters were thus obtained (Tab. 8).

RF models were built with these hyperparameters. To make the models more easily comparable in terms of other influences, these hyperparameters were not further modified.

Table 8. Hyperparamteres values returned by the randomized search with cross validation.

Dataset	Max_depth	Max_features	Max_samples	N_estimators
Balanced	9	0.6	0.8	125
Imbalanced	9	0.7	0.8	110
Precipitation	5	0.1	0.3	50

2.2.2. Feature selection

Correct selection of predictors (features) significantly affects model performance. It was done in three phases during this research. Firstly, predictors with insufficient counts or different frequency (*e.g.* maximum temperature, reported once daily) were dropped from the dataset. Also, irregularly appearing predictors or predictors with many missing values were discarded (*e.g.* highest cloud layer). As a second step, predictors obviously redundant for visibility forecasting (*e.g.* report type information) were also dropped. The third step was the actual search for the ideal combination of predictors. Again, this can be done using manually written code or imported functions from the Sci-kit learn library.

To get best predictors combination automatically, recursive feature elimination with cross-validation (RFECV) method was chosen. This estimator is initially trained on a complete set of features (Tab. 9) and the importance of each

Table 9. Original and complete set of predictors for visibility forecast in three hours.

Predictor	Time	Series label	Symbol	Remark
Month	of Forecast	mm	M	Observed
Cyclical value of month	of Forecast	month_sin month_cos	M_{sin}, M_{cos}	Calculated
Cyclical hour of the day	of Forecast	HH	H	Calculated
Visibility	3 and 4 hours ago	visshift3, visshift4	VIS_{n-3}, VIS_{n-4}	Observed
RH	3 and 4 hours ago	rhshift3, rhshift4	RH_{n-3}, RH_{n-4}	Observed
Temperature	3 and 4 hours ago	Tshift3, Tshift4	T_{n-3}, T_{n-4}	Observed
Dew point temperature	3 and 4 hours ago	Tdshift3, Tdshift4	Td_{n-3}, Td_{n-4}	Observed
Wind speed	3 and 4 hours ago	ffshift3, ffshift4	Ws_{n-3}, Ws_{n-4}	Observed
Cloud base height	3 and 4 hours ago	h1shift3, h1shift4	Ch_{n-3}, Ch_{n-4}	Observed
Cloud amount	3 and 4 hours ago	n1shift3, n1shift4	N_{n-3}, N_{n-4}	Observed
Hourly precipitation sum	3 and 4 hours ago	r1shift3, r1shift4	$R1_{n-3}$, $R1_{n-4}$	Observed
6-hours precipitation sum	3 and 4 hours ago	r6shift3, r6shift4	$R6_{n-3}$, $R6_{n-4}$	Observed
Visibility category	3 and 4 hours ago	vis_cshift3, vis_cshift4	$VISC_{n-3}$, $VISC_{n-4}$	Observed
Category of phenomena	3 and 4 hours ago	ww_catshift3, ww_catshift4	WWC_{n-3}, WWC_{n-4}	Calculated
U-component of wind	3 and 4 hours ago	ushift3, ushift4	U_{n-3}, U_{n-4}	Calculated
V-component of wind	3 and 4 hours ago	vshift3, vshift4	V_{n-3}, V_{n-4}	Calculated
Duration of phenomena	3 hours ago	WW_l	WW_L	Calculated

feature is calculated. Then the least important feature is dropped. This procedure is repeated recursively on the set of predictors until the defined number of features is finally reached. (sklearn.feature_selection.RFE, 2022).

As the most suitable parameter for cross validation, f1 parameter was chosen (see Appendix 1). The main advantage is that f1 reflects the ideal ratio of accuracy and bias very well.

2.2.3. Output accuracy evaluation

The four most representative evaluation metrics were used for evaluating the model. As usual in various other meteorological works, the Jolliffe and Stephenson contingency table method will be applied (Jolliffe and Stephenson, 2012). In a very comprehensive manner, the output statistics will show accuracy, false alarm, miss rate, and fog detection probability. Their characteristics are listed in Tab 10. It should be noted that a variety of other accuracy metrics exist, but in the interest of maintaining clarity of the procedure, these have been selected as representative and relevant. It should be recalled that this evaluation of forecasts is somehow different from the usual binary classification, *i.e.*, dichotomous forecasts, as it is rather multi-categorical. Abbreviations have been defined for this purpose: n for number of all tested cases, F_i for category i forecasted, O_i for category i observed.

Table 10. Accuracy metrics for model evaluation.

Metrics	Equation	Explanation
Accuracy	$Acc = \frac{1}{n}\sum F_i O_i$	What fraction of forecasted values were correct?
Total False Alarm Rate	$FAR = \frac{1}{n}\sum F_i O_j$ For $F_i < O_j$	What fraction of all the forecasts were alarming – predicting lower visibility?
Miss Ratio	$Miss = \frac{1}{n}\sum F_i O_j$ For $F_i > O_j$	What fraction of all the forecasts were erroneous – predicting higher visibility?
Fog Probability Detection	$PoFgD = \frac{1}{n_i}\sum F_i O_i$ For $i = 1$	What fraction of fog observations were forecasted correctly?

Other very relevant metrics are mentioned *e.g.*, in (Jolliffe and Stephenson, 2012), particularly Gerrity skill score, Heidke skill score or Peirce's skill score that can be very conveniently used for multicategory forecasts. However, the potential ability to accurately assess leads to the use of a high number of metrics, which reduces the clarity of the assessment. Therefore, for the purposes of this article, rather basic metrics are used. However, if interpreted correctly, their predictive capability is preserved.

3. Results

Different combinations of predictor sets were tested in the experimental phase. The experiments were initiated with the entire data set listed in Tab 9.. Each set of predictors was tested on different model configurations (balanced, imbalanced, with different hyperparameters or, in the case of balanced RF, with different sampling strategies). The importance of each predictor (feature) was estimated as the permutation importance. This is calculated as the reduction in the model score when one predictor value is randomly shuffled (Breiman, 2001). Besides numerical value, importance was also visualized (Fig. 6) since it in some cases yielded quite different results.

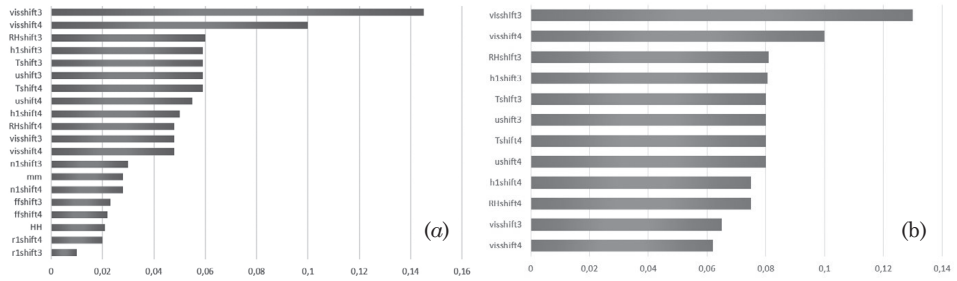


Figure 6. Permutation importance visualization for (*a*) full dataset, and (*b*) experimental balanced dataset by imblearn.ensemble.BalancedRandomForest.

After the experiments, it was possible to highlight four subsets of predictors that achieved the highest accuracy or produced other interesting value, such as very low FAR or high fog detection capability. Predictors in these subsets were selected based on permutation importance.

D1. Complete dataset (as shown in the Tab. 9.)

D2. M_{sin}, VIS_{n-3}, VIS_{n-4}, RH_{-3}, T_{n-3}, U_{n-3}, V_{n-3}, Ch_{n-3}, VIS_{n-4}, RH_{n-4}, T_{n-4}, U_{n-4}, V_{n-4}, Ch_{n-4}, WW_L;

D3. H, VIS_{n-3}, VIS_{n-4}, $VISC_{n-3}$, WWC_{n-3}, RH_{-3}, T_{n-3}, U_{n-3}, V_{n-3}, Ch_{n-3}, VIS_{n-4}, RH_{n-4}, T_{n-4}, U_{n-4}, V_{n-4}, Ch_{n-4}, WW_L;

D4. M_{sin}, VIS_{n-3}, VIS_{n-4}, RH_{-3}, T_{n-3}, U_{n-3}, V_{n-3}, Ch_{n-3}, VIS_{n-4}, RH_{n-4}, T_{n-4}, U_{n-4}, WW_L;

Some results that represent well the differences between the subsets, sampling strategies and balancing are shown in the Tab. 11.

The table shows that it is possible to achieve almost 80% accuracy on the test set. The balancing of the dataset is of great importance and the values of underestimation and overestimation of the model (Miss, FAR) are flipped accord-

Table 11. Accuracy metrics for the random forest method applied to the 3h visibility forecasting (applied on four defined subsets).

Dataset no.	Balanced	Remark	Acc	Miss	FAR	PoFgD
D1	Yes	Sampling_strategy = auto	72%	7%	21%	69%
D1	Yes	Sampling_strategy = majority	60%	8%	32%	65%
D1 (WWL excl.)	Yes	Sampling_strategy = auto	70%	8%	22%	65%
D2	Yes	Sampling_strategy = auto	72%	7%	21%	68%
D3	No	----	78%	15%	6%	61%
D4	No	----	78%	14%	8%	54%
D4	Yes	Sampling_strategy = auto	71%	9%	19%	59%

ing to the procedure used. At first sight, the worst success rate accomplished the 'majority' sampling strategy used, which reduces the number of only the most numerous categories. This significantly alters the data set. In contrast, the 'auto' sampling strategy, which resamples all classes, performed better. Obviously, the sampling strategy must also be handled cautiously. The following key statements can be drawn from the tables:

1. When balancing the dataset, the FAR increases significantly and the Miss value decreases. This is due to the nature of the balancing. It creates dataset, where significantly more low visibility situations occur during training than in reality.
2. The balanced dataset shows the highest value of fog detection potential. It can be assumed that a model specialized for fog prediction can also be created, but it will be inapplicable for the other categories.
3. The selection of an appropriate set of attributes is one of the main factors in the percentage evaluation. However, even with a manually calibrated set of features, similar values can be achieved. It depends on what accuracy metrics are used and what is the capture of the model author.

The thresholds were set somewhat subjectively to make the results universally applicable. However, when considering the aviation application and directly the ICAO Annex 3 thresholds (800, 1500, 3000, 5000 m) for TAF change groups, the accuracy changes. PoFgD has of course been calculated for the 0–800 m category, so the results in the Tab. 12 are not directly comparable to those previously presented.

Table 12. Output accuracy metrics of model for 3h visibility forecasting of ICAO Annex 3 categories.

Dataset no.	Balanced	Remark	Acc	Miss	FAR	PoFgD
4	Yes	Sampling_strategy = auto	82%	3%	15%	62%
4	No	–	90%	3%	7%	57%

The results presented in the table look promising, especially in the case of an imbalanced dataset. It should be recalled that there was a very high frequency of visibilities greater than 5 km. Therefore, for these visibility categories, the model was slightly advantaged over those predicting smaller categories.

3.1. One-hour model results

As the forecast above is meant only to forecast value after three hours, it is also valid to test how well we can forecast values before this term. Therefore, one-hour time-span was also modelled in order to see its possible accuracy improvement (Tab. 13).

Table 13. Output accuracy metrics of model for 1h visibility forecasting.

Dataset no.	Balanced	Remark	Acc	Miss	FAR	PoFgD
4	No	Sampling_strategy = auto	85.6%	7.80%	6.50%	77%
4	Yes	Sampling_strategy = auto	81.7%	5.70%	12.50%	77%

As expected, the model accuracy for the hourly forecast is substantially better. Since the hourly forecast was not the focus of this paper, no in-depth analysis was performed, neither exploratory nor output statistics. An interesting value regarding the random forest in general, in terms of the three-hour model results, is that the FAR and Miss ratio values remained much lower for both the balanced and unbalanced data sets.

3.2. Persistent fog forecasting comparison

To check if the model does not merely copy the persistent forecast, an analysis of the theoretical success of the persistent forecast was performed (see Appendix 2 – Fog observations analysis for a more detailed discussion of visibility values and their effect on fog forecast accuracy). This was based on the analysis of observations. The frequency with which fog would persist at 1, 2 and 5 hours ahead of the forecast was calculated.

The following figure shows the percentage of observed visibility categories 1 to 5 hours before the hour when the fog was observed (Fig. 7). In the first hour, 77% of the observed fog corresponds to the accuracy of the persistent forecasts. In contrast, in the three-hour time range, the random forest clearly outperformed the probability of fog detection for a persistent forecast. In the five-hour time span, the persistent forecast would only reach 40%.

3.3. Influence of phenomena on fog forecasting

The outputs were analyzed not only in terms of visibility but also in terms of association with phenomena. It is assumed that if the fog was not well pre-

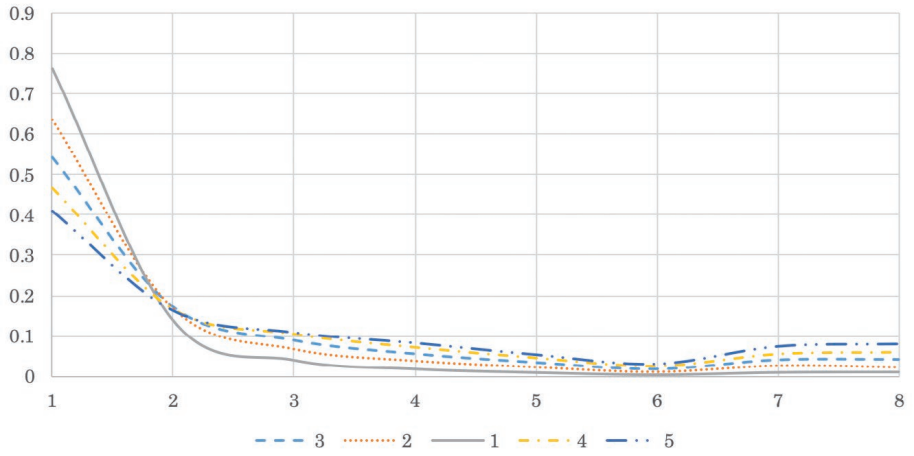


Figure 7. Relative frequency of occurrence of visibility categories (0 represents fog, 7 represents 10–100 km visibility) in the time range 1–5 hours before fog observation (represented by five lines).

dicted, it could have been caused by sudden precipitation, advective factors or sudden radiative cooling.

Of RF (D3, imbalanced) forecasting results set, those where category 0 visibility (below 1 km) was observed but not predicted were selected. These forecast errors were separated into columns according to the visibility category that was incorrectly predicted. One column was assigned to each visibility category, which is divided according to the phenomena observed 3 hours before the not predicted fog (Fig. 8). All phenomena categories that were observed during not forecasted fog were involved.

Analysis of the events preceding the incorrect fog forecasts showed that in terms of the three-hour forecast, most of the inaccurate fog forecasts are associated with category 1, which reports no phenomena, therefore this category has minor or no apparent influence. This could be attributed to sudden radiative cooling or advection that would not be picked up by the model. Category 3 is also often represented, most often in the 1–2 km forecasts. This represents recent precipitation in the previous hour, which could indicate a fluctuating weather pattern and very difficult to capture effects of water vapour and surface moisture. Precipitation impacts are captured by categories 6, 7 and 8, followed by rare occurrences of categories 10 and 13.

3.4. Current phenomena analysis

Particularly during sudden episodes of precipitation associated with showers and thunderstorms, a very sharp drop in visibility can be expected. Therefore, although this cannot be included in the forecast, current (observation time) phe-

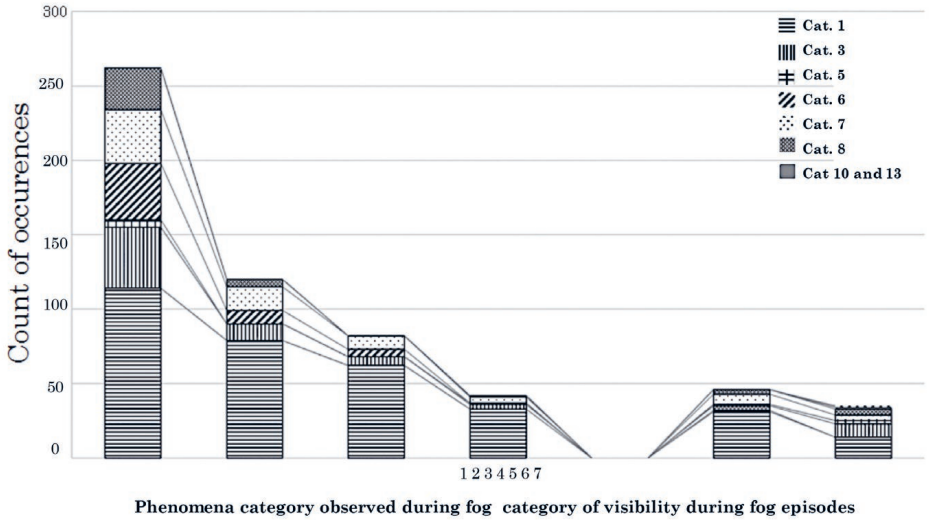


Figure 8. The visibility categories predicted at the time the fog was predicted. Each column represents one category and is divided by the categories of phenomena that were observed. (1 for visibility 1–2 km, 7 for category 10-100 km, corresponding to the Table *6*, phenomena corresponding to the Table 7).

nomena have also been investigated (Tab. 14). The aim was to reveal if the method can deal with phenomena that suddenly cause a drop of visibility within one hour.

The high counts of categories 6 and 7 suggest insufficient modelling of radiative cooling, as the model predicts very good visibility when in fact it is below 1 km. It can also be concluded that the category 1 prediction (visibility 1–2 km) models the process very promisingly with a relatively small error. Concerning rain showers, it is obvious that model did not expect any visibility drop, although it dropped under 1 km. Here is place for improvement in terms of phenomena

Table 14. Counts of phenomena categories observed during incorrectly forecasted visibility (visibility lower than 1 km forecasted, 1–7 standing for visibility categories, 1 for visibility 1–2 km, 7 for category 10–100 km, corresponding with Tab. 6, phenomena categories corresponding with Tab. 7.

WW category	Shortened description	Visibility category						
		1	2	3	4	5	6	7
1	No precipitation	9	5	2	0	0	0	3
3	Precipitation in preceding hour	4	2	0	0	0	1	0
5	Fog	144	83	54	36	0	60	53
6	Drizzle	48	17	16	2	0	2	2
7	Rain	32	7	7	2	0	3	3
8	Snow	24	5	4	3	0	0	10
11	Showers	0	0	0	0	0	0	5

classification. It should be noted that this method basically cannot predict that a storm or rain shower will come. From the set of predictors, it can only be a change in wind characteristics, a storm or shower that has already occurred that day, or cloud height characteristics. However, convection is far too complex a phenomenon to be perfectly predicted from a set of observational data alone.

3.5. Visibility in precipitation

The effect of precipitation was also tested (Tab. 15). The purpose of this testing was to determine if the RF method can help in predicting decreasing visibility during precipitation. The model has only been trained and tested based on observations of precipitation and has not been thoroughly calibrated as part of this work, so it can undoubtedly be improved. Obviously, it will be necessary to include parameters to predict precipitation so that its effect can be better predicted by the model. These would probably be the amount and height of cloud base and ideally the form of precipitation.

Table 15. Output accuracy metrics of model for 3 h visibility forecasting when precipitation observed.

Dataset no.	Balanced	Remark	Acc	Miss	FAR	PoFgD
D1	Yes	Sampling_strategy = auto	76%	5%	19%	89%

The rainfall observation model showed relatively good accuracy. The probability of detecting very low visibility increased to 89%. In any case, it did not avoid a high FAR due to the balanced data set. The better performance in the lowest visibility region (below 1 km) can be attributed to the higher proportion of observations where fog is followed by rain or drizzle. The main problem with using this model is the assumption that we actually know whether it will rain in the following hours. Rather, this is an example to guide further research.

4. Discusion

Visibility modelling using the random forest method for a three-hour forecast interval based on observational data only has achieved quite promising results. For the unbalanced data set, it achieved an accuracy of 78% with 14–15% error rate and approximately 6–8% false alarms. The balanced random forest achieved an accuracy of 71–72% and reduced the error rate to only 7–9% but 19–20% false alarms in comparison.

Such a success rate shows that even with only knowledge of the observation statistics, one can make a reasonably good estimate of the likely future trend in visibility. Another very important finding is that the ratio of false alarms and misses is reversed when balancing the dataset. This can be of considerable use

for further applications of the random forest method in visibility prediction. These findings also show where the basic accuracy of RF classifiers for visibility prediction can be expected and where it can be further improved by adding additional data.

The research results have several limitations that were known in advance. However, they are primarily based on the data set itself. Firstly, the modelling was done on data from a single station, so it is not possible to use a set of predictors or perhaps model hyperparameters. What can be used is the procedure used.

As far as point forecasts are concerned, they usually take the form of probabilities or a more specific determination, *e.g.* they can take the form "visibility over 10 km, initially 3–5 km", *i.e.* they allow more values. Most forecasts also predict for longer intervals, not exactly one hour. This approach was therefore strictly deterministic.

5. Conclusions

The main research question that was asked beforehand was: How will the random forest method perform on observational data alone?

As anticipated, the research provided a way for interested individuals to handle and analyze visibility. The work established the basic accuracy of an observation-only machine learning model for different predictors, hyperparameter configurations, or target categories thresholds. The work also identified the limitations of the Random Forest method on the observational data set. Analyzing the results, the work identified some situations where the RF model is wrong and where human error in prediction can occur.

The work formed the basis for the development of a combined or hierarchical system to support forecasting in practice. Based on the procedures used, the method can be applied to other groups of predictors and other locations.

The results of the paper are still limited by the limitations of the data used. The work does not present the possibility of implementing other different datasets such as radar data, forecasts, etc. Also, the possibility of using several ML models simultaneously is not applied.

The future use of this method can be considered for the construction of a comprehensive tool to support aviation forecasting (*e.g.* take-off forecast, TREND forecasts). These are sometimes issued every 30 minutes and can be very difficult to concentrate on fully. Obviously, other data would need to be factored into the calculation, such as radar to capture the effects of convection and incoming precipitation, processed radar imagery, etc. It would certainly be useful to add data from nearby stations, weather type data, information from upper-air sounding, etc. Adding numerical model outputs would be a separate topic.

References

BalancedRandomForestClassifier (2012): Imbalanced learn, https://imbalanced-learn.org/stable/references/generated/imblearn.ensemble.BalancedRandomForestClassifier.html.

Breiman, L. (2001): Random forests, *Mach. Learn.*, **45**, 5–32, https://doi.org/10.1023/A:1010933404324.

Brownlee, J. (2017): *Long short-term memory networks with Python.* Jason Brownlee, p. 246.

Brownlee, J. A. (2019): A gentle introduction to imbalanced classification, *Machine learning mastery,* https://machinelearningmastery.com/what-is-imbalanced-classification/.

Chase, R., Harrison, D., Burke, A., Lackmann, G. M. and McGovern, A. (2022): A machine learning tutorial for operational meteorology. Part I: Traditional machine learning, *Wea. Forecast.*, **37**, 1509–1529, https://doi.org/10.1175/WAF-D-22-0070.1.

Cordeiro, F., Gutemberg, B. F., Leite de Albuquerque Neto, F. and Gultepe, I. (2021): Visibility and ceiling nowcasting using artificial intelligence techniques for aviation applications, *Atmosphere*, **12**(12), 1657, https://doi.org/10.3390/atmos12121657.

Deng, T., Cheng, A., Han, W. and Lin, H. (2019): Visibility forecast for airport operations by LSTM neural network, *ICAART,* **2**, 466–473, https://doi.org/10.5220/0007308204660473.

Gultepe, I., Müller, M. D. and Boybeyi, Z. (2006): A new visibility parameterization for warm-fog applications in numerical weather prediction models, *J. Appl. Meteorol. Climatol.*, **45**, 1469–1480, https://doi.org/10.1175/JAM2423.1.

Hill, A. J., and Schumacher, R. S. (2021): Forecasting excessive rainfall with random forests and a deterministic convection-allowing model, W*ea. Forecast.*, **36**(5), 1693–1711, https://doi.org/10.1175/WAF-D-21-0026.1.

International Civil Aviation Organization (ICAO) (2010): *Annex 3 to the Convention on International Civil Aviation. Meteorological Service for International Air Navigation*. Montreal, Quebec, Canada.

Jolliffe, I. and Stephenson, D. (Editors) (2012): *Forecast verificaton: A practitioner's guide in atmospheric science.* Oxford: Wiley-Blackwell, p. 296.

Patriarca, R., Simone, F. and Di Gravio, G. (2023): Supporting weather forecasting performance management at aerodromes through anomaly detection and hierarchical clustering, *Expert Syst. Appl.*, **213**(C), 119210, https://doi.org/10.1016/j.eswa.2022.119210.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weis, R., Dubourg, V., Vanderplas, J., Passos, A. and Cournapeau, D. (2011): Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.,* **12**, 2825–2830. Retrieved from Scikit learn: https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf.

Precision-Recall (2007): Retrieved from Scikit learn: https://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html?highlight=precision.

Salman, A. G. and Kanigoro, B. (2021): Visibility forecasting using autoregressive integrated moving average (ARIMA) models, *Procedia Comput. Sci.*, **179**, 252–259, https://doi.org/10.1016/j.procs.2021.01.004.

Schultz, M. G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L. H., Mozaffari, A. and Stadtler, S. (2021): Can deep learning beat numerical weather prediction? *Phil. Trans. R. Soc.*, **A379**, 20200097, https://doi.org/10.1098/rsta.2020.0097.

Schumacher, R. S., Hill, A. J., Klein, M., Nelson, J. A., Erickson, M. J., Trojniak, S. M. and Herman, G. R. (2021): From random forests to flood forecasts: A research to operations success story, *B. Am. Meteorol. Soc.*, **102**(9), pE1742–E1755, https://doi.org/10.1175/BAMS-D-20-0186.1.

sklearn.ensemble.RandomForestClassifier (2012): Retrieved from Scikit Learn: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html?highlight=random+forest+classifier.

sklearn.linear_model.LogisticRegression (2012): Retrieved from Scikit Learn: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html?highlight=logistic+regression.

sklearn.metrics.f1_score (2011): Retrieved from Scikit learn: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html.

sklearn.model_selection.GridSearchCV (2022): Retrieved from Scikit Learn: https://scikit-learn.org/stable/modules/grid_search.html.

sklearn.model_selection.RandomizedSearchCV (2022): Retrieved from Scikit Learn: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RandomizedSearchCV.html.

sklearn.naive_bayes.GaussianNB (2012): Retrieved from Scikit Learn: https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html.

sklearn.feature_selection.RFE (2022): Retrieved from Scikit Learn: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE.html.

sklearn.svm.SVC (2012): Retrieved from Scikit Learn: https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html#sklearn-svm-svc.

Sládek, D. (2021): Automation of the quality assesment of aeronautical meteorological reports, Ph.D. Thesis. University of Defence, Brno, Czech Republic (in Czech, Automatizace hodnoceni kvality leteckych meteorologickych zprav), p. 112.

Sruthi, E. R. (2021): Understanding Random Forest, Retrieved from Analytics Vidhya: https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/.

Tsai, Y.-Z., Hsu, K.-S., Wu, H.-Y., Lin, S.-I., Yu, H.-L., Huang, K.-T., Hu, M.-C. and Hsu, S.-Y. (2020): Application of random forest and ICON models combined with weather forecasts to predict soil temperature and water content in a greenhouse, *Water,* **12**, 1176, https://doi.org/10.3390/w12041176.

Wantuch, F. (2001): Visibility and fog forecasting based on decision tree method, Időjárás, **105**, 29–38.

WMO (2019): *Manual on Codes - International Codes*, Volume I.3 – Annex II to the WMO Technical Regulations: Part D – Representations derived from data models. Geneva, Switzerland, p. 239, https://library.wmo.int/doc_num.php?explnum_id=10530.

Zhu, L., Zhu, G., Han, L. and Wang, N. (2017): The application of deep learning in airport visibility forecast, *Atmos. Climate Sci.*, **7**, 314–322, https://doi.org/10.4236/acs.2017.73023.

SAŽETAK

Kratkoročna prognoza vidljivosti određena metodom slučajne šume

David Sládek

Točno predviđanje vidljivosti ključno je za sigurne operacije zrakoplova. Ova studija ispituje kako različite konfiguracije modela slučajne šume (eng. *Random Forest*) mogu poboljšati predviđanja vidljivosti. Koriste se tehnike predprocesiranja, uključujući analizu korelacije za prepoznavanje temeljnih odnosa u promatranjima vremena. Podaci *vremenskih nizova* pretvaraju se u redoviti podatkovni okvir kako bi se olakšala analiza. Ova studija predlaže klasifikacijski okvir za organiziranje podataka o vidljivosti i meteoroloških pojava. Taj okvir se zatim koristi za razvoj prognoze vidljivosti korištenjem metode slučajne šume. Studija također prikazuje postupke za podešavanje hiperparametara, odabir značajki, uravnotežavanje podataka i procjenu točnosti za taj skup podataka. Glavni rezultati su parametri modela slučajne šume za trosatnu prognozu vidljivosti te analiza pogrešaka prognoze slabe vidljivosti. Dodatno, ispitani su i modeli za jednosatnu prognozu i prognozu vidljivosti u slučaju oborine. Dobiveni modeli pokazuju točnost determinističke prognoze od približno 78%, uz oko 6% lažnih uzbuna, dajući sveobuhvatan pregled mogućnosti modela slučajne šume za predviđanje vidljivosti. Kao što se i očekivalo,

model je pokazao ograničenja pri simulaciji brzog radijacijskog hlađenja i pri naglom smanjenju vidljivosti uzrokovanom oborinama. Naime, u odnosu na oborine, točnost modela je bila 79%, ali stopa lažnih uzbuna iznosila 19%. Dodatno, metoda slučajne šume postavlja temelje za poboljšanje točnosti prognoza uključivanjem dodatnih prognostičkih podataka, dok njezina primjena na skupove realnih podataka proširuje primjenu tehnika strojnog učenja na na meteorološke probleme.

Ključne riječi: zrakoplovna meteorologija, predviđanje vidljivosti, prognoza sadašnjeg vremena, prognoza pri slijetanju (trend), strojno učenje, slučajna šuma, odabir značajki, podešavanje hiperparametara

Corresponding author's address: David Sládek, Department of Military Geography and Meteorology, University of Defence, Kounicova 65, Brno, the Czech Republic; e-mail: david.sladek@unob.cz

Appendix 1 – F1 score

"The F1 score can be interpreted as a harmonic mean of the precision and recall, where an F1 score reaches its best value at 1 and worst score at 0. The relative contribution of precision and recall to the F1 score are equal" (sklearn.metrics.f1_score, 2011). *F*1 score can be calculated as:

$$F1 = \frac{2(p * r)}{(p + r)},$$

where p is precision and r recall. Precision is defined by scikit.learn as the number of true positives (Tp) over the number of true positives plus the number of false positives (Fp);

$$P = \frac{T_p}{T_p + F_p}.$$

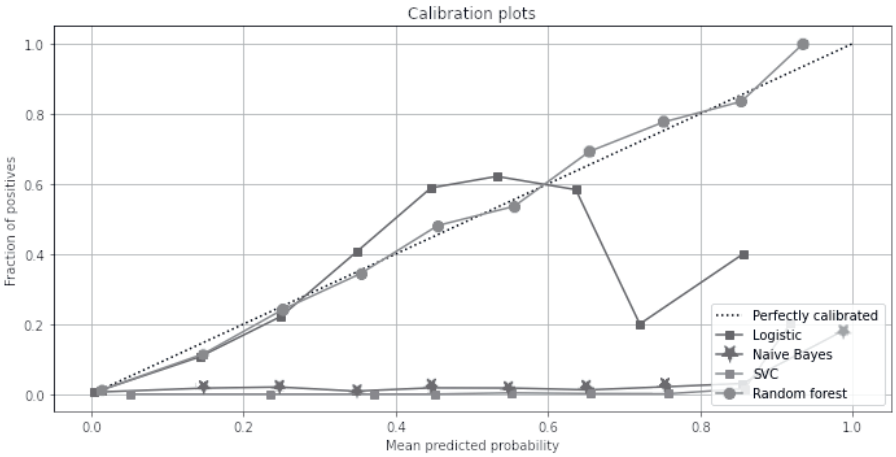
Recall is defined as the number of true positives (Tp) over the number of true positives plus the number of false negatives (Fn) (Precision-Recall, 2007):

$$R = \frac{T_p}{T_p + F_n}.$$

Appendix 2 – Methods comparison

The following three-year comparison (2012–2015) was used to support an assertion of the appropriateness of using the RF method. It shows presumable model performances of:

1. Random Forest Classifier
2. Logistic Regression
3. Gaussian Naive Bayes classifier
4. Linear SVC (support vector classification)[1]



Appendix 3 – Fog observations analysis

Following table composition presents overview of all hours of fog forecasted. Boxplots show values of visibility observations (from left to right) 5 hours - 1 hour preceding the fog observation.

---

[1] Sources: (sklearn.svm.SVC, 2012; sklearn.naive_bayes.GaussianNB, 2012; sklearn.ensemble.RandomForestClassifier, 2012; sklearn.linear_model.LogisticRegression, 2012)

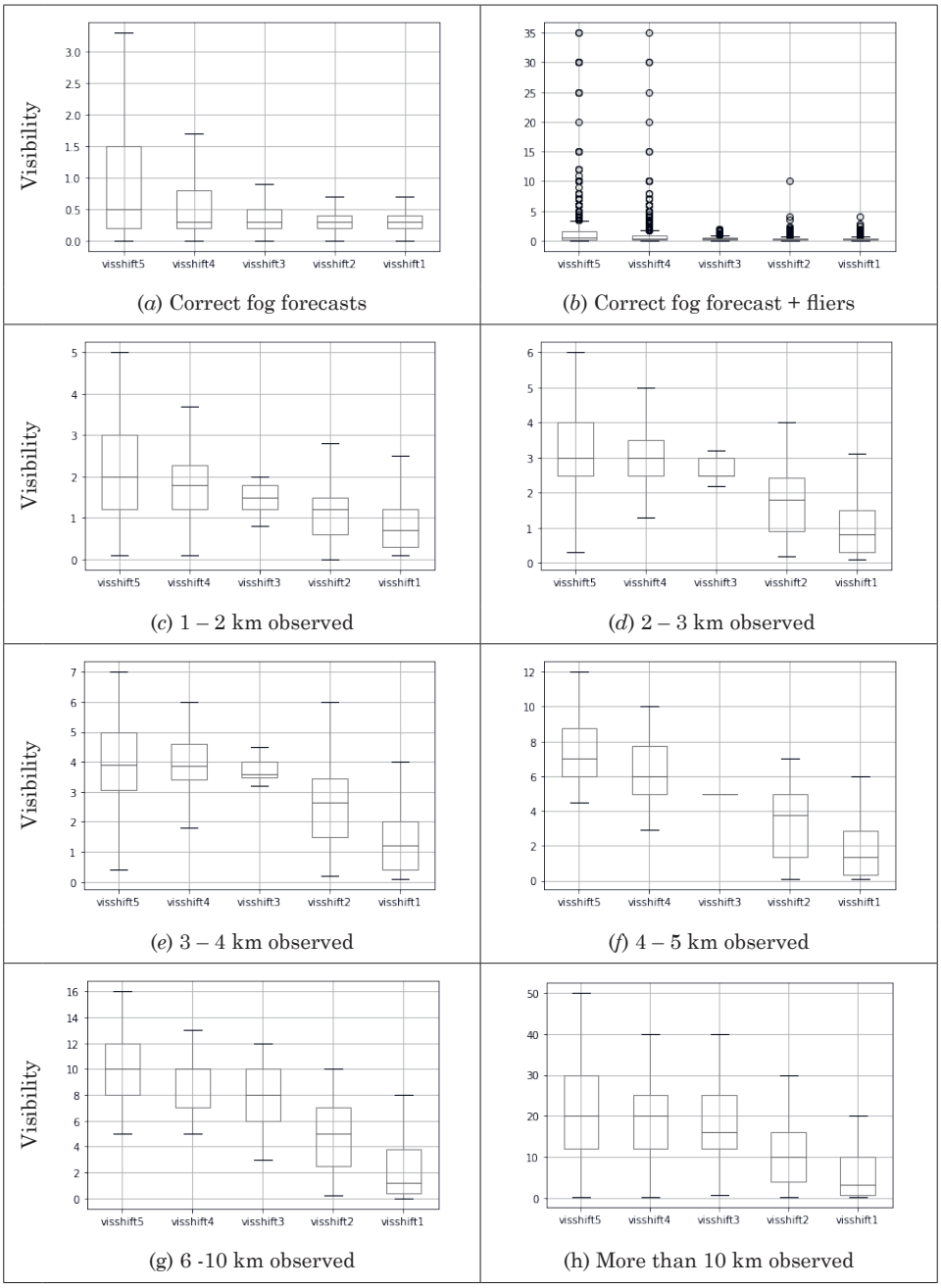


Figure 9. Boxplots of observed visibility before fog was observed. Categorized by correct (a), (b) and incorrect forecasts (in categories as captioned).

From the table composite in the Fig. 9, it is obvious that fog is correctly forecasted in case it is observed three hours in advance of observation (a). However, there are also situations (b), when fog is forecasted correctly although 5 and 4 hours before, visibility is about 20 or 30 km. Extremely problematic, on the other hand, are situation when visibility is more than 10 km – even 20-40 km, as shown in (h) – and fog is observed.