

Predictability of fog using surface observation-based machine learning models: evaluation at Argentinean airport

Melina Sol Yabra

`melina.s.yabra@nasa.gov`

National Aeronautics and Space Administration

Juan Jose Ruiz

National Scientific and Technical Research Council

Ramon de Elia

National Meteorological Service

Luciano Vidal

National Meteorological Service

Research Article

Keywords: fog, forecast, surface observations, machine learning, aviation, Argentina

Posted Date: February 10th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8503721/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at CEAS Aeronautical Journal on August 28th, 2026. See the published version at <https://doi.org/10.1007/s13272-026-01003-8>.

Predictability of fog using surface observation-based machine learning models: evaluation at Argentinean airport

Melina Sol Yabra^{1,2,a}, Juan Ruiz^{3,4,b}, Ramon de Elia², Luciano Vidal^{2,c}

¹Climate and Radiation Laboratory, Earth Science Division, NASA Goddard Space Flight Center, Greenbelt, MD 20770, USA

²Servicio Meteorológico Nacional, C1425GBE, Buenos Aires, Argentina.

³Centro de Investigaciones del Mar y la Atmósfera (CIMA), CONICET – UBA, C1428EGA, Buenos Aires, Argentina.

⁴Instituto Franco Argentino Sobre Estudios de Clima y Sus Impactos (UMI-IFAECI/CNRS-CONICET-UBA), C1428EGA, Buenos Aires, Argentina.

Corresponding author: Melina Sol Yabra melina.s.yabra@nasa.gov

^a: 0000-0003-1614-778X

^b: 0000-0002-5079-641X

^c: 0000-0001-6715-2648

Acknowledgements

This work was carried out within the framework of a Doctoral Thesis in Atmospheric and Ocean Sciences of the Faculty of Exact and Natural Sciences of the Universidad de Buenos Aires, Argentina, based at the Servicio Meteorológico Nacional, Argentina (SMN), and funded by the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) of Argentina. The authors also want to thank the Centro de Informaciones Meteorológicas of the SMN for the data provided.

Statements and Declarations

The authors declare no conflicts of interest.

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Abstract

Low visibility conditions due to the presence of fog can lead to delays or cancellations of arriving and departing flights to avoid possible incidents or accidents at airports. Therefore, accurate visibility forecasts are required to keep airport capacity as high as possible. Recently, some studies explored the use of machine learning algorithms and routine surface meteorological observations to produce fog forecasts. This study explores the potential of machine learning (ML) models trained solely on surface observations for short-term fog nowcasting at Ezeiza International Airport in Argentina, within 6 hours from initialization time. Three simple ML methods—a logistic regression, a decision tree, and a random forest—were evaluated using 20 years of data. Results show that nonlinear methods (decision tree and random forest) required fewer, simpler predictors to match or exceed the forecasting skill of the logistic regression, which relied more heavily on complex variable transformations. All methods significantly outperformed climatological and persistence-conditional reference models. In all three methods, the most important predictor was the fog condition at initialization time. A seasonal and daily data-splitting strategy improved performance at almost all forecasting time horizons. The models trained in this study represent a promising operational tool for aviation forecasting and could also complement operational numerical models, whose ability to resolve fog-scale processes remains limited.

Keywords: fog, forecast, surface observations, machine learning, aviation, Argentina

1. Introduction

The reduction of visibility associated with fog and mist, and their consequent reduction of visibility, can greatly impact several weather-dependent human activities such as agriculture (Baldocchi and Walker, 2014) and transportation (Cook et al., 2009; Rudra et al., 2015; Gultepe et al., 2014; 2017) and can also affect human health, contributing to enhance pollution (Tanaka et al., 1998; Kampa and Castanas, 2008). Particularly in air transportation, fog is one of the main safety and logistical disruptors as it can cause incidents during aircraft landing and departure, generating material and human losses in the most adverse cases (Gultepe et al., 2019; Gultepe, 2023). To ensure safe operations under low-visibility conditions, airports implement Low Visibility Procedures (LVP), which reduce airport capacity and often lead to flight delays, cancellations, or diversions, with substantial economic consequences (Anaman et al., 2000; Tabernier et al., 2021). The negative impacts of fog are primarily mitigated in two ways:

61 through improved instrumentation and operational procedures for reduced visibility, and
62 through accurate fog forecasting.

63 An accurate fog forecast enables airport traffic controllers and airline operators to
64 reorganize flight schedules and load additional fuel on aircraft, anticipating potential delays
65 caused by reduced visibility. To achieve this, aviation meteorological forecasters must predict
66 airport weather conditions at least 24 hours in advance, which is the maximum validity period
67 for the Terminal Aerodrome Forecast (TAF), which is updated every 6 hours (International
68 Civil Aviation Organization, ICAO, 2010). Therefore, the initial 1- to 6-hour forecast must
69 provide the highest short-term reliability. However, the required forecast lead time and
70 accuracy may vary depending on factors such as flight duration and preparation time, meaning
71 that some mid- and long-haul flights may still be affected if high-accuracy forecasts are not
72 available well in advance.

73 The literature presents various approaches and techniques for predicting fog and
74 associated visibility values. One of the most commonly used tools for this purpose is numerical
75 weather prediction (NWP) models (e.g., Bergot and Guedalia, 1994; Gultepe et al., 2007).
76 However, NWP models still face several challenges when forecasting fog, including
77 insufficient vertical and horizontal resolution, poor representation of the planetary boundary
78 layer and microphysical processes, systematic errors, and a lack of accurate observations to
79 generate initial conditions (Bari et al., 2023). As a faster and more computationally efficient
80 alternative, data-driven statistical methods have been widely used and improved over the past
81 decades (Weyn et al, 2020; Wu and Xue, 2024). There are also hybrid approaches where NWP
82 models provide input data for statistical models (Kamangir et al., 2021, Ruiz et al., 2018).

83 In recent decades, various data-driven statistical methods have been tested to generate
84 visibility and fog occurrence forecasts. For example, Vislocky and Fritsch (1997) used multiple
85 linear regression to generate visibility forecasts from observations. Their forecasting system
86 was later improved by Leyton and Fritsch (2003, 2004) by increasing the frequency of surface
87 observations. Ghirardelli and Glahn (2010) developed an operational visibility forecasting
88 system for multiple lead times and locations, again using multiple linear regression. Given the
89 complexity and nonlinear nature of the problem, characterized by the temporal and spatial
90 variability of fog, considerable efforts have been dedicated to applying machine learning (ML)
91 approaches, as they have proven to be effective solutions in other areas of weather forecasting,
92 such as deep convection (McGovern et al., 2023), volcanic ash dispersion (Torrìsi et al., 2022),

among others. In this field, ML-based methods used to forecast visibility include neural networks (Pasini et al., 2001; Marzban et al., 2007; Miao et al., 2012; 2020), Bayesian model averaging (Roquelaure et al., 2009; Boneh et al., 2015), support vector regression (Cornejo-Bueno et al., 2018; 2021; Zhu et al., 2018), and decision trees (Bartoková et al., 2015; Dutta and Chaudhuri, 2015). Herman and Schumacher (2016) compared various statistical methods and datasets for airport visibility predictions and found that no single model consistently outperforms the others.

One of the main issues with ML-driven models is the selection of predictors. Most studies found the highest predictor power in visibility, humidity variables, wind speed, and cloud ceiling (Castillo-Boton et al, 2022; Pelaez-Rodriguez et al, 2023; Bajaj et al, 2025). Other studies (Bremnes and Michaelides, 2007; Schütz et al., 2023) augmented the number of predictors, including temporal variations and other transformations, to incorporate temporal aspects into the problem. On the other hand, Bari et al. (2023), in their review of decision support systems for fog, emphasized the importance of providing clear and concise information about fog forecast-related uncertainties, as this is crucial for decision-makers (Keith and Leyton, 2007; Herman and Schumacher, 2016).

In this context, the objective of the present study is to assess the predictability of fog and mist occurrence (both referred to as “fog” hereafter) in the 1–6-hour horizon at Ministro Pistarini International Airport in Ezeiza (Argentina), using surface weather observations and ML techniques with varying levels of complexity. The study evaluates the predictive potential of available data by applying logistic regression, decision trees, and random forest methods (see Yabra et al., 2023 and 2025a for information regarding data availability). This work contributes to improve the understanding of fog predictability based on surface observations, provides a comparison of statistical and ML techniques, and offers results from a region that has been scarcely explored in terms of short-term fog forecasting.

The article is structured as follows: Section 2 describes the dataset used and outlines the predictors, including new variables created, while Section 3 define ML methods and performance metrics. Section 4 presents the model performances from different perspectives and Section 6 summarizes the study’s main findings, limitations, and future research directions.

4. Data

The main dataset for this study consists of hourly synoptic observations (SYNOP) for the period 2000-2019 from the database of the National Meteorological Service of Argentina (SMN) for the Ministro Pistarini International Airport in Ezeiza, Province of Buenos Aires. This airport was selected because it has the highest frequency of fog and the highest air traffic in Argentina (Yabra et al. 2021, Yabra et al., 2023; 2025a). Based on the previously mentioned studies, this airport is mostly affected by radiative fog (i.e., Lala et al., 1975) and less frequently by advection fog (i.e., Roach, 1995).

The original dataset includes the following variables:

1. Month of the observation (integer between 1 and 12).
2. Hour of the observation in UTC (integer between 0 and 23).
3. Air temperature at 2 m in °C.
4. Dew point temperature at 2 m in °C.
5. Surface pressure in hPa.
6. Total cloud cover (integer between 0 and 8).
7. Wind speed at 10 m in kts.
8. Wind direction at 10 m in degrees.
9. Visually estimated minimum visibility (VIS, between 0 and 70 km reported in a non-linear scale. See World Meteorological Organization-WMO-, 1954).
10. Cloud base height above the station is encoded in 9 categories (from 0 m to 2500 m reported in a non-linear scale. WMO, 1954).
11. Fog-related SYNOP present weather condition (PW-Fog) defined as 1 if the present weather is 10, 11, 28, or 40-47, and 0 otherwise.
12. Precipitation-related SYNOP present weather condition (PW-Precipitation) defined as 1 if the present weather is 14-16, 20-29, 36-39, or 50-99, and 0 otherwise.
13. Lithometeor-related SYNOP present weather condition (PW-Lithometeor) defined as 1 if the present weather is 4-9, 30-35, or 98, and 0 otherwise.
14. No significant event-related present weather condition (PW-Nothing) defined as 1 if the present weather is 0-3, and 0 otherwise.
15. Dew point depression in °C.
16. Relative humidity in %.
17. Sin of the hour (Sin-hour): Calculated as:

$$\sin(\text{hour}) - \sin(\text{hour} - 1) = \sin(\text{hour}) \left(2\sin\left(\frac{\pi}{24}\right) + \frac{\pi}{4} \right)$$

18. Sin of the month (Sin-month): Calculated as:

$$\sin(\text{month} - \text{month}_{\text{start}}) = \sin(2\pi \frac{\text{month} - \text{month}_{\text{start}}}{12} + \frac{\pi}{4})$$

19. 6h-visibility trend: Calculated as the difference between the current visibility and its value 6 hours earlier.

20. 6h-temperature trend: As in the previous predictor, but for the 2-m air temperature.

21. 6h dew point depression trend: As in the previous predictor, but for the dew point depression.

22. 6h cloud base height trend: As in the previous predictor, but for the cloud base height.

23. 24h minimum visibility: Minimum visibility value recorded in the past 24 hours.

24. Median of the wind speed in the previous 24 hours.

25. Difference between the 2-m temperature and the crossing temperature (Baker et al. 2002).

26. 2-m temperature 24 hours before the forecasted time.

27. Visibility 24 hours before the forecasted time.

The target variable of this study is the fog condition at the forecast time, which is 1 if fog conditions are present and 0 otherwise. Fog hours are identified as in Yabra et al. (2025a) as those SYNOP reports where the present weather conditions code is any of 10, 11, 28, 40, 41, 42, 43, 44, 45, 46, or 47, visibility is under 5000 m, and dew point depression is under 2°C. The frequency of occurrence of fog hours, based on yearly information, is around 4%.

Present weather was disaggregated into 4 dichotomous variables based on the main phenomenon, since it is the only nominal categorical variable (variables 11-14). The Hour and Month variables exhibit a strong but nonlinear relationship with fog occurrence (see the hourly and monthly probability of fog in the first panels of Fig. 8 and 9), and nonlinear methodologies were in fact chosen in this study to capture such nonlinear relationships between predictors and the target variable. However, the Sin-hour and Sin-month transformations are included to evaluate whether explicitly transforming these variables further enhances predictor selection and model performance. A phase shift of $\pi/4$ was applied in both cases to better align the results with the climatological occurrence of fog, thereby facilitating the comparison. As mentioned in Section 1, the development of forecasting techniques must also include the definition of variables that could be useful for this particular problem. Based on the findings in previous studies (Yabra et al., 2025a; 2025b), fog formation at this airport is mainly associated with the

effects of the daily radiative cycle, advection of temperature or humidity, or the lowering of the cloud base. Several authors have already shown the benefits of including trends of certain variables to account for temporal variability (Bremnes and Michaelides, 2007; Schütz et al., 2023). For this reason, we include, as possible predictors, the temporal variations in the last 6 hours for visibility, temperature, dew point depression, and cloud base height (variables 19-22). Also, based on the discussions in Yabra et al (2025a, 2025b) regarding the persistence of meteorological conditions on consecutive days, variables incorporating information from the previous 24 hours are proposed (variables 23-27).

The dataset was divided into training, validation, and testing sets. The initial 70% was used for training (from 1/1/2000 to 31/12/2014), while the remaining 30% was employed to evaluate the models (from 1/1/2015 to 31/12/2019). In subsections 3.2 and 3.3, during the process of selecting hyperparameters and predictors, a new partition is applied to the training set using the same proportion to separate it from the validation sets.

3. Methodology

3.1. ML methods

As mentioned in Section 1, this study evaluates the potential of surface meteorological observations to predict the probability of fog occurrence by applying logistic regression, decision trees, and random forest methods. To this end, the probability of occurrence is estimated as the result of a binary classification problem that distinguishes between fog and no-fog hours. Figure 1 illustrates an artificial example of data classification based on two predictors (Fig. 1a) and the probability function generated by each method (Fig. 1b-d).

3.1.1. Logistic Regression

Logistic regression (LR) is an analytical method that allows the prediction of event occurrence probabilities with low computational cost. Numerous authors have already used this methodology for fog prediction (Marzban et al., 2005; Kneringer et al., 2017; Negishi and Kusaka, 2021). Equation 1 describes a multivariate logistic regression model, where p is the probability of occurrence of fog, X_i is the i -th predictor, β_i is the coefficient associated with the i -th predictor, and β_0 is the intercept term.

$$p(fog|X) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}} \quad (\text{Eq. 1})$$

The coefficients β are defined as those that minimize the distance between the estimated probability and the observed outcome for a given training dataset. This minimization is performed iteratively using a gradient descent approach. To avoid the overfitting problem in this technique, a regularization term that penalizes the magnitude of the coefficients can be added to the cost function.

Although the probability obtained from this method is a nonlinear function of the predictors, regions of equal probability are described by linear functions. Figure 1b shows an example of this for the case of two predictors, where equal probability contours are straight lines.

3.1.2. Decision Tree

Decision trees (DT) are widely used ML methods for classification problems. This methodology was used to forecast the occurrence of fog by Dietz et al. (2022) and Bartokova et al. (2015), among others, yielding promising results. By applying multiple linear partitions to the predictor space, a nonlinear probability function is obtained, as shown in Figure 1c. Each dividing line in the space corresponds to a node in the tree where a new separation is made based on a single predictor. To determine the optimal combination of variables and partitions, various metrics can be employed. In this study, the chosen metric is the entropy, which quantifies the impurity of a node by measuring the disorder or uncertainty in the data distribution. A node with zero entropy is perfectly pure, meaning all samples belong to the same class, while maximum entropy indicates a mix of classes (Breiman et al., 1984).

In this methodology, the training data is used to calculate the optimal predictors and thresholds for each partition that best divide the predictor space to optimize the classification. The probability of each class is estimated based on the samples at the final nodes. DTs can also suffer from overfitting. Typically, an overfitted DT model will have multiple partitions and numerous final nodes, each with very small samples. To avoid this, different criteria can be used to limit the growth of the tree or to "prune" the overfitted tree, such as limiting the number of levels, setting a minimum amount of data required in a node for further partitioning, or limiting the maximum number of final nodes, among others (Esposito et al, 1997).

3.1.3. Random Forest

The Random Forest (RF) method is a widely used ensemble approach. It has been used for forecasting fog occurrence by Pernov and Gueroa (2023) and Zhen et al. (2023). This method consists of the generation of an ensemble of DTs. Individual trees in the ensemble tend to overfit the data. The overfitting is reduced by combining the output of individual trees (e.g., by averaging the probability estimated by each tree for a given set of predictors), thus resulting in a more generalizable model than the single DT model (Fig. 1d).

RF improves performance by introducing diversity among the decision trees in the ensemble. This is achieved by training each tree on different random subsets of the data and by applying additional strategies, such as limiting the number of predictors considered at each split. When this number is smaller than the total number of available predictors, the algorithm selects a random subset of predictors for each split. As a result, even predictors that might not be useful for separating the entire dataset (as often happens with DTs) can still contribute meaningfully by improving splits in specific contexts within individual trees. The samples used to train the trees are obtained by bootstrapping (i.e., randomly resampling with repetition) from the original training sample (Wilks, 2006).

3.2. Validation metrics and reference models

The Brier Score (BS) and the Brier Skill Score (BSS) are usually used to evaluate the skill of probabilistic predictions (Wilks, 2011). However, Benedetti (2009) and Jewson (2004), among several other authors, showed the limitation of the BS when the absolute probability of the event is very small. In these situations, the BS loses resolution, demonstrating little sensitivity to improvements in the model. For this reason, the log-likelihood (Wilks, 2011) was proposed by the latter author to evaluate the performance of probability prediction models:

$$LL = -\sum_{n=1}^N \ln[\theta_n y_n + (1 - y_n)(1 - \theta_n)] \quad (\text{Eq. 4}).$$

The log-likelihood score (LS) of a perfect forecast would be $LS = -\infty$, and a higher value indicates a worse forecast. Also, a log-likelihood skill score (LSS) is defined as:

$$LSS = 1 - \frac{LS}{LS_{ref}} \quad (\text{Eq. 5})$$

where LS_{ref} is the LS of the reference model. In this study, the performance of the prediction models is measured by LS and LSS using a climatological model as a reference.

The climatological model estimates the probability of fog occurrence using its climatological probability (P_c) and the climatological probability conditioned on the month of the year ($P_{c|m}$) and the time of the day ($P_{c|h}$), which are given by the following expressions:

$$P_c = \frac{n_f}{N} \text{ (Eq. 6)}$$

$$P_{c|m} = \frac{n_{f|m}}{N_m} \text{ (Eq. 7)}$$

$$P_{c|h} = \frac{n_{f|h}}{N_h} \text{ (Eq. 8),}$$

with N the total number of samples, N_m the total number of samples for a given month of the year and N_h the total number of samples for a given time of the day (independently of the month), and n_f , $n_{f|m}$, and $n_{f|h}$ the corresponding number of samples with fog conditions.

For radiative fogs, which are one of the most common fog types in this airport (Yabra et al., 2021; 2023; 2025a), persistence is considered a good forecast model for the first few hours. Therefore, the probability of fog conditioned on fog occurrence at the forecast initial time is used as an additional reference model, included for comparison purposes only. This conditional probability is computed as:

$$P_{(f_t|f_0)} = \frac{n_{ff}}{n_{f_0}} \text{ (Eq. 9)}$$

$$P_{(f_t|f_t-f_0)} = \frac{n_{ff}}{n_{ff-f_0}} \text{ (Eq. 10)}$$

In turn, these probabilities were also conditioned on the month of the year and on the hour of the day, separately.

To estimate the statistical significance of the validation metrics, we follow the approach used by Herman and Schumacher (2018) where the bootstrapping technique is used to estimate the confidence limits (10th and 90th percentiles of the score distribution) with 1000 random samples.

3.3. Hyperparameters and features selection

The structure of the ML methods (e.g., the depth of the DT or the number of members in the RF ensemble) and how training is carried out (e.g., the maximum number of iterations before converging to the minimum of the cost function in LR) are problem-dependent parameters and are referred to as hyperparameters, which must be set before training the

models. To this end, the optimal combination of these hyperparameters was sought by using all the available variables for all methods and forecasting horizons. This search was carried out through 10-fold cross-validation of the training set (using a random 70-30 ratio to separate between training and validation), evaluating the LSS corresponding to each combination of hyperparameters. Each hyperparameter was first tested across 10 values spanning a broad range. A second search was then conducted using 10 additional values within the interval that produced the highest LSS. Table 1 shows the best combination of hyperparameters for the three techniques used in this study. It is noteworthy that the optimal trees within the RF are significantly larger than the individual DT, illustrating how the ensemble approach helps to counteract overfitting. It is important to mention that, in the case of RF, not all combinations of hyperparameters were evaluated, but rather they were assessed one at a time, considering the previously evaluated optimal values for each parameter. This was due to time optimization and available computational resources. In most cases, no significant changes were found that were coherent with the forecasting horizon, and they seemed to be random fluctuations due to sampling errors.

Another preliminary step in model configuration involves identifying the optimal combination of predictors for each forecasting horizon and technique. This approach enables efficient model training by excluding predictors that provide little useful information or could negatively affect model performance. To achieve this, an iterative forward feature selection method is applied (Wilks, 2006). The process starts with a model that includes only the single predictor showing the highest predictive power based on the LSS metric. At each iteration, the predictor that yields the greatest improvement in model performance when combined with the existing predictors is added to the set. The new model's skill is then evaluated, and if the improvement is statistically significant at the 90% confidence level (assessed using a bootstrap approach), the iteration continues. Otherwise, the procedure stops, and the model from the previous iteration is retained.

4. Results

4.1. Preset of ML models

Figure 2 shows the Pearson correlation matrix between all the predictors. Some of them exhibit a considerable correlation with fog occurrence in the next 6 hours (e.g., visibility, PW-Fog, PW-Nothing, and 6h-Vis-variation), with correlation decreasing with lead time. The

correlation between the target variables is particularly notable, a consequence of the typical persistence of the fog conditions for 2 or more consecutive hours. The most substantial correlations between predictors and fog occurrence are found in visibility and fog at the forecast initiation time. Other important predictors at the forecast initial time include wind speed with a negative correlation, as stronger winds inhibit fog formation, humidity with a positive or negative correlation depending on whether relative humidity or dew point depression is used, since moist conditions favor fog, and the absence of relevant weather phenomena showed a negative correlation. Additionally, the visibility trend over the past 6 hours and the minimum visibility in the last 24 hours show a strong correlation with fog occurrence. This may be related to the persistence of weather patterns across consecutive days, as previously discussed by Yabra et al. (2025a), and to negative visibility trends that often precede fog formation. The correlation analysis and the definition of the variables were used to guide and simplify the predictor selection method described in Section 3.2. However, the absence of a linear correlation with the target variable does not imply that a variable lacks predictive value, since this study employs nonlinear modeling approaches. For this reason, additional nonlinear analyses are included to better characterize the relationships between the predictors and the target variables.

To complement the predictor importance, considering the possibility of nonlinear relationships between the predictors and the target we analyze whether the probability distribution of a given predictor is different under fog and no-fog conditions in the following hours. To evaluate this, Figure 3 shows the distribution of each variable, conditioned by the occurrence of fog at each forecast lead time. To present all variables in a single figure, the data were standardized using their sample mean and standard deviation. To quantitatively estimate the differences between the distributions, the Kullback–Leibler divergence (KL) is shown below each variable. The KL is a non-negative quantity, defined on the interval $[0, \infty)$. A value of 0 indicates that the two probability distributions are identical, while larger values indicate increasing dissimilarity between them (Kullback and Leibler, 1951). The more different the two distributions are, the higher the KL divergence. As this difference decreases with increasing forecasting horizon, only the 1-hour and 6-hour forecasts are included. The KL divergence is commonly used to quantify how different and biased two distributions are; here, it is used as a proxy for the nonlinear relationship between the predictors and the target variables. Visibility could be a highly effective predictor of fog 1 hour ahead, as both distributions appear to be categorically separated. This suggests that low visibility values are

directly associated with the occurrence or continuation of fog in the following hour. However, this clear separation was not observed in later forecast hours. Another interesting variable is the 6h-Vis variation since the values corresponding to the absence of fog at all forecast times appear to be concentrated around 0. This is often the case when visibility remains optimal (reported as 10 km) for an extended period, resulting in a visibility variation of 0 km. In such cases, fog is not expected in the following hours. Conversely, a decrease in visibility over the past 6 hours is strongly associated with the occurrence of fog in the subsequent hours, especially in the next 1 to 2 hours. Other variables that exhibit distinct distributions are dew point depression and relative humidity. In these cases, the variable ranges show some overlap, but certain values are more likely to correspond to fog than to no-fog conditions in the following hours.

Figure 4 shows the optimal combination of predictors for each forecast lead time and ML technique, indicating the order in which they were selected by the method (Section 3.2). It can be observed that the LR selects a greater number of predictors compared to the other methods, which use a maximum of six predictors. By order of selection, fog occurrence prediction at all forecast lead times is highly influenced by the presence of fog at initialization, indicating the importance of persistence. This is evidenced by the fact that Visibility and PW-Fog were selected among the first two predictors in most models, which also agrees with the high correlations shown in Figure 2. In particular, visibility is the best single predictor for the first three forecast hours in all the models considered. The PW-Fog variable adds new information by distinguishing hours of reduced visibility due to fog from other phenomena.

Another highly relevant variable is the time of day, as fog exhibits a marked diurnal cycle (see first panel of Fig. 8), and its selection order ranks between second and third in almost all models. Humidity and wind speed follow as the next most important features. Information on the occurrence of precipitation at the time of forecast initialization is particularly important for the RF method. While the LR method selects it as a predictor, it does not rank it among the most important, and the DT method does not select it for any forecast horizon. This information may be related to the fog formation mechanism described by Tardif and Rasmussen (2007; Yabra et al., 2025b), where the evaporation of fallen precipitation contributes to air humidification. Similarly, although cloud base height and its variation are not among the main predictors, they were selected to improve model performance in LR and DT. This result could be linked to fog formation caused by the lowering of the cloud base, as also discussed for other region by Tardif and Rasmussen (2007) and locally by Yabra et al. (2025b).

On the other hand, many of the newly introduced variables (6h-Temp and -Vis variation) were selected among the most important predictors by the LR method whereas variables like 24h-Vis-min were selected at all forecasting horizons, although not as the primary predictor. However, their inclusion in the remaining models did not lead to significant improvements, and consequently, they were not selected.

4.2. Evaluation of ML models

All methods, with their respective hyperparameters (Table 1), were trained for each forecast lead time using the predictors previously selected for each model (Fig. 4). Figure 5 presents the reliability diagrams along with the corresponding forecast probability distributions. In a well-calibrated model, forecast probabilities (e.g., 70%) should match the observed relative frequencies (e.g., 70%), which would appear as a diagonal line in the figure. Beyond calibration, the distribution of forecast probabilities (gray bars) indicates the sharpness of the forecast, which refers to the concentration of forecasts near 0 or 1 while remaining calibrated. Ideally, this distribution exhibits an asymmetric U-shape, with a predominance of values near zero and a mean consistent with the climatological probability of fog occurrence. While hints of this U-shaped pattern can be observed at the 1-hour lead time, the occurrence of forecasts near 100% probability decreases with increasing lead time. For the first 3 hours of the forecast, the ML models provide a good reproduction of the probability of fog occurrence. However, from the fourth forecast hour onward, the models show a subtle underestimation of probabilities, more noticeable in the LR method compared to the other two models. However, it is worth mentioning that the highest probability values still provide valuable information for forecasting, given that the climatological probability of fog occurrence is only 4%. On the other hand, the probability histogram of the DT method shows a lack of continuity, which may be because the probabilities obtained from this method correspond to the occurrence probabilities at each terminal node (Fig. 1). In this context, the output variable of this model can be considered categorical ordinal rather than numerical and continuous. The diagrams corresponding to the RF method show that assembling multiple trees partially resolves this issue by producing a more continuous histogram.

Figure 6 shows the performance of each model and forecast horizon for a specific period in September 2018. This period was chosen due to the presence of isolated fog hours within the testing subset. For the first fog event (03-06 UTC), it is evident that the presence of fog at the initialization time is a key factor for a successful forecast, as it directly triggers the forecast

probability, with a 1-hour delay in the 1-hour forecast and up to a 6-hour delay in the 6-hour forecast. In contrast, other variables appear to drive the high probability of occurrence for the second fog event, beginning on the 5th at 10 UTC. The dissipation of this event is also well captured, with all forecasts showing probabilities dropping below 0.4 by 16 UTC. Finally, the isolated fog hour increases the forecast probability in subsequent hours, producing false alarms.

The performance variation of each model as a function of the forecast horizon is illustrated in Figure 9. The LSS quantifies the improvement of the new models over the climatological model. At first sight, it can be observed that both the persistence-conditional model and the ML models lose predictability as the forecast horizon increases. However, all ML models outperform the climatological and persistence-conditional models. The performance of the DT and RF techniques is nearly identical, being indistinguishable in the plot at all forecast lead times. This result seems to suggest that the fog predictability may have reached its limit with this dataset and algorithms, as the ensemble of several models does not demonstrate an improvement over a single model. However, in the following section, it is proved that this is not the case. Meanwhile, the LR technique performs slightly worse than the other models, with the greatest differences between 4-5 hour forecasts. Considering the 1-hour forecast, all trained models showed an improvement of approximately 70% which decreases monotonically to around 32%, for the 6-hour forecast.

Considering that fog occurrence exhibits a marked daily and annual cycle, the previous analysis was extended to include the hourly and monthly variation of the metrics, as shown in Figures 8 and 9, respectively. From the daily variation of LSS, it is evident that prediction skill is closely tied to the hourly frequency of fog occurrence (Fig. 8, first panel). While the 1-hour forecast consistently outperforms the climatological model across nearly all hours of the day, the performance of all ML models decreases mostly during daytime hours (between 15 and 03 UTC), when the probability of fog occurrence is low. In contrast, during the period of maximum fog probability and the following hours (mostly between 09 and 12 UTC), all models show substantial improvement, with skill increasing slightly with forecast horizon. At 03 UTC, there is a noticeable drop in predictability across all models, only shown in the 1 and 2-hour forecasts due to plot limits. As discussed in Yabra et al. (2025a), this may be partly related to the greater weight meteorological observers assign to observations at this time compared to surrounding hours, although this effect is imperceptible in the climatology. As mentioned in Section 4.1, all trained models rely primarily on the presence of fog at the initialization time.

Consequently, the models tend to perform poorly when fog conditions at initialization differ from those at the forecast validity time. This behavior helps explain why the highest LSS values are observed during the hours of maximum fog probability. Finally, although there are no significant differences among the ML models, all of them significantly outperform the persistence-conditional model at every hour.

The annual variation of the LSS is shown in Figure 11. As the forecast horizon increases, the predictability of the models decreases, showing a slight improvement during the winter-spring months (July-November) compared to the summer months. All the ML models exhibit a similar but weak annual cycle, while this cycle is not present in the persistence-conditional model. While this behavior complicates the measurement of the individual performance of each model, it demonstrates that the hourly fog forecast models at this airport would not be sensitive to the annual climatological cycle, thus reducing the complexity of the problem.

4.3. Strategies for improving the forecast

To increase the predictability of the models, different strategies were evaluated that can be implemented without adding significant computational cost. The question was raised whether it is better to train a single general model using all the data or to train multiple models using different smaller subsets. Two strategies were evaluated to answer this: the first divides the data between more and less probable seasons using the monthly probability of occurrence (April to September, and October to March, respectively), and the second divides the data between more and less probable hours using the hourly probability of occurrence (2 to 14 UTC, and 15 to 01 UTC, respectively). The comparison was carried out using the DT technique, chosen for its simple and nonlinear nature and for having shown one of the highest performances among all the studied techniques, and using its combination of predictors that had been selected. Figure 10 shows the performance of each of these strategies and compares them with the basic DT model (Section 4.2). The comparison between the baseline DT model and the variants stratified by season and by time of the day cycles reveals important differences in forecast skill across lead times. For the 1-hour horizon, the baseline model outperforms both stratified approaches, reflecting the dominant role of persistence at very short timescales: current atmospheric conditions alone provide sufficient information, and splitting the data reduces sample size without adding meaningful context. In contrast, at horizons beyond 2 hours, both stratified models consistently surpass the baseline, as the additional seasonal and

diurnal context helps the classifier capture more homogeneous fog-formation regimes and reduces variance within training subsets. The time-of-day model, in particular, shows up to a 20% improvement at the 2-hour lead time, though its performance advantage decreases with longer horizons, while the seasonal split yields more moderate but consistent gains. These results suggest that persistence dominates very short-range fog predictability, whereas incorporating seasonal and diurnal structure becomes increasingly beneficial at longer lead times.

5. Conclusion

The study carried out aimed to establish the predictive potential of a tool based purely on surface observations for short-term fog occurrence forecasting. On the one hand, the hypothesis was confirmed that certain transformations are necessary for the variables reported in ordinary surface observations, although not to the same extent for all ML methods. Techniques such as LR may not handle categorical variables well, like present weather, as it is difficult to establish a logical order, whereas DT and RF should not encounter major problems since space partitions do not require ordering. Also, instead of using month and hour of the day directly, conditional probabilities as a function of month and hour could be used in future studies. This would resolve the issues associated with discontinuities in these variables. Nonlinear methods showed a need for fewer predictors, particularly fewer of the predictors defined in this study. This result means that these methods can find relationships between ordinary variables whose efficiency surpasses the use of more complex variables (such as those defined here). In fact, the analytic and simplest method evaluated in this study largely relied on the new variables for all forecasting horizons. The results also imply that combinations of variables, such as the dew point depression, outperform using individual variables like air and dew point temperatures as predictors in all models. This is because it reduces the number of predictors to consider and simplifies the "boundary" that divides the predictor space between fog and no-fog. From these three methods, it can be inferred that as the complexity of nonlinear models increases, the same forecasting skill can be achieved using simpler predictors. This result raises the future question of whether it is worth choosing an optimal set of hyperparameters for each forecasting horizon or whether a single combination could be used for all horizons simultaneously.

All the evaluated methods demonstrated an improvement in short-term forecasting ability compared to a purely climatological model. This ability decreases as the forecasting

horizon increases up to 6 hours. All the trained models outperformed a persistence-conditional model, with LR being worse than the other two. No significant differences were found between DT and RF. The variability of performances highlights that fog predictability at Ezeiza airport is strongly modulated by the diurnal cycle, but only weakly by the annual cycle. The improvement of all ML models during hours of high fog occurrence confirms the central role of persistence at short lead times, while the marked decline during daytime hours underscores the challenge of predicting rare fog events. The absence of a clear annual cycle in predictability, despite seasonal variations in fog frequency, suggests that the models are more sensitive to short-term atmospheric conditions than to broad climatological signals. However, while differences among the ML models remain insignificant, they consistently outperform the conditional and climatological reference models, demonstrating the robustness of data-driven approaches for very short-range fog forecasting.

The exploration of different improvement strategies found that dividing the problem into two, using months with high and low fog frequency, showed better predictive ability than using a single general model. Other strategies, such as training multiple models according fog monthly and hourly frequency of occurrence, show significant improvements compared to the results obtained from a single model trained from a 2-hour forecast, but not in the 1-hour forecast. Based on these results, the question of whether it is better to train a single general model using all data or train multiple models using different sub-samples seems not to have a single answer. However, dividing the main problem into several models seems to be a promising idea.

From these results, it can be inferred that the prediction ability with this dataset has not reached its maximum potential. Although in the future, this study could be enhanced by using new non-linear methods, such as neural networks, or incorporating other types of surface data, like hourly surface observations from closest airports, it is estimated that the inclusion of additional measurement instruments, such as satellite imagery or output variables from NWP models, could provide greater improvements to the forecasting system by considering the spatial and temporal evolution of fog and the processes that govern it. The models trained in this work could serve as a new operational forecasting tool in the aviation meteorological offices of the SMN, complementing their results with other tools and the experience of forecasters.

559 **Bibliography**

- 560 Anaman, K. A., Lelleyett, S. C., & Avsar, G. S. (2000). Assessing the effect of aviation
561 weather forecasts on fuel expenditures of an international airline. *International Journal of*
562 *Transport Economics*, 27(2), 257–277.
- 563 Baker, R. J., Cramer, J., y Peters, J., 2002. Radiation fog: Ups airlines conceptual models and
564 forecast methods. 11th Conference on Aviation, Range, and Aerospace Meteorology,
565 Portland, OR, Amer. Meteor. Soc., UPS Airlines.
- 566 Bajaj, K., Ushakiran Mannam, Prasad Deshpande, Yash Patil, Arnab Bhattacharya, and
567 Shivam Tripathi. 2025. Forecasting of Fog Index and Prediction Interval using Bayesian
568 Methods. In *Proceedings of the 8th International Conference on Data Science and*
569 *Management of Data (12th ACM IKDD CODS and 30th COMAD) (CODS-COMAD '24).*
570 Association for Computing Machinery, New York, NY, USA, 270–278.
571 <https://doi.org/10.1145/3703323.3703738>
- 572 Baldocchi, D., & Waller, E. (2014). Winter fog is decreasing in the fruit-growing region of the
573 Central Valley of California: Decreasing FOG. *Geophysical Research Letters*, 41(9), 3251–
574 3256. <https://doi.org/10.1002/2014GL060018>
- 575 Bari, D., Bergot, T., y Tardif, R., 2023. Fog Decision Support Systems: A Review of the
576 Current Perspectives. *Atmosphere*, 14 (8), 1314, doi:10.3390/atmos14081314.
- 577 Bartoková, I., Bott, A., Bartok, J., y Gera, M., 2015. Fog Prediction for Road
578 Traffic Safety in a Coastal Desert Region: Improvement of Nowcasting Skills by the
579 Machine-Learning Approach. *Boundary-Layer Meteorology*, 157 (3), 501–516,
580 doi:10.1007/s10546-015-0069-x.
- 581 Benedetti, R., 2009. Scoring rules for forecast verification. 138, 203–211.
- 582 Boneh, T., G. T. Weymouth, P. Newham, R. Potts, J. Bally, A. E. Nicholson, and K. B. Korb,
583 2015: Fog Forecasting for Melbourne Airport Using a Bayesian Decision Network. *Wea.*
584 *Forecasting*, 30, 1218–1233, <https://doi.org/10.1175/WAF-D-15-0005.1>.
- 585 Bergot, T., y Guedalia, D., 1994. Numerical Forecasting of Radiation Fog. Part I: Numerical
586 Model and Sensitivity Tests. *Monthly Weather Review*, 122 (6), 1218, doi:10.1175/1520-
587 0493(1994)122<1218:NFORFP>2.0.CO;2.
- 588 Breiman, L., Friedman, J., Stone, C. J., y Olshen, R., 1984. *Classification and Regression*
589 *Trees*. Chapman and Hall/CRC.
- 590 Bremnes, J. B., y Michaelides, S. C., 2007. Probabilistic Visibility Forecasting Using Neural
591 Networks. *Pure and Applied Geophysics*, 164 (6-7), 1365–1381, doi:10.1007/s00024-007-
592 0223-6.
- 593 Castillo-Botón, C., Casillas-Pérez, D., Casanova-Mateo, C., Ghimire, S., Cerro-Prada, E.,
594 Gutierrez, P. A., Deo, R. C., & Salcedo-Sanz, S. (2022). Machine learning regression and
595 classification methods for fog events prediction. *Atmospheric Research*, 272, 106157.
596 <https://doi.org/10.1016/j.atmosres.2022.106157>

597 Cook, L., Wood, B., Klein, A., Lee, R., y Memarzadeh, B., 2009. Analyzing the Share of
598 Individual Weather Factors Affecting NAS Performance Using the Weather Impacted Traffic
599 Index. doi:10.2514/6.2009-7017, URL <https://arc.aiaa.org/doi/abs/10.2514/6.2009-7017>.

600 Cornejo-Bueno, L., Casanova-Mateo, C., Sanz-Justo, J., Cerro-Prada, E., y Salcedo-Sanz, S.,
601 2017. Efficient Prediction of Low-Visibility Events at Airports Using Machine-Learning
602 Regression. *Boundary-Layer Meteorology*, 165 (2), 349–370, doi: 10.1007/s10546-017-0276-
603 8.

604 Cornejo-Bueno, S., Casillas-Pérez, D., Cornejo-Bueno, L., Chidean, M. I., Caamaño, A. J.,
605 Cerro-Prada, E., Casanova-Mateo, C., y Salcedo-Sanz, S., 2021. Statistical Analysis and
606 Machine Learning Prediction of Fog-Caused Low-Visibility Events at A-8 Motor-Road in
607 Spain. *Atmosphere*, 12 (6), 679, doi:10.3390/atmos12060679.

608 Dietz, S. J., Kneringer, P., Mayr, G. J. & Zeileis, A. (2019). Low-visibility forecasts for
609 different flight planning horizons using tree-based boosting models. *Advances in Statistical
610 Climatology, Meteorology and Oceanography*, 5, 101–114. [https://doi.org/10.5194/ascmo-5-](https://doi.org/10.5194/ascmo-5-101-2019)
611 101-2019

612 Dutta, D., y Chaudhuri, S., 2015. Nowcasting visibility during wintertime fog over the airport
613 of a metropolis of india: decision tree algorithm and artificial neural network approach. *Nat.
614 Hazards.*, (75), 1349–1368, doi:doi.org/10.1007/s11069-014-1388-9.F.

615 Esposito, D. Malerba, G. Semeraro and J. Kay, "A comparative analysis of methods for
616 pruning decision trees," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*,
617 vol. 19, no. 5, pp. 476-491, May 1997, doi: 10.1109/34.589207.

618 Ghirardelli, J. E., y Glahn, B., 2010. The Meteorological Development Laboratory's Aviation
619 Weather Prediction System. *Weather and Forecasting*, 25 (4), 1027–1051,
620 doi:10.1175/2010WAF222312.1.

621 Gultepe, I., et al, 2007. Fog Research: A Review of Past Achievements and Future
622 Perspectives. *Pure and Applied Geophysics*, 164 (6-7), 1121–1159, doi:10.1007/s00024-007-
623 0211-x

624 Gultepe, I., et al, 2014. Ice fog in arctic during fram ice fog project: Aviation and nowcasting
625 applications. *Bulletin of the American Meteorological Society*, 95 (2), 211–226,
626 doi:10.1175/BAMS-D-11-00071.1.

627 Gultepe, I., Heymsfield, A. J., Field, P. R., y Axisa, D., 2017. Ice phase precipitation.
628 *Meteorological Monographs*, 58, 6.1–6.36, doi:10.1175/AMSMONOGRAPHS-D-16-
629 0013.1, URL [https://journals.ametsoc.org/view/journals/amsm/58/1/amsmonographs-d-16-](https://journals.ametsoc.org/view/journals/amsm/58/1/amsmonographs-d-16-0013.1.xml)
630 0013.1.xml.

631 Gultepe, I., y coautores, 2019. A review of high impact weather for aviation meteorology.
632 *Pure and Applied Geophysics*, 176 (5), 1869–1921, doi:10.1007/s00024-019-02168-6.

633 Gultepe, E., Wang, S., Blomquist, B., Fernando, H. J. S., Kreidl, O. P., Delene, D. J., &
634 Gultepe, I. (2024). Machine learning analysis and nowcasting of marine fog visibility using
635 FATIMA Grand Banks campaign measurements. *Frontiers in Earth Science*, 11, 1321422.
636 <https://doi.org/10.3389/feart.2023.1321422>

- 638 Herman, G. R., & Schumacher, R. S. (2016). Using reforecasts to improve forecasting of fog
639 and visibility for aviation. *Weather and Forecasting*, 31(2), 467–482.
640 <https://doi.org/10.1175/WAF-D-15-0072.1>
- 641 ICAO, 2010: Technical specifications related to meteorological observations and reports:
642 Appendix 3. annex 3 to the convention on international civil aviation: Meteorological service
643 for international air navigation, 17th ed.
- 644 Jewson, S. (2004). The problem with the Brier score [Preprint]. arXiv.
645 <https://doi.org/10.48550/arXiv.physics/0401046>
- 646 Kamangir, H., Collins, W., Tissot, P., King, S. A., Dinh, H. T. H., Durham, N., y Rizzo, J.,
647 2021. Fognet: A multiscale 3d cnn with double-branch dense block and attention mechanism
648 for fog prediction. *Machine Learning with Applications*, 5, 100 038,
649 doi:<https://doi.org/10.1016/j.mlwa.2021.100038>.
- 650 Kampa, M.; Castanas, E. Human health effects of air pollution. *Environ. Pollut.* 2008, 151,
651 362–367. [Google Scholar] [CrossRef] [PubMed]
- 652 S. Kullback. R. A. Leibler. "On Information and Sufficiency." *Ann. Math. Statist.* 22 (1) 79 -
653 86, March, 1951. <https://doi.org/10.1214/aoms/1177729694>
- 654 Keith, R., & Leyton, S. M. (2007). An experiment to measure the value of statistical
655 probability forecasts for airports. *Weather and Forecasting*, 22(4), 928–935
- 656 Kneringer, P., Dietz, S. J., Mayr, G. J., y Zeileis, A., 2019. Probabilistic Nowcasting of Low-
657 Visibility Procedure States at Vienna International Airport During Cold Season. *Pure and*
658 *Applied Geophysics*, 176 (5), 2165 2177, doi:10.1007/s00024-018-1863-4.
- 659 Lala, G. G., Mandel, E., y Jiusto, J. E., 1975. A numerical evaluation of radiation fog
660 variables. *Journal of Atmospheric Sciences*, 32 (4), 720 728, doi:10.1175/1520-
661 0469(1975)032<0720:ANEORF>2.0.CO;2.
- 662 Leyton, S. M., y Fritsch, J. M., 2003. Short-Term Probabilistic Forecasts of Ceiling and
663 Visibility Utilizing High-Density Surface Weather Observations. *Weather and Forecasting*,
664 18 (5), 891, doi:10.1175/1520-0434(2003)018<0891:SPFOCA>2.0.CO;2.
- 665 Leyton, S. M., y Fritsch, J. M., 2004. The Impact of High-Frequency Surface Weather
666 Observations on Short-Term Probabilistic Forecasts of Ceiling and Visibility. *Journal of*
667 *Applied Meteorology*, 43 (1), 145 156, doi:10.1175/1520-
668 0450(2004)043<0145:TIOHSW>2.0.CO;2.
- 669 Marzban, C., Leyton, S., y Colman, B., 2007. Ceiling and Visibility Forecasts via Neural
670 Networks. *Weather and Forecasting*, 22 (3), 466, doi:10.1175/WAF994.1.
- 671 McGovern, A., Chase, R. J., Flora, M., Gagne, D. J., Lagerquist, R., Potvin, C. K., Snook, N.,
672 y Loken, E., 2023. A Review of Machine Learning for Convective Weather. *Artificial*
673 *Intelligence for the Earth Systems*, 2 (3), e220077, doi:10.1175/AIES-D-22-0077.1.

675 Miao, Y., Potts, R., Huang, X., Elliott, G., y Rivett, R., 2012. A Fuzzy Logic Fog Forecasting
676 Model for Perth Airport. *Pure and Applied Geophysics*, 169 (5-6), 1107–1119,
677 doi:10.1007/s00024-011-0351-x.

678 Miao, K., Han, T., Yao, Y., Lu, H., Chen, P., Wang, B., & Zhang, J. (2020). Application of
679 LSTM for short-term fog forecasting based on meteorological elements. *Neurocomputing*,
680 408, 285–291. <https://doi.org/10.1016/j.neucom.2019.12.129>

681 Negishi, M., y Kusaka, H., 2022. Development of statistical and machine learning models to
682 predict the occurrence of radiation fog in Japan. *Meteorological Applications*, 29 (2), 2048,
683 doi:10.1002/met.2048.

684 Pasini, A., y Potestà, S., 1995. Short-range visibility forecast by means of neural network
685 modelling: a case-study. *Nuovo Cimento C Geophysics Space Physics C*, 18 (5), 505–516,
686 doi:10.1007/BF02506781.

687 Peláez-Rodríguez, C., Pérez-Aracil, J., Casanova-Mateo, C., & Salcedo-Sanz, S. (2023).
688 Efficient prediction of fog-related low-visibility events with machine learning and
689 evolutionary algorithms. *Atmospheric Research*, 295, 106991.
690 <https://doi.org/10.1016/j.atmosres.2023.106991>

691 Penov, N., y Guerova, G., 2023. So a Airport Visibility Estimation with Two Machine-
692 Learning Techniques. *Remote Sensing*, 15 (19), 4799, doi:10.3390/rs15194799.

693 Roach, W. T., 1995. Back to basics: Fog: Part 2 - The formation and dissipation of land fog*.
694 *Weather*, 50 (1), 7–11, doi:10.1002/j.1477-8696.1995.tb06053.x.

695 Roquelaure, S., y Bergot, T., 2008. A Local Ensemble Prediction System for Fog and Low
696 Clouds: Construction, Bayesian Model Averaging Calibration, and Validation. *Journal of*
697 *Applied Meteorology and Climatology*, 47 (12), 3072, doi:10.1175/2008JAMC1783.1

698 Rudra, R. P., Dickinson, W. T., Ahmed, S. I., Patel, P., Zhou, J., Gharabaghi, B., y Khan, A.
699 A., 2015. Changes in rainfall extremes in ontario. *International Journal of Environmental*
700 *Research*, 9 (4), 1117–1126, doi:10.22059/ijer.2015.1000.

701 Ruiz, J., Schonholz, T., y Saulo, C., 2018. Generación de pronósticos probabilísticos de
702 visibilidad a partir de pronósticos numéricos retrospectivos y observaciones. *Meteorologica*,
703 43 (1), 73–96, URL [http://www.meteorologica.org.ar/wp-content/uploads/2018/07/Ruiz_y-](http://www.meteorologica.org.ar/wp-content/uploads/2018/07/Ruiz_y-otros_Vol43N1.pdf)
704 [otros_Vol43N1.pdf](http://www.meteorologica.org.ar/wp-content/uploads/2018/07/Ruiz_y-otros_Vol43N1.pdf).

705 Schütz, M., Schütz, A., Bendix, J., y Thies, B., 2024. Improving classification-based
706 nowcasting of radiation fog with machine learning based on filtered and preprocessed
707 temporal data. *Quarterly Journal of the Royal Meteorological Society*, 150 (759), 577–596,
708 doi:10.1002/qj.4619.

709 Tabernier, L., E. C. Fernández, A. Tautz, R. Deransy, and P. Martin. 2021. “Fuel Tankering:
710 Economic Benefits and Environmental Impact for Flights up to 1500 NM (Full Tankering) and
711 2500 NM (Partial Tankering).” *Aerospace* 8, no. 2: 37.
712 <https://doi.org/10.3390/aerospace8020037>

713 Tanaka, H.; Honma, S.; Nishi, M.; Igarashi, T.; Teramoto, S.; Nishio, F.; Abe, S. Acid fog
714 and hospital visits for asthma: An epidemiological study. *Eur. Respir. J.* 1998, 11, 1301–
715 1306.

716 Tardif, R., and Rasmussen, R. M., 2007. Event-Based Climatology and Typology of Fog in
717 the New York City Region. *Journal of Applied Meteorology and Climatology*, 46 (8), 1141,
718 doi:10.1175/JAM2516.1.

719 Torrisi, F., Amato, E., Corradino, C., Mangiagli, S., y Del Negro, C., 2022. Characterization
720 of Volcanic Cloud Components Using Machine Learning Techniques and SEVIRI Infrared
721 Images. *Sensors*, 22 (20), 7712, doi:10.3390/s22207712.

722 Vislocky, R. L., and Fritsch, J. M., 1997. An Automated, Observations-Based System for
723 Short-Term Prediction of Ceiling and Visibility. *Weather and Forecasting*, 12 (1), 31–43,
724 doi:10.1175/1520-0434(1997)012<0031:AAOBSF>2.0.CO;2. Wilks, 2006

725 Wilks, D., 2019. *Statistical methods in the atmospheric sciences*. Elsevier.

726 Weyn, J. A., Durran, D. R., & Caruana, R. (2020). Improving data-driven global weather
727 prediction using deep convolutional neural networks on a cubed sphere. *Journal of Advances*
728 *in Modeling Earth Systems*, 12, e2020MS002109. <https://doi.org/10.1029/2020MS002109>

729 WMO, 1954: WMO no-8: Guide to Instruments and Methods of Observation.
730 <https://library.wmo.int/idurl/4/32158>

731 Wu, Y.; Xue, W. Data-Driven Weather Forecasting and Climate Modeling from the
732 Perspective of Development. *Atmosphere* 2024, 15, 689.
733 <https://doi.org/10.3390/atmos15060689>

734 Yabra, M. S., et al, 2021. Las nieblas en los aeropuertos argentinos: revisión de literatura y
735 perspectiva de los pronosticadores. Repositorio Institucional del Servicio Meteorológico
736 Nacional Argentino, (Nota Técnica 2021-89), URL <http://hdl.handle.net/20.500.12160/1540>.

737 Yabra, M. S., de Elía, R., Vidal, L., Nicolini, M. (2023). Fog climatology in Argentina's airports (in
738 Spanish). *Meteorologica*, Vol. 48 N°1 (2023), <https://doi.org/10.24215/1850468Xe020>

739 Yabra, M. S., de Elía, R., Vidal, L., Nicolini, M. (2025a). Observational characterization of fog events
740 over the main Argentine airports. *International Journal of Climatology*. 45, no. 9: e8858.
741 <https://doi.org/10.1002/joc.8858>

742 Yabra, M. S., Nicolini, M., Vidal, L. (2025b). Characterization of influential mechanisms in
743 the life cycle of fog: a study of two extreme events at airports in central-eastern Argentina.
744 *Journal of Southern Hemisphere Earth Systems Science* 75, ES25006.
745 <https://doi.org/10.1071/ES25006>

746 Zhen, M., Yi, M., Luo, T., Wang, F., Yang, K., Ma, X., Cui, S., y Li, X., 2023. Application of
747 a Fusion Model Based on Machine Learning in Visibility Prediction. *Remote Sensing*, 15 (5),
748 1450, doi:10.3390/rs15051450.

750 Zhu, L., Zhu, G., Han, L., y Wang, N., 2017. The application of deep learning in airport
751 visibility forecast. Atmospheric and Climate Sciences, (7), 314 322,
752 doi:10.4236/acs.2017.73023.
753

754 **Tables**

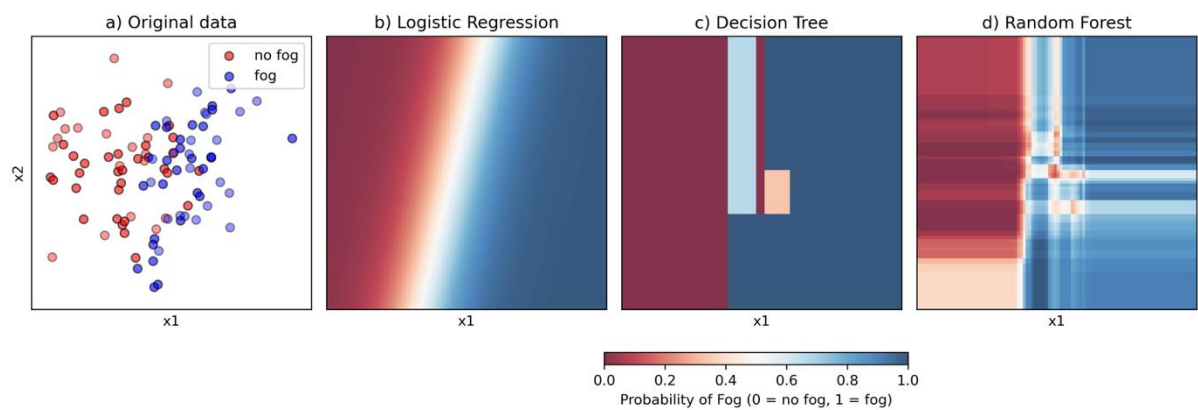
755 *Table 1: Description of the optimal hyperparameters found for each forecast horizon and ML*
756 *method.*

ML method	Hyperparameters	Forecast horizon					
		1h	2h	3h	4h	5h	6h
DT	Splitting criterion	Entropy					
	Maximum depth	7	6	7	6	7	4
	Maximum leaf nodes	40	34	32	28	25	16
RF	Members	100	400	300	300	400	300
	Splitting criterion	Entropy					
	Maximum depth	15	10	15	15	15	10
	Maximum leaf nodes	1000	1000	500	1000	500	500
	Maximum number of variables (%)	0.4	0.4	0.4	0.5	0.5	0.4
	Maximum number of samples used to train each estimator for bootstrap sampling (%)	0.5	0.5	0.7	0.5	0.8	0.3
	Minimum number of data in splitting node (%)	0.001					

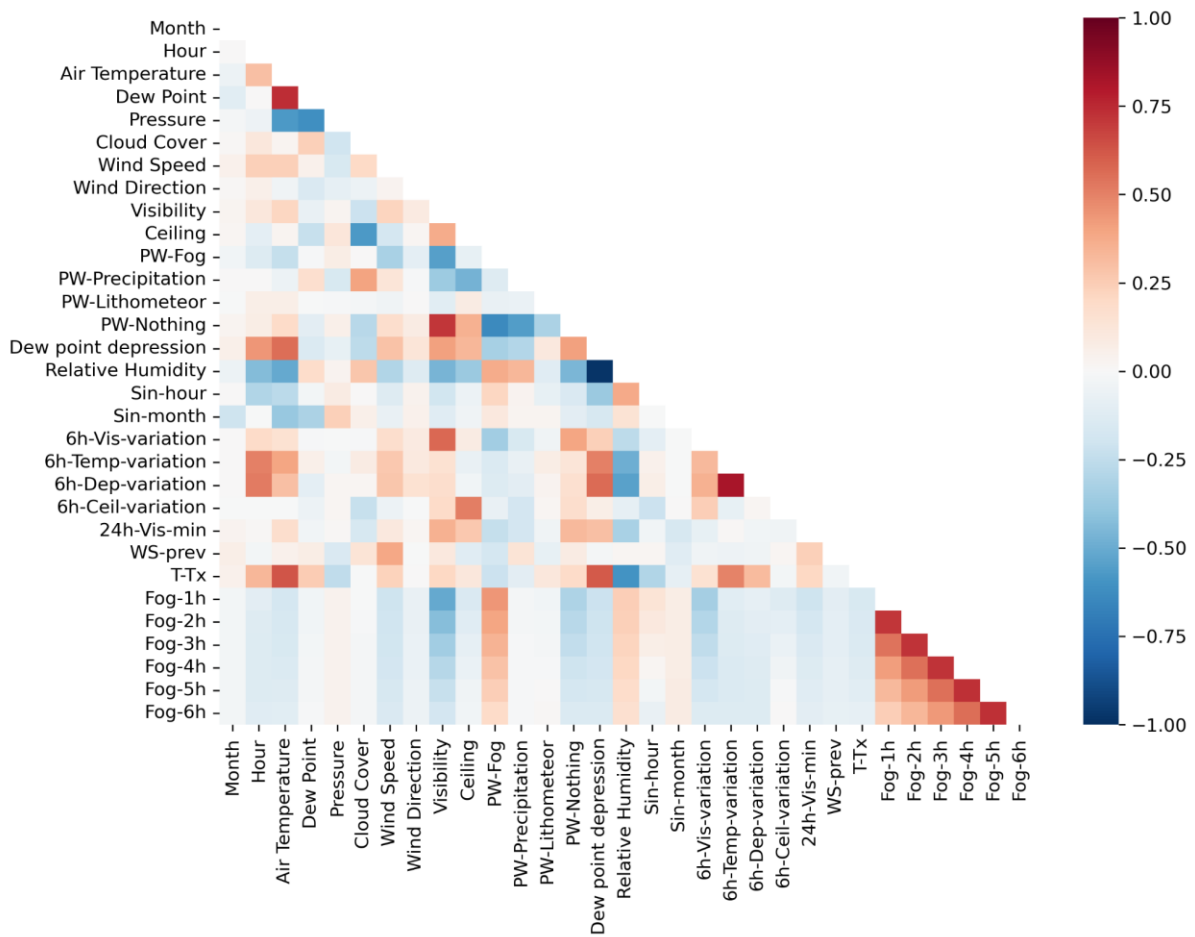
757

758

759 **Figures**



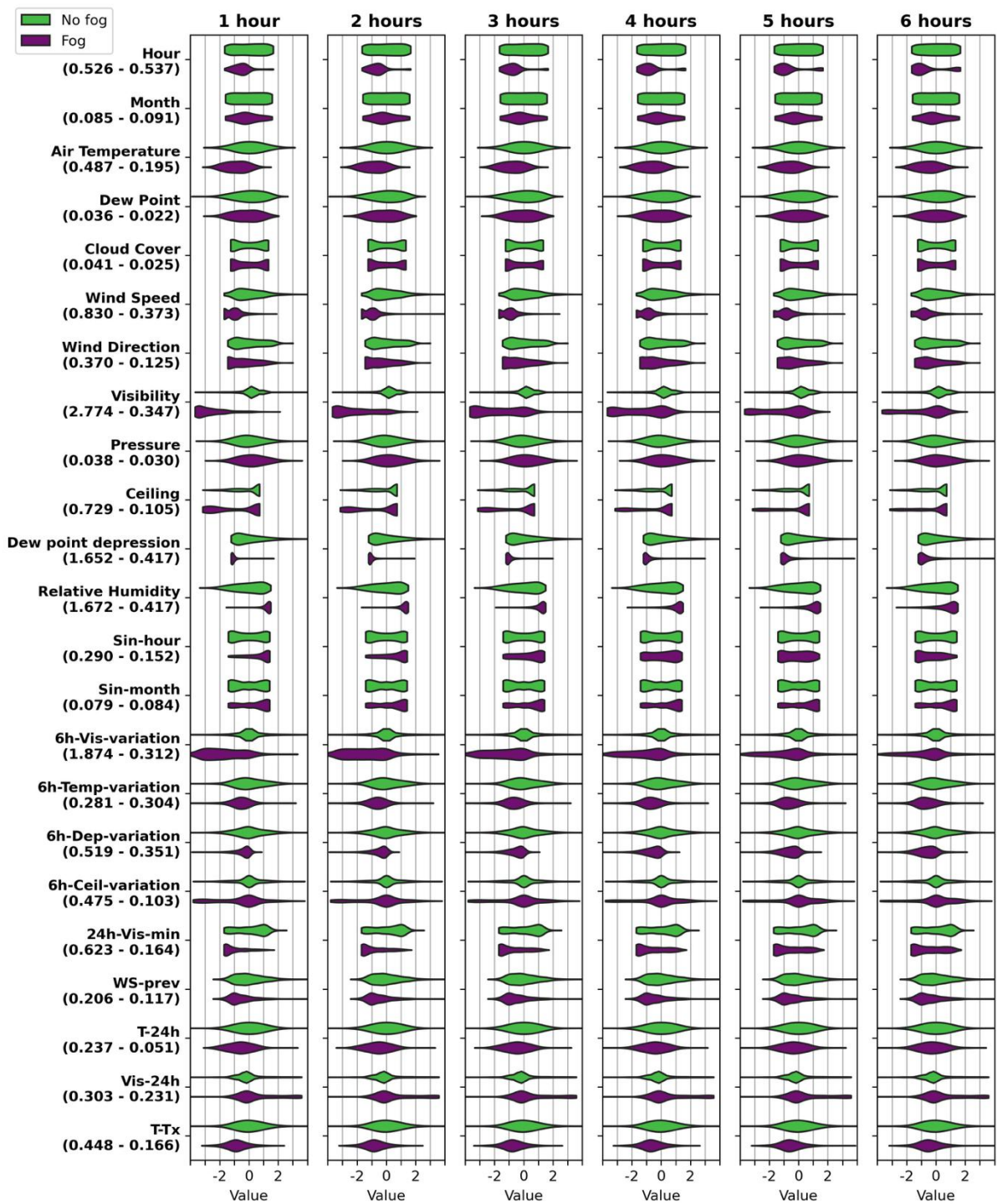
760
761 *Figure 1: Simplified example of the resulting probability estimation for two categories*
762 *represented by the red and blue points (e.g., fog and no fog) in panel a) and their distribution*
763 *as a function of two predictors (x1 and x2, e.g., relative humidity and wind): b) logistic*
764 *regression, c) decision tree, and d) random forest. The color represents the probability of*
765 *having a fog hour.*



770

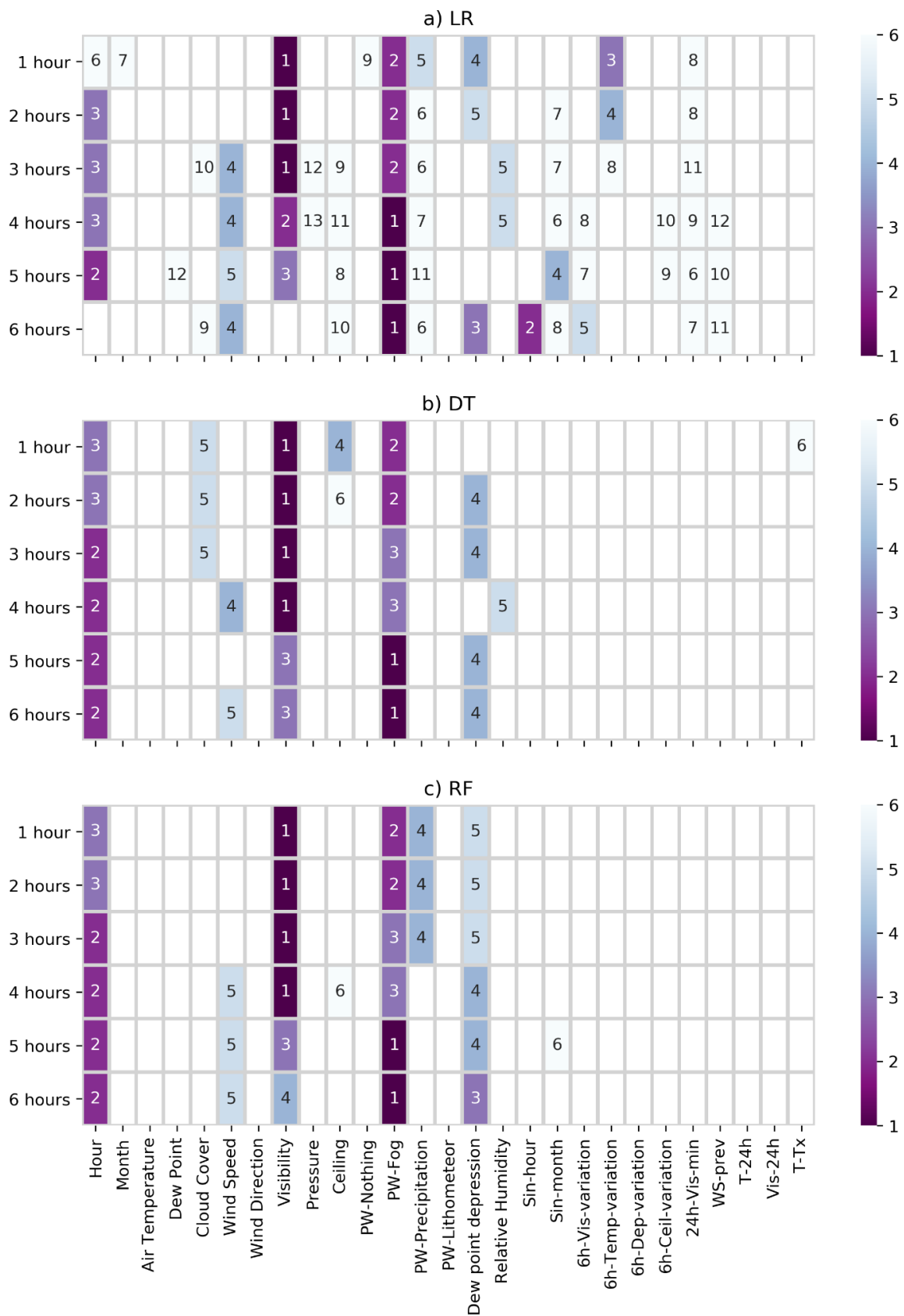
771 *Figure 2: Pearson correlation coefficients between all predictors and target at different*
 772 *forecast lead times between 1 (Fog-1h) and 6 (Fog-6h) hours.*

773



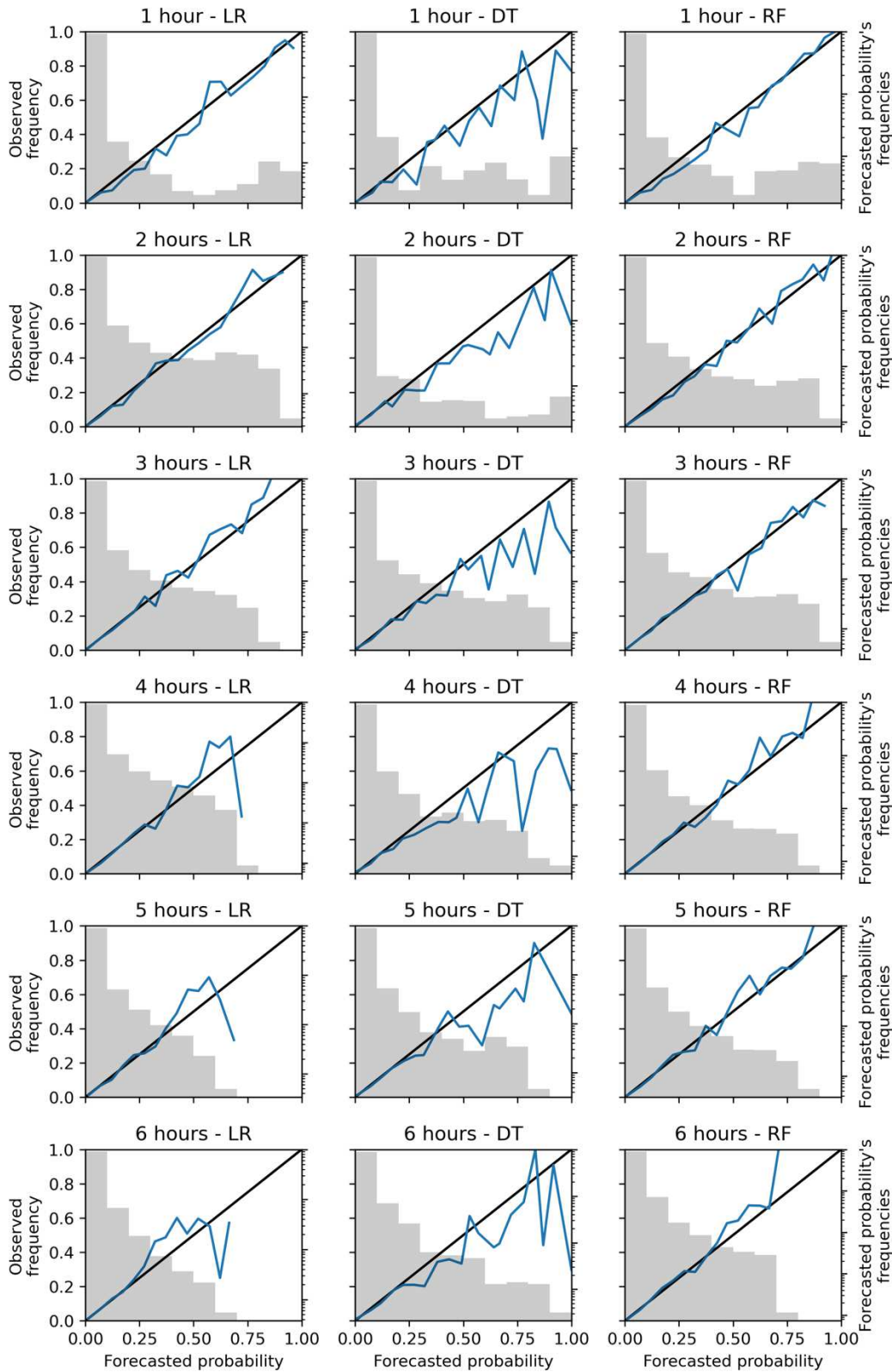
774

775 *Figure 3: Violin plots displaying the distributions of standardized variables, grouped by fog*
 776 *conditions at the forecast time: fog in violet and no-fog in green. Beneath each variable label,*
 777 *the KL-divergence between the two distributions is shown for the 1-hour and 6-hour forecasts,*
 778 *respectively (values in parentheses)*



779

780 *Figure 4: Order in which each variable was selected as a predictor for each forecast horizon*
 781 *and method.*



782

783

784

785

Figure 5: Reliability diagram for all the trained models (blue line), while the black line represents a forecast with perfect calibration. The frequency distribution of forecasted probabilities by the models is shown in gray bars with a logarithmic scale.

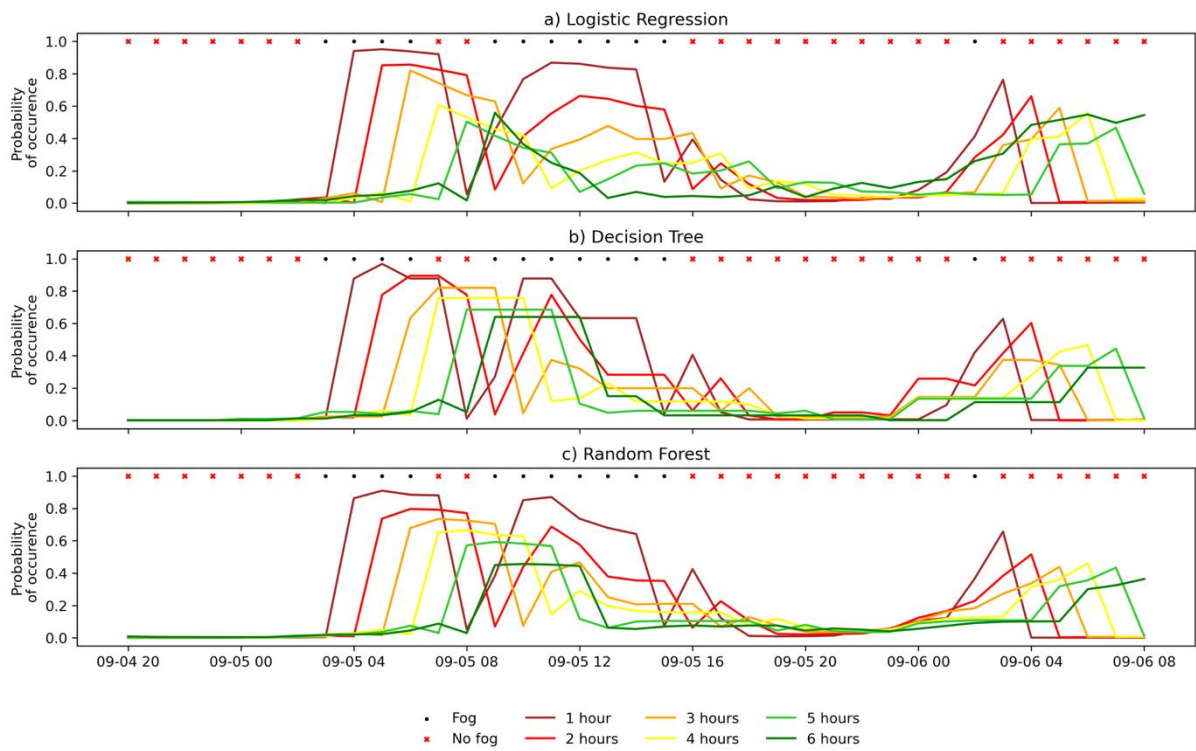
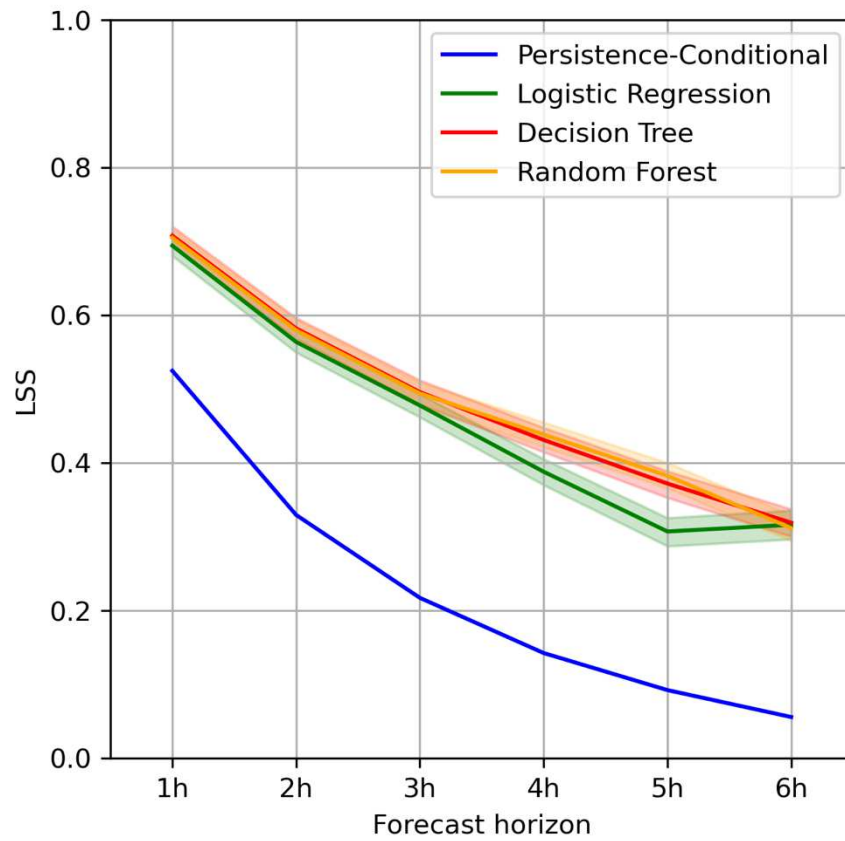


Figure 6: Forecasted fog probability at different lead times (lines) and by different ML methods: LR (a), DT (b), RF (c) over a fog event case in September 2018 (month-day UTC time). The observed fog conditions at each time is shown at the top of each panel: fog as black dots and no fog as red crosses.



792

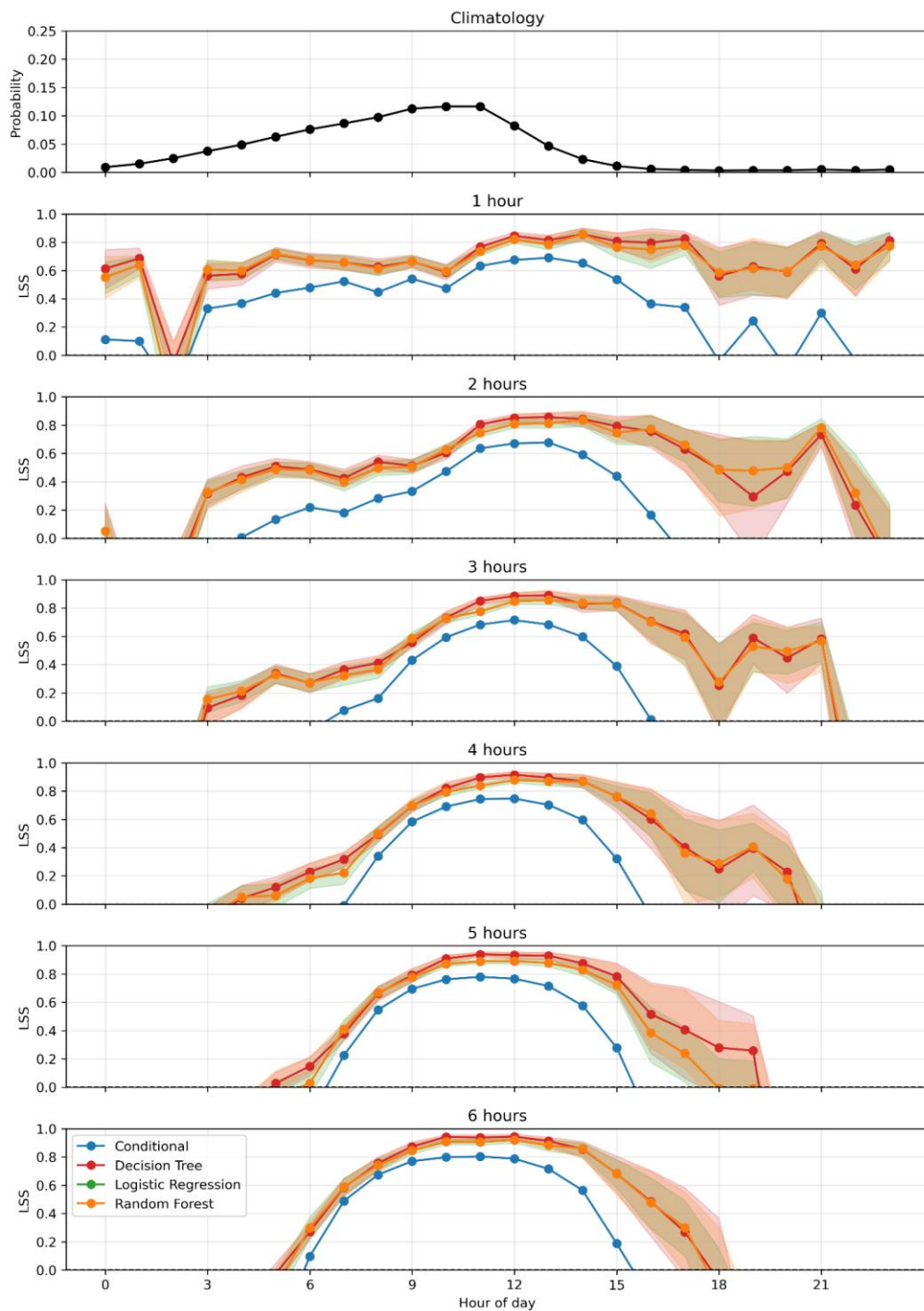
793 *Figure 7: LSS relative to the climatological model, for each model and forecast horizon. The*
 794 *10%-90% confidence interval for the trained models is shown (shaded).*

795

796

797

798



799

800 *Figure 8: Daily variation of the LSS relative to the climatological model, for each model and*
 801 *forecast horizon (panels 2-7). The 10%-90% confidence interval for the trained models is*
 802 *shown (shaded). Pannel 1 shows the fog probability of occurrence (consider the difference in*
 803 *the vertical axis regarding the rest of the panels).*

804

805

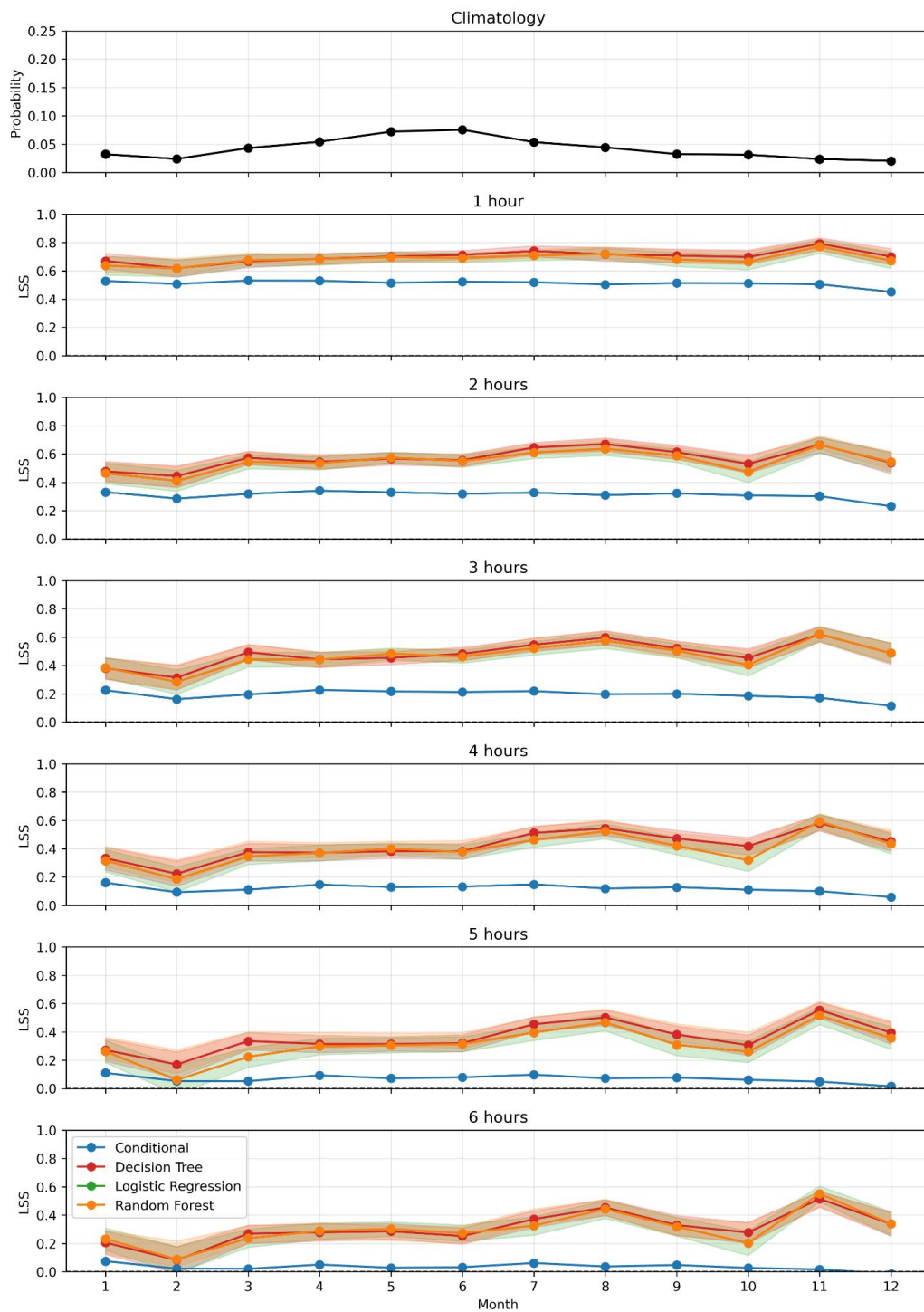
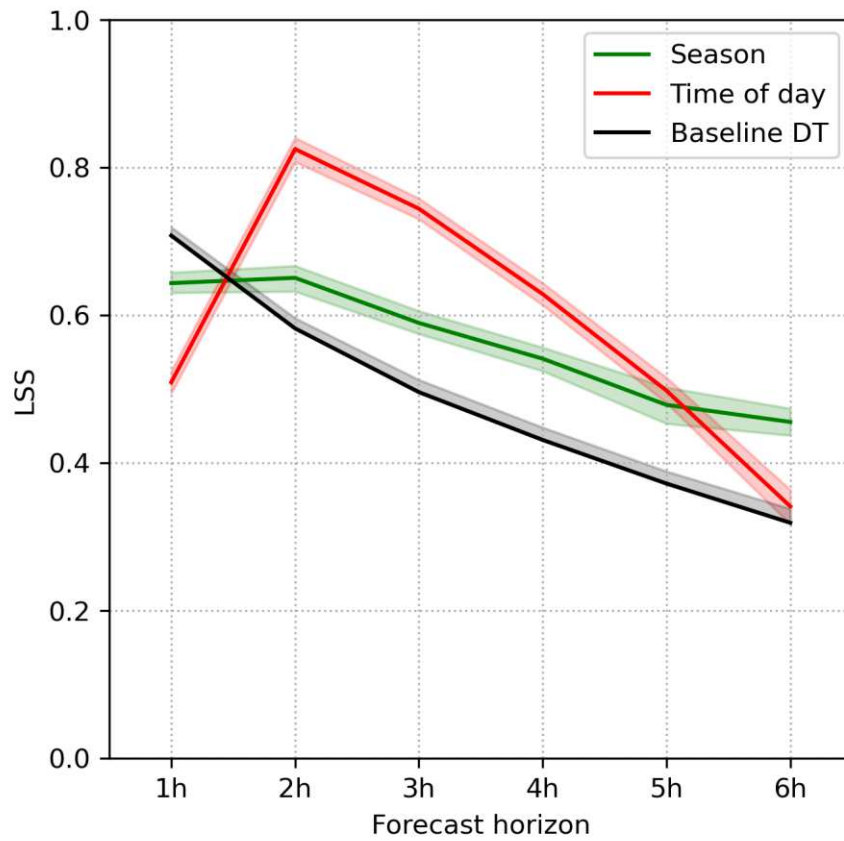


Figure 9: Monthly variation of LSS for each model and forecast horizon (panel 2-7). The 10%-90% confidence interval for the trained models is provided (shaded). Pannel 1 shows the fog probability of occurrence (consider the difference in the vertical axis regarding the rest of the panels).



812

813 *Figure 10: LSS for each forecast horizon for the time of the day strategy (red), and seasonal*
814 *strategy (green), compared to the baseline DT model (black). The 10%-90% confidence*
815 *interval for the trained models is provided (shaded).*

816

817