Pure Appl. Geophys. 178 (2021), 4651–4663

https://doi.org/10.1007/s00024-021-02889-7

Pure and Applied Geophysics

Study of GOES-R Thermodynamic Indices for Short-Term Forecasting of Convective Weather Events Using Machine Learning

Caroline Menegussi Soares,[1] Gutemberg Borges França,[1] Manoel Valdonel de Almeida,[1] and Vinícius Albuquerque de Almeida[1]

Abstract—Forecast models (6–8 h) are proposed for predicting meteorological convective events (MCEs) based on machine learning algorithms for two preestablished areas I and II that correspond to the flight routes from Rio de Janeiro to São Paulo, Brazil. As input data, the models take five atmospheric instability indices, extracted from remotely sensed data retrieved from the GOES-R satellite, for January to March of 2018 and 2019. Temporal characterization for the events was carried out using lightning (i.e., atmospheric discharge) data from 2001 to 2019, and the hourly distribution defined the forecast period of the study. Five machine learning algorithms were selected and trained using 400 experiments aimed at forecasting MCEs and their severity. Statistics indicated that the multilayer perceptron and simple logistic algorithms showed the best forecast performance, with probability of detection of 0.86 and 0.94, false alarm rate of 0.14 and 0.08, bias of 1.00 and 1.01, *F*-measure of 0.86 and 0.94, and Kappa of 0.72 and 0.85 in area I and II, respectively. The hindcasting results for April 2020 reveal that the models can capture physical characteristics related to the convective event patterns and correctly forecast 96% of the days, and are tools that could be used to improve operational forecasts for aviation purposes in this region and elsewhere.

Keywords: Flight routes, atmospheric discharges, convective events, machine learning.

1. Introduction

Severe meteorological convective events (MCEs) have a serious impact on aviation and are often responsible for accidents and incidents worldwide (CENIPA, 2018; Gultepe et al., 2019). MCEs can impact a flight in its different phases; that is, during landing and take-off with the presence of hail and/or heavy rain, gusts, and wind shear, and en route, by the formation of ice and/or clear-air convection-induced turbulence (Gultepe et al., 2019).

Aviation meteorology aims to contribute to ensuring the safety standards, economy, and efficiency of flights. Therefore, it is necessary to determine whether these weather conditions will cause significant variations, such as ice formation, turbulence, thunderstorms (cumulonimbus clouds), low ceilings, low visibility, and wind gusts (or the appearance of wind shears), which may result in deviations in the route and, consequently, higher fuel consumption. As a result, traffic control requires assertive predictions regarding the estimated time for the beginning of these variations and the expected period of their duration.

With the aim of improving short-term predictions, several early studies regarding weather forecasting, such as the ones by Wilson (1966) and Wilk and Gray (1970), suggested forecasting approaches based on extrapolations of radar data limited to subjective interpretation to predict storms.

For applications in aviation, Isaac et al. (2006, 2011, 2012) presented a sequence of analyses that resulted in a refined forecasting system for aviation using data from numerical models, surface observations, radar, satellite, and a microwave radiometer to generate projections of up to approximately 6 h for major airports in Canada.

Nascimento (2004) used profiles of numerical models and radiosonde data to verify atmospheric conditions for the prediction of severe storms and concluded that the convective parameters and indices used in mid-latitudes can be used to elaborate or adjust indices that are more appropriate for the

[1] Laboratory for Applied Meteorology, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil. E-mail: caroline@lma.ufrj.br; gutemberg@lma.ufrj.br; valdonel@lma.ufrj.br; vinicius@lma.ufrj.br

tropical region. Nascimento (2005) presented a comprehensive discussion on the issue of severe weather forecasting in Brazil and a description of atmospheric parameters useful to help identify atmospheric conditions conducive to the occurrence of severe convective storms, and stated that atmospheric discharge (AD) is essential for weather forecasting and climate monitoring. Pinto et al. (2006) presented a description of the spatiotemporal distribution of AD considering data recorded from 1999 to 2004 in the southeast region of Brazil, where the Rio de Janeiro to São Paulo air route studied here is located. They concluded that high-frequency occurrence of AD is associated with the effects of urban heat islands and orography, which enhance the convection mechanism.

Recently, Bender (2018) studied the atmospheric conditions that generate strong storms in the metropolitan region of São Paulo and analyzed the potential for increased convective activity with urban expansion until 2030, concluding that vertical and horizontal urbanization will generate increased precipitation and displacement of severe storms for the region of São Paulo City considering the current projections. Conversely, Paulucci et al. (2019) studied the spatial and temporal variability of cloud-to-ground rays in the metropolitan region of Rio de Janeiro using data from 2001–2016 and observed that, of 258,794 ADs, 64.3% occurred in summer, 20.5% in spring, 12.9% in autumn, and 2.3% in winter.

The area between Rio de Janeiro and São Paulo, located in southeastern Brazil, is highly impacted by MCE, characterized by AD records, primarily during the hottest months of the year between November and April (Pinto et al., 2006).

Presently in Brazil, nowcasting of MCEs is done by meteorologists, who use their experience to integrate different in situ meteorological observations and/or atmospheric model outputs using conceptual models of how the atmosphere works. It is known that this nowcasting procedure suffers from in a high degree of subjectivity and uncertainty. Therefore, the objective of this work is to study the occurrence of MCEs along the Rio de Janeiro–São Paulo air route and evaluate the use of machine learning (ML) algorithms as a tool for short-term forecasting of these events and their severity, based on remotely sensed atmospheric thermodynamic indices from satellites.

This article is part of a sequence of related studies successfully delivered by the Laboratory of Applied Meteorology of the Federal University of Rio de Janeiro (Almeida, 2009; Almeida et al., 2020a, b; França et al., 2016, 2018; Hermsdorff, 2018; Silva et al., 2016) that have utilized machine learning techniques for forecasting convective events.

2. Data and Methods

2.1. Study Region

The study region is a rectangular area of 160,908 km^2, corresponding to the Rio de Janeiro–São Paulo route, which is approximately represented by the geographical position of the two airports, Antonio Carlos Jobim (Rio de Janeiro) and Guarulhos (São Paulo). Considering the representativeness of the atmospheric profile in a 150 km radius circle, according to the World Meteorological Organization, the area was divided using the free computational software QGIS (version 3.14.0, available at https://qgis.org/en/site/), for study purposes, into two areas (I and II), as shown in Fig. 1.

2.2. Data

The data used are from two sources, as described in Table 1.

1. The Geostationary Operational Environmental Satellite (GOES) series formally began on 16 October 1975 with the launch of GOES-A or GOES I. This series is a joint effort between the National Aeronautics and Space Administration (NASA) and the National Oceanic and Atmospheric Administration (NOAA) for the development of idealized geostationary meteorological satellites, of which 19 have been produced (GOES A–U). GOES-Q was not built, and GOES-T and U are planned to be launched in 2021 and 2024, respectively. All previous satellites were successfully launched (https://www.nasa.gov/content/goes-overview/index.htmlabout:blank).

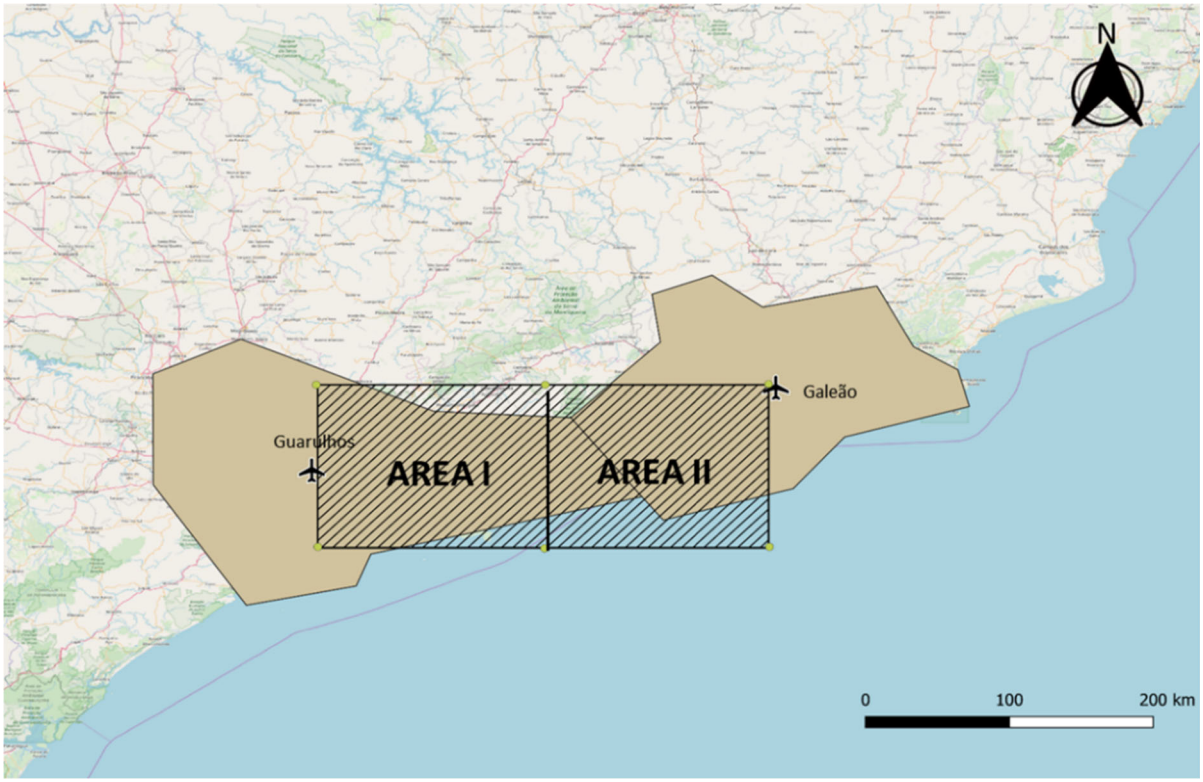

Figure 1
Representation of areas I and II

Table 1

Data used in the study

Source	Description	Temporal resolution	Time (UTC)	Qty	Period
Atmospheric profiles from GOES-R (https://www.ncdc.noaa.gov/airs-web/search)	The values of the set of indices K, Showalter, lifted index, total totals, and CAPE represent 10 km × 10 km pixels in areas I and II	15 min	Hourly from 12 to 16	96,400[a]	January to March of 2018 and 2019
Atmospheric discharge from RINDAT (http://www.rindat.com.br)	Time and location (lat., long.) of atmospheric discharges in areas I and II	300 ns	–	9,945,882 (AD)	2001–2019

[a]The number 96,400 means the number of thermodynamic index values extracted in the period established for the study. On the other hand, 9,945,882 is the number of AD that occurred in the period and in the study area

The GOES-R Algorithm Working Group (http://www.star.nesdis.noaa.gov/goesr/documentation_ATBDs.php) provides information that can be used to calculate derived stability indices from the

retrieved atmospheric profiles, such as the lifted index (LI), convective available potential energy (CAPE), total totals index (TT), Showalter index (SI), and K index values (K). To determine favorable sites for formation of intense storms that give rise to heavy rains, the following indices and thresholds were considered in this study: K index values (George, 1960) above 30 °C have high potential for occurrence of storms; negative LI values (Galway, 1956) are related to unstable conditions; low or negative SI values (Showalter, 1953) indicate more atmospheric instability; TT values above 40 °C (Miller, 1972) imply storm formation potential; and CAPE values greater than 1000 J/kg (Houze, 1993) are related to instability. The indices are determined using only noncloud profiles, representing a 10 km × 10 km pixel, and have been available for download since December 2017. In this study, the listed indices were used as predictor variables of the atmospheric conditions that characterize the potential for MCE occurrence in the regions of interest, areas I and II in Fig. 1.

2. The National Network of Atmospheric Discharge Detection (RINDAT) has been recorded continuously since December 2000, according to the information available at http://simepar.br/rindat/internas/institucional.shtml. The AD data are used here for MCE characterization.

2.3. *Methods*

Considering the potential of GOES-R indices as MCE predictors, two important aspects were addressed: (1) Local thermodynamic indices are not, by themselves, unique predictors to identify the atmospheric pattern of severe MCEs, which have complex thermodynamic formation mechanisms, since some other aspects such as advection, shear, and others could contribute to the mechanisms. Thus, the idea is that the model based on ML can be identified in the set of values of the indices related to MCE atmospheric patterns. (2) If a given MCE appeared at the edge of the area of interest and could impact the area, it is assumed here that the averages of each index used by area, representing the volume of the atmosphere as approximately recommended by the World Meteorological Organization (2013), are a potential predictor of severe MCEs for areas I and II.

Therefore, the challenge in this work is to identify the atmospheric conditions that precede the onset of an MCE for a given period of the day, in other words providing a YES–NO classification for MCE occurrence. The WEKA software package (version 3.7.12) developed by the University of Waikato in New Zealand (Witten et al., 2016) was chosen because it has several classification algorithms based on ML. These algorithms can be trained and tested to identify the atmospheric thermodynamic patterns associated with MCE onset in target areas I and II. The steps of the applied method are as follows:

1. The nowcasting period was analyzed and defined based on the distribution of daily AD occurrence using the dataset in Table 1.
2. The spatial averages of the GOES-R instability indices at 12, 13, 14, 15, and 16 UTC were calculated for areas I and II (Table 1).
3. The dataset of average thermodynamic indices (input of the algorithms to be trained) were concatenated hourly, for 12, 13, 14, 15, and 16 UTC, for area I and II and classified according to the number of recorded AD (YES when $AD > 0$ and NO when $AD = 0$). In other words, the cumulative number of AD was computed between 18:00 and 23:59 h local time. If the cumulative value was greater than zero, the event is classified as YES (MCE event); if there is no atmospheric discharge, the classification is NO (non-MCE event).
4. Testing was carried out via a classical cross-validation procedure consisting of dividing the total dataset into k mutually exclusive subsets (folds) of the same size; from there, a fold is used for testing and the remaining $k - 1$ folds are used for training. For training and testing of all the ML classification algorithms of WEKA, the weights of the occurrence of YES and NO events are adjusted (class balance) in the training dataset, considering the following event proportions (where MCE and non-MCE occurrence are artificially replicated or reduced considering the behavior of the observed value indices); (1) original data, (2) 50% YES and

50% NO, (3) 60% YES and 40% NO, (4) 70% YES and 30% NO, and (5) 80% YES and 20% NO for the different forecast lead times from 12, 13, 14, 15, and 16 UTC, respectively. To balance the output of the model so that the model learns the physical patterns related to MCE and non-MCE events efficiently, the output weights of each model were varied artificially; That is, if the model output has many more values of NO, a greater weight is applied at the YES output. The output weights were chosen randomly based on trial and error. Many tests were carried out to determine the proportion of weight variation at the output of the model that resulted in the best performance.

5. The optimal ML classification algorithm (or model, meaning that a given algorithm was trained and evaluated via the statistics in Table 4) with the best performance in the previous step was selected, considering the values of the statistics specified in Table 3, for further analysis of the results and discussion.
6. Using a trial-and-error procedure, the AD per hour (AD/h) limit for MCE was determined by gradually varying the value of AD/h (threshold) and observing the performance of the ML algorithms in classifying an MCE with an AD equal to or above the limit until reaching the optimum severity limit.
7. Finally, an optimal model was applied to a hindcasting experiment using independent data sampling.

Table 2

Classifiers used

Classifier	Description	Reference in WEKA
Random forest	A collection of tree classifiers that are trained on different subsets of input features; the one with the best performance is chosen	Breiman (2001)
J48	Creates a single decision tree based on all input resources available	Quinlan (1993)
MPL	Multilayer perceptron with one or more hidden layers and an output layer of computational nodes	–
Hoeffding tree	Grants a certain level of confidence in the best attribute to split the tree	Hulten et al. (2001)
Logistic	Predicts the probability of an event belonging to the standard class, which can be of type 0 or 1	Brownlee (2016)
Simple logistic	Builds linear logistic regression models. LogitBoost with simple regression functions as base learners was used for fitting the logistic models	Sumner et al. (2005)

2.4. Selected ML Algorithms

The ML algorithms used in step 5 are briefly described in Table 2.

2.5. Evaluation

The predictions obtained using the ML algorithms were evaluated in comparison with the target values using a two-dimensional contingency table (Table 3) according to Wilks (2006). Thus, it was possible to determine five categorical statistics, namely (1) probability of detection (POD), (2) false alarm (FAR), (3) BIAS, (4) *F*-measure, and (5) KAPPA, as detailed in Table 4.

3. Results

To train and test the classification algorithms for MCE short-term forecasting, as described in Sect. 2.3, the distribution of AD occurrence was initially analyzed for the Rio de Janeiro–São Paulo route. Figure 2a shows the seasonal distribution of ADs from 2001 to 2019. As expected, it was observed that, in the spring–summer period, because of the greater availability of energy, approximately 89% of

Table 3

Contingency matrix

Observed			
	Yes	No	Total
Forecast			
Yes	Hits (a)	False alarm (c)	
No	Misses (b)	Correct negatives (d)	
Total	Observed Yes	Observed No	Total

Table 4

Statistics retrieved from the contingency matrix

Statistic	Formula	Description	References
Probability of detection (POD)	$\mathrm{POD} = \frac{a}{a+c}$	Measures the fraction of observed events that were correctly predicted. A perfect score is 1	Wilks (2006)
False alarm (FAR)	$\mathrm{FAR} = \frac{b}{b+d}$	Measures the fraction of YES predictions in which the event did not occur. A perfect score is 0	–
BIAS	$B = \frac{a+b}{a+c}$	Measures the predicted event frequency as a proportion of the observed event frequency. The perfect score is 1	–
F-measure[a]	$F = \frac{2P \cdot \mathrm{POD}}{P+\mathrm{POD}}$	Measures the accuracy of a test. It considers the accuracy and recall of the test to calculate the score. The F-measure is the harmonic average of the precision and recall, having a best value of 1 (perfect precision)	Sasaki (2007)
Kappa[b]	$k = \frac{p_0 - p_e}{1 - p_e}$	Measures the performance of the classification process, where perfect agreement is 1. This coefficient can be defined as a measure of association It is used to describe and test the degree of agreement in the classification, that is, its reliability and accuracy	Cohen (1960)

[a]where the statistics used are defined as $P = \frac{a}{a+b}$,
[b]$k = \frac{a+d}{a+b+c+d}$ and $= \frac{[(a+b)(a+c)]+[(c+d)(b+d)]}{(a+b+c+d)^2}$.

ADs occur: 57.1% and 31.7% in summer and spring, respectively. In autumn, 6.3% of ADs occurred, whereas 4.8% occurred in winter. This decrease in the number of ADs compared with spring–summer can be explained by the absence of local convective forcing, as these seasons are usually drier. They are characterized by a westward displacement of the South Atlantic Subtropical High, and the associated subsidence hinders convective activities in the study area; thus, ADs are typically related to mass passages (cold or hot fronts).

Figure 2b represents the hourly distribution of AD in the study area and indicates that convective activity increases from 3 pm to peak at 6 pm local time. It is noteworthy that most ADs occurred between 6 pm and 0 am local time, as also found by Almeida et al. (2020c) for the flight terminal area of Rio de Janeiro. Remotely sensed data from GOES-R were retrieved for January–March of 2018 and 2019 (Table 1). Figure 3 shows the number of MCE and AD range for 413 occurrences in the study area. Notably, 79.4% of the events had AD of 130 or less. In this study, severe MCE was defined when an MCE had more than 130 AD.

Training and testing any artificial intelligence algorithm is a time-consuming task (França et al., 2016). Following the method, 56 classification algorithms available in WEKA were trained and tested via cross-validation. To select the classifier(s) algorithm(s) to explore in detail, all of them were initially evaluated with input (from to 12 to 16 UTC for areas I and II). After 280 experiments, the algorithms called naive Bayes (John & Langley, 1995), multilayer perceptron (Quinlan, 1993), Hoeffding tree (Hulten et al., 2001), logistic (Brownlee, 2016), and simple logistic (Sumner et al., 2005) produced superior statistical results and were thus selected.

The experiments are described separately for the MCE forecast (or detection) (method steps 7 and 8) and the severity forecast.

3.1. MCE Forecast

A total of 400 MCE prediction experiments were performed considering the following four aspects: (1) five selected algorithms, (2) dataset configuration 1 to 5, as per step 4 of the method, (3) different lead times (i.e., 12, 13, 14, 15, and 16 UTC), and (4) two areas (I and II) in step 4 of the method.

Table 5 presents the best test results (cross-validation) for the input times at 12, 13, 14, 15, and 16 UTC, identifying the configuration of the training dataset. Therefore, all of the statistics have as the unit the ideal value; 1 – FAR was adopted instead of the classic statistic FAR (whose ideal value is zero). The

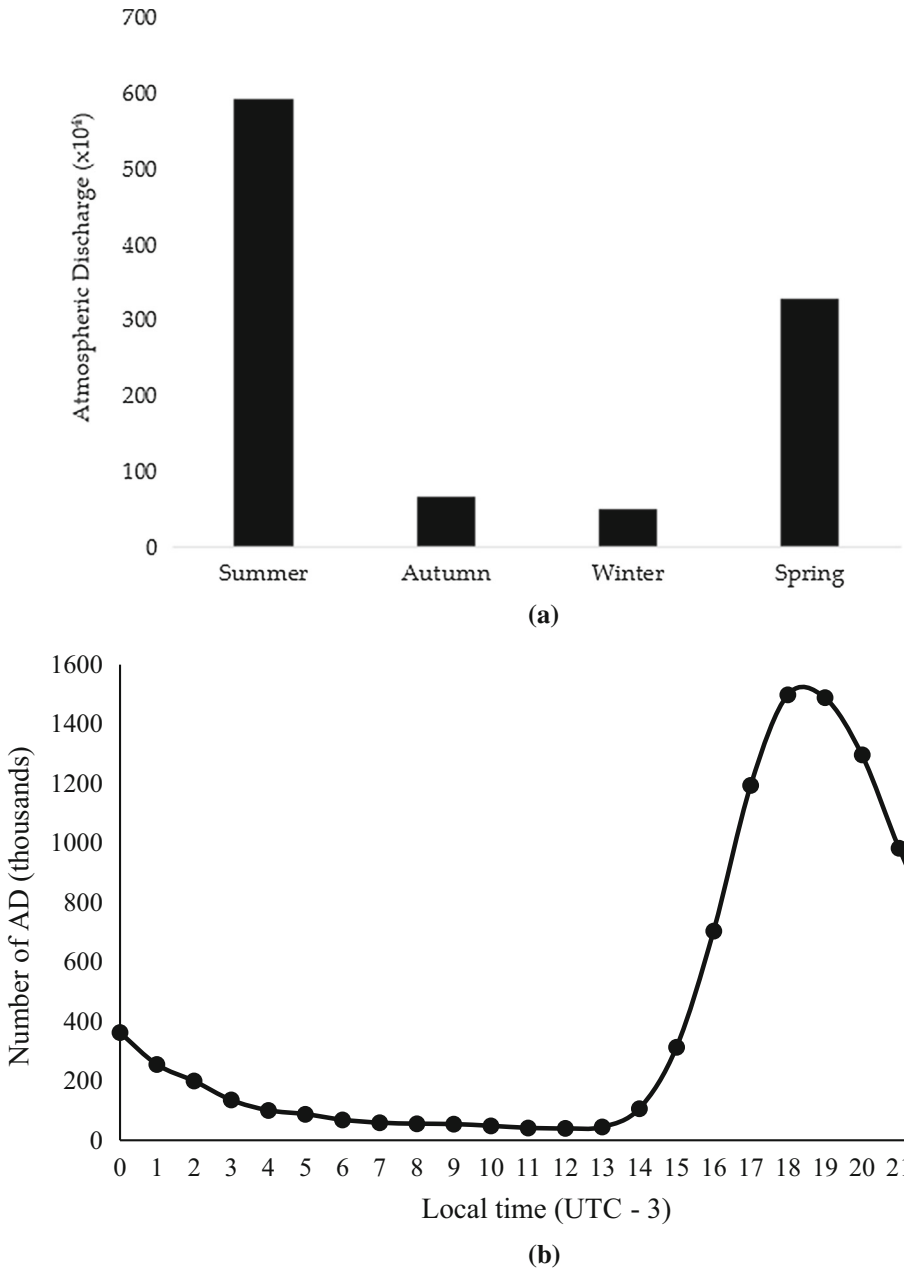

Figure 2
a Seasonal and **b** hourly distribution of atmospheric discharge (AD) for the Rio de Janeiro–São Paulo flight route

performance of the five algorithms was similar regardless of the lead time or area, because the values of POD, 1 – FAR, *F*-measure, BIAS, and KAPPA statistics were in the small ranges of [0.81, 0.94], [0.80, 0.92], [0.81, 0.94], [0.95, 1.01], and [0.61, 0.85], respectively. However, the class weight variation used in step 4 of the method proved to be an interesting strategy for improving the training of the algorithms because the best result (the MLP algorithm whose 6 h probability of detection was on the order of 86%, with a 6 h false alarm rate of 13%, although its forecast was still slightly skewed, but

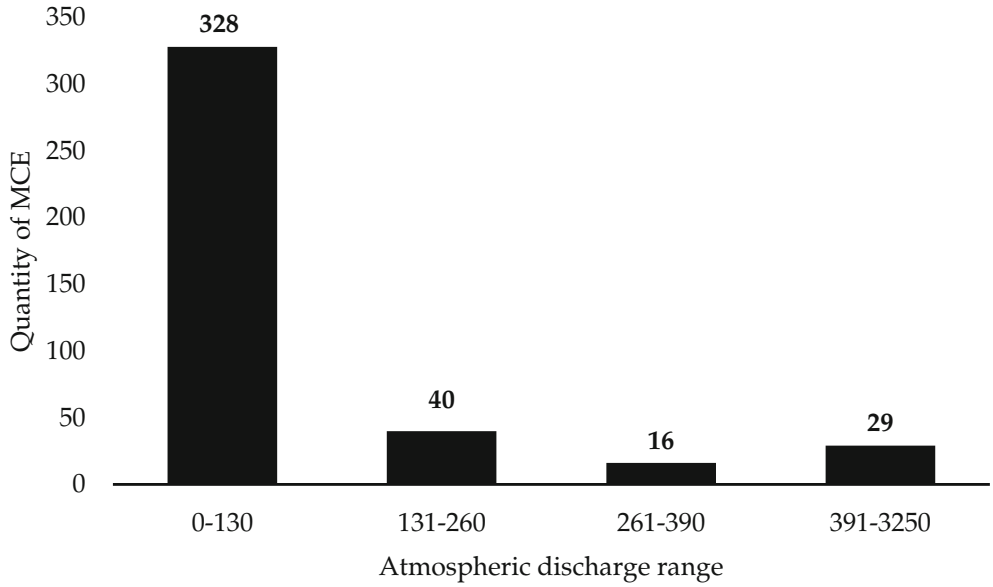

Figure 3
Distribution of 413 MCRs by AD range per hour from January to March of 2018 and 2019

noted only in the third decimal place of the BIAS value) for area I was obtained with the 15 UTC input and dataset configuration 2. For area II, all algorithms showed better performance (the simple logistic algorithm generated an 8 h forecast with slightly skewed MCE with BIAS = 1.01, and a high detection probability and low false alarm rate of 94% and 8%, respectively) with the 13 UTC input and dataset configuration 4.

3.2. Storm Severity

In summary, here two models were developed. The first is a model to predict whether the condition is favorable to convective events. The second is a model to predict the severity of the convective event, assuming that the first model predicted conditions favorable to its formation. For the severity of these events, several procedures were performed, by percentiles, by trial and error, analyzing how the distribution of significant events behaved. It was found that most events, that is, those with the lowest or least significant impacts, occur more frequently (as observed in Fig. 3). After several tests, it was identified that a threshold higher than 130 AD (those of lower occurrence) can be used to determine whether an event is severe or not. In other words, an AD number greater than 130 AD/h indicates a severe event.

Following step 6 of the method, the variation of AD/h from 1 to 3000 was considered in increments of 10 AD/h, to calculate the evaluation statistics in Table 4. The distribution presented in Fig. 3 shows that the highest occurrence of AD was in the range of up to 130 discharges per hour. Therefore, 81 and 328 convective meteorological events were classified as severe and nonsevere, respectively (according to Table 6). This means that the training record sets generated an output of YES for severe MCE but NO for all others. Step 4 was then used to generate five dataset configurations. Finally, the algorithms were trained and tested until the optimal algorithms were obtained.

The J48 (Breiman, 2001) and RF (Quinlan, 1993) algorithms were the ones that obtained the best performance; the values of their statistics are presented in Table 6. Generally, the results for area II are superior to those for area I. Both algorithms were capable of a 6-h forecast of severe MCE (with dataset configuration 2) with POD = 0.78 and FAR = 0.23, with low bias for area I. The same algorithms produced more expressive results in the forecast of

Table 5

Tests of the five selected algorithms at 12, 13, 14, 15, and 16 UTC (and dataset configuration) for areas I and II. The columns represent the statistics probability of detection (P), the difference of false alarm from unit (1 − F), F-measure (F1), BIAS (B), and KAPPA (K)

Area I algorithm	Dataset configuration: (3) for 12 UTC					Dataset configuration: (2) for 13 UTC					Dataset configuration: (2) for 14 UTC				
	P	1 − *F*	*F*1	*B*	*K*	*P*	1 − *F*	*F*1	*B*	*K*	*P*	1 − *F*	*F*1	*B*	*K*
Naive Bayes	0.82	0.80	0.81	1.00	0.61	0.82	0.82	0.82	0.98	0.64	0.81	0.81	0.81	1.00	0.63
Logistic	0.83	0.80	0.83	0.98	0.64	0.81	0.81	0.81	1.00	0.62	0.82	0.82	0.82	1.00	0.63
MPL	0.86	0.80	0.85	0.96	0.69	0.85	0.85	0.84	1.00	0.69	0.82	0.82	0.82	1.00	0.63
Simple logistic	0.82	0.78	0.82	0.98	0.62	0.81	0.81	0.81	1.01	0.62	0.82	0.82	0.82	1.00	0.64
Hoeffding tree	0.82	0.80	0.82	1.00	0.61	0.82	0.82	0.82	1.00	0.64	0.81	0.81	0.81	1.00	0.63

Area I algorithm	Dataset configuration: (2) for 15 UTC					Dataset configuration: (2) for 16 UTC				
	P	1 − *F*	*F*1	*B*	*K*	*P*	1 − *F*	*F*1	*B*	*K*
Naive Bayes	0.83	0.83	0.83	1.00	0.66	0.82	0.82	0.82	1.00	0.64
Logistic	0.85	0.85	0.85	1.00	0.69	0.83	0.83	0.83	1.02	0.65
MPL	0.86	0.86	0.86	1.00	0.72	0.84	0.84	0.84	1.00	0.69
Simple logistic	0.83	0.83	0.83	1.00	0.67	0.85	0.85	0.85	1.00	0.70
Hoeffding tree	0.83	0.83	0.83	1.00	0.66	0.82	0.82	0.82	1.00	0.64

Area II algorithm	Dataset configuration: (2) for 12 UTC					Dataset configuration: (4) for 13 UTC					Dataset configuration: (3) for 14 UTC				
	P	1 − *F*	*F*1	*B*	*K*	*P*	1 − *F*	*F*1	*B*	*K*	*P*	1 − *F*	*F*1	*B*	*K*
Naive bayes	0.83	0.84	0.83	1.00	0.66	0.90	0.84	0.90	0.98	0.76	0.87	0.85	0.87	1.00	0.73
Logistic	0.82	0.82	0.82	1.00	0.64	0.91	0.85	0.91	0.98	0.77	0.86	0.83	0.85	0.98	0.69
MPL	0.82	0.82	0.81	1.00	0.63	0.90	0.79	0.90	0.92	0.74	0.91	0.87	0.91	0.97	0.80
Simple logistic	0.83	0.83	0.83	1.00	0.65	0.94	0.92	0.94	1.01	0.85	0.86	0.84	0.86	0.99	0.71
Hoeffding tree	0.83	0.83	0.83	1.00	0.66	0.90	0.84	0.90	0.98	0.75	0.87	0.85	0.87	0.99	0.73

Area II algorithm	Dataset configuration: (2) for 15 UTC					Dataset configuration: (4) for 16 UTC				
	P	1 − *F*	*F*1	*B*	*K*	*P*	1 − *F*	*F*1	*B*	*K*
Naive bayes	0.84	0.84	0.84	1.00	0.68	0.86	0.82	0.86	1.00	0.67
Logistic	0.83	0.83	0.83	1.00	0.66	0.88	0.81	0.87	0.95	0.70
MPL	0.83	0.83	0.83	1.00	0.66	0.93	0.86	0.93	1.00	0.83
Simple logistic	0.86	0.86	0.85	1.00	0.71	0.88	0.79	0.88	0.96	0.70
Hoeffding tree	0.84	0.84	0.84	1.00	0.68	0.87	0.82	0.87	1.01	0.68

the severity of MCEs for area II. The RF generated an 8 h prediction of severe MCE with POD = 92% (14% higher than area I) and a false alarm rate of 7% (16% higher than area I) with a positive bias of 6% (approximately 6% less than area I).

3.3. Hindcast Results

Considering the definition of severe MCE, the RF algorithm was applied to both areas. Tables 7 and 8 show the daily records of the thermodynamic index values at 15 UTC (area I) and 13 UTC (area II), corresponding to columns 2–6, the amount of AD recorded (column 7), and the MCE forecast output (column 8) for days with data available for areas I and II in April 2019, respectively. The results reveal that the models obtained expressive and similar hindcasts and could correctly identify, 8 h in advance, 100% of non-MCE days and 96% of MCE days. The observed forecast errors correspond to only one MCE day for each area, that is, day 13, in area I, and day 28 in area II, as seen in column 8 of Tables 7 and 8, respectively. To evaluate the model failures, the instability index values were analyzed, and it is noteworthy that these are compatible with those that occur in typical MCE days, except for the CAPE

Table 6

Tests of performance of the two selected algorithms at 15 UTC (for area I) and 13 UTC (for area II) with a specific dataset configuration

Area I algorithm	Dataset configuration: (1) for 15 UTC					Dataset configuration: (2) for 15 UTC				
	POD	1 – FAR	*F*-measure	BIAS	KAPPA	POD	1 – FAR	*F*-measure	BIAS	KAPPA
J48	0.90	0.56	0.89	0.80	0.58	0.78	0.78	0.77	1.00	0.58
Random forest	0.90	0.56	0.89	0.80	0.58	0.78	0.78	0.77	1.00	0.58
Area II algorithm	Dataset configuration: (1) for 13 UTC					Dataset configuration: (2) for 13 UTC				
	POD	1 – FAR	*F*-measure	BIAS	KAPPA	POD	1 – FAR	*F*-measure	BIAS	KAPPA
J48	0.88	0.90	0.89	1.08	0.73	0.92	0.92	0.92	1.00	0.84
Random forest	0.92	0.93	0.92	1.06	0.80	0.93	0.93	0.93	1.00	0.85

Table 7

Values of thermodynamic indices and atmospheric discharges for area I in April 2019

DAY	CAPE	KI	LI	SI	TT	AD	Forecast result
01	4.54	26.47	0.52	4.73	40.66	0	Correct
02	0.00	13.23	4.68	8.35	35.49	0	Correct
03	0.00	20.81	2.00	6.95	37.89	0	Correct
04	0.00	22.28	3.75	6.00	37.21	0	Correct
05	44.78	35.28	−0.99	0.25	44.45	985	Correct
06	65.06	35.61	−1.06	−0.30	43.66	29,271	Correct
07	159.55	35.32	−1.94	−0.02	43.94	3655	Correct
08	No information						
09	4.98	32.04	0.01	0.66	43.16	9	Correct
10	1.62	25.81	1.43	3.22	41.29	3	Correct
11	0.01	23.10	2.63	5.06	38.61	0	Correct
12	0.00	18.55	3.16	6.27	37.06	0	Correct
13	0.00	24.24	0.54	3.56	41.09	175	Not correct
14	No information						
15	No information						
16	92.83	31.97	−0.88	0.98	43.85	26,244	Correct
17	247.71	25.40	−0.33	1.54	43.63	5450	Correct
18	2.24	1.49	4.38	8.83	36.48	0	Correct
19	2.59	−9.03	5.32	10.94	35.96	0	Correct
20	0.00	3.89	7.44	10.90	31.99	0	Correct
21	0.00	23.85	3.77	3.85	40.04	88	Correct
22	5.25	32.04	0.41	1.39	43.59	0	Correct
23	No information						
24	No information						
25	20.84	27.45	0.66	4.16	39.51	4	Correct
26	78.03	29.23	−0.60	2.31	42.23	849	Correct
27	5.32	31.71	−0.03	1.23	42.92	0	Correct
28	7.45	29.49	−0.05	1.98	43.30	6286	Correct
29	0.00	18.86	2.42	5.70	38.34	4	Correct
30	0.00	3.28	3.53	8.37	37.46	0	Correct

Table 8

Values of thermodynamic indices and atmospheric discharges for area II in April 2019

Day	CAPE	KI	LI	SI	TT	AD	Forecast result
01	16.30	25.71	0.37	5.60	40.80	0	Correct
02	0.00	12.21	3.61	9.02	35.42	0	Correct
03	0.27	15.31	0.66	7.59	38.97	0	Correct
04	0.00	15.44	2.31	8.75	33.70	0	Correct
05	21.07	31.21	−0.77	2.46	42.41	2620	Correct
06	16.70	35.89	−0.75	−0.02	43.42	345	Correct
07	282.99	26.86	−2.57	2.87	43.59	5735	Correct
08	No information						
09	55.59	33.98	−0.66	0.91	42.94	1743	Correct
10	5.18	29.42	0.25	2.66	41.72	0	Correct
11	0.00	23.31	2.13	6.83	36.82	0	Correct
12	0.00	18.25	3.40	8.46	34.02	0	Correct
13	4.89	25.47	0.62	4.07	39.55	0	Correct
14	No information						
15	No information						
16	31.69	29.03	−0.29	2.58	41.99	28	Correct
17	410.52	27.65	−1.44	1.34	44.29	17,407	Correct
18	23.58	9.39	2.58	7.67	36.42	11	Correct
19	53.33	0.81	2.35	7.71	39.74	0	Correct
20	0.00	−18.02	3.66	14.15	29.21	0	Correct
21	0.00	19.00	5.12	5.92	37.27	0	Correct
22	1.72	29.14	0.40	3.09	41.75	0	Correct
23	No information						
24	No information						
25	37.47	25.44	0.24	5.38	38.68	0	Correct
26	16.43	24.55	0.43	5.52	39.09	0	Correct
27	3.36	28.72	1.25	3.72	40.22	0	Correct
28	0.00	26.38	1.67	3.78	40.56	1205	Not correct
29	1.82	21.07	0.72	4.72	39.72	5	Correct
30	0.70	10.60	2.93	5.72	38.16	0	Correct

index value, which was null with positive LI (atypical for MCE days according to the data in Sect. 2.2). Intuitively, a plausible reason is that there was little quality control in the data and, for YES results, the calculation of the CAPE and LI values was obtained with corrupted data, resulting in erroneous hindcasts for the specific days.

4. Conclusions

A forecast model based on machine learning algorithms for short-term meteorological convective event forecasting was developed for the Rio de Janeiro–São Paulo route. The main results are summarized below:

- The 6 pm local time period was identified as the period with the highest convective activity.
- The models with input data at 10 am and 12 pm local time showed the best performance.
- The developed models demonstrate the ability to predict and classify the severity of MCE and non-MCE with high probability of detection, low false alarm rate, and bias close to unity (unbiased model).
- Atmospheric profiles extracted from orbital remote sensors are promising for characterizing the thermodynamic states of the atmosphere at high frequency (every 15 min), thus providing essential input data for any predictive models for MCEs, since the results of the 6 h and 8 h hindcasts of the model (trained algorithms) offered satisfactory nowcasting prediction performance for identifying MCE and non-MCE in the studied area.

To use these models in the future as an operational tool by Brazilian aviation, the data period has been extended. In addition, the methodology developed here will be applied to other flight routes and areas with scarce sounding data.

A great advantage of using machine learning is that numerical time models do not solve phenomena on a spatial and time (nowcasting) scale. In addition, this approach can be utilized to obtain a local forecast without using parameterizations, using only the observed data. Another advantage is that ML models have lower computational cost, in addition to reducing the forecast uncertainty.

Atmospheric physics and dynamics suggest a good representation, because the model is fed with local thermodynamic conditions.

Studies such as those by Almeida et al. (2020c) and Anochi et al. (2021), among others, show how important it is to consider machine learning models for forecasting in operational centers to improve the quality of forecasts and reduce computing costs.

Acknowledgements

The authors are grateful to FURNAS Centrais Elétricas for providing the AD data and DECEA for insights from weekly discussions and biannual conferences.

Author contributions C.M. spent much effort writing the original draft and suggested corrections. In addition, G.F. played a critical role in funding acquisition and project administration, in addition to his technical contributions. The remaining authors contributed equally to reviewing and editing the manuscript and the investigation. All authors read and agreed to the published version of the manuscript.

Funding

This study was funded by the Department of Airspace Control (DECEA) via the Brazilian Organization for Scientific and Technological Development of Airspace Control (CTCEA) (grant 002-2018/COPPETEC_CTCEA).

Data availability

All data and materials as well as software application or custom code support their published claims and comply with field standards.

Declarations

Conflict of Interest There are no conflicts of interest.

References

Almeida, M. V. (2009). Aplicação de técnicas de redes neurais artificiais na previsão de curtíssimo prazo da visibilidade e teto para o aeroporto de Guarulhos–SP. Doctoral thesis. Federal University of Rio de Janeiro. http://www.coc.ufrj.br/teses-de-doutorado/153-2009/1186-manoel-valdonel-de-almeida. Accessed 20 August 2020.

Almeida, V. A., França, G. B., & Velho, H. F. C. (2020a). Short-range forecasting system for meteorological convective events in Rio de Janeiro using remote sensing of atmospheric discharges. *International Journal of Remote Sensing*. https://doi.org/10.1080/01431161.2020.1717669

Almeida, V. A., França, G. B., & Velho, H. F. C. (2020b). Data assimilation for nowcasting in the terminal area of Rio de Janeiro. *Ciência e Natura*. https://doi.org/10.5902/2179460X53224

Almeida, V. A., França, G. B., Velho, H. F. C., & Ebecken, N. F. F. (2020c). Artificial neural network for data assimilation by WRF model in Rio de Janeiro, Brazil. *Revista Brasileira De Geofísica*. https://doi.org/10.22564/rbgf.v38i2.2042

Anochi, J. A., Almeida, V. A., & Velho, H. F. C. (2021). Aprendizado de máquina para modelagem de previsão de precipitação climática na América do Sul. *MDPI Remote Sensing*. https://doi.org/10.3390/rs13132468

Bender, A. (2018). Condições Atmosféricas Conducentes a Tempestades Severas e sua Relação com a Urbanização na RMSP. Doutoral thesis. Universidade de São Paulo. https://doi.org/10.11606/T.14.2019.tde-18042019-101002

Breiman, L. (2001). Random forests. *Machine Learning*. https://doi.org/10.1023/A:1010933404324

Brownlee, J. (2016). *Master machine learning algorithms*. Melbourne: Machine Learning Mastery.

CENIPA (2018). Aeródromos—Sumário Estatístico 2008–2017. Available online: file:///C:/Users/Windows%2010/Downloads/sumario_estatistico_aerodromos.pdf. Accessed 08 July 2020.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. https://doi.org/10.1177/001316446002000104

França, G. B., Almeida, M. V., & Rossete, A. C. (2016). An automated nowcasting model of significant instability events in the flight terminal area of Rio De Janeiro, Brazil. Atmospheric Measurement Techniques. https://doi.org/10.5194/amt-9-2335-2016

França, G. B., Bonnet, S. M., & Albuquerque Neto, F. L. (2018). Nowcasting model of low wind profile based on neural network using SODAR data at Guarulhos Airport, Brazil. *International Journal of Remote Sensing*. https://doi.org/10.1080/01431161.2018.1425562

Galway, J. G. (1956). The lifted index as a predictor of latent instability. *Bulletin of the American Meteorological Society*. https://doi.org/10.1175/1520-0477-37.10.528

George, J. J. (1960). *Weather forecasting for aeronautics*. Academic Press.

Gultepe, I., Sharman, R., Williams, D., Zhou, P., Ellrod, B., Minnis, G., et al. (2019). A review of high impact weather for aviation meteorology. *Pure and Applied Geophysics*. https://doi.org/10.1007/s00024-019-02168-6

Hermsdorff, J. (2018). Previsão de instabilidade atmosférica significativa usando árvore de decisão na região metropolitana do Rio de Janeiro. Masters' dissertation. Federal University of Rio de Janeiro. Accessed 20 August 2020.

Houze, R. A. (1993). *Cloud dynamics*. Academic Press.

Hulten G., Spencer L., & Domingos, P. (2001). Mining time-changing data streams. *KDD '01: Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining*. https://doi.org/10.1145/502512.502529

Isaac, G. A., Bailey, M., Boudala, F., Burrows, W. R., Cober, S. G., Crawford, R. W., et al. (2012). The Canadian airport nowcasting system (CAN-Now). *Meteorological Applications*. https://doi.org/10.1002/met.1342

Isaac, G. A., Bailey, M., Boudala, F., Cober, S. G., Crawford, R., Donaldson, N., et al. (2011). Decision making regarding aircraft de-icing and in-flight icing using the Canadian airport nowcasting system (CAN-Now). *SAE International*. https://doi.org/10.4271/2011-38-0029

Isaac, G. A., Bailey, M., Cober, S. G., Donaldson, N., Driedger, N., Glazer, et al. (2006). Airport Vicinity Icing and Snow Advisor (AVISA). *44th AIAA Aerospace Sciences Meeting and Exhibit*. https://doi.org/10.2514/6.2006-1219

John, G., H. & Langley, P. (1995). Estimating continuous distributions in Bayesian classifiers. *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann.

Miller, R. C. (1972). Notes on analysis and severe storm forecasting procedures of the Air Force Global Weather Central. *Technical Report 200*. Air Weather Service: United States Air Force.

Nascimento, E. L. (2004). Identifying Severe Thunderstorm Environments in Southern Brazil: Analysis of Severe Weather Parameters. *Proceedings of the 22nd Conference on Severe Local Storms*. American Meteorological Society. Retrieved June 26, 2021, from file:///C:/Users/Windows%2010/Downloads/81745.pdf

Nascimento, E. L. (2005). Previsão de tempestades severas utilizando-se parâmetros convectivos e modelos de mesoescala: uma estratégia operacional adotável no Brasil? *Revista Brasileira De Meteorologia, 20*(1), 121–140.

Paulucci, T. B., França, G. B., Libonati, R., & Ramos, A. M. (2019). Long-term spatial–temporal characterization of cloud-to-ground lightning in the metropolitan region of Rio de Janeiro. *Pure and Applied Geophysics*. https://doi.org/10.1007/s00024-019-02216-1

Pinto, O. J., Naccarato, K. P., Pinto, I. R. C. A., Fernandes, W. A., & Pinto, O. N. (2006). Monthly distribution of cloud-to-ground lightning flashes as observed by lightning location systems. *Geophysical Research Letters*. https://doi.org/10.1029/2006GL026081

Quinlan, J. R. (1993). *C4.5. Programs for machine learning*. Morgan Kaufmann.

Sasaki, Y. (2007). The truth of the F-measure. Retrieved August 15, 2020, from https://www.cs.odu.edu/~mukka/cs795sum09dm/Lecturenotes/Day3/F-measure-YS-26Oct07.pdf

Showalter, A. K. (1953). A stability index for forecasting thunderstorms. *Bulletin of the American Meteorological Society*. https://doi.org/10.1175/1520-0477-34.6.250

Silva, W. L., Albuquerque, F. A. N., França, G. B., & Matschinske, M. (2016). Conceptual model for runway change procedure in Guarulhos international airport based on SODAR data. *The Aeronautical Journal*. https://doi.org/10.1017/aer.2016.33

Sumner, M., Frank, E., & Hall, M. (2005). *Speeding up logistic model tree induction*. Springer. https://doi.org/10.1007/11564126_72

Wilk, K. E. & Gray, K. C. (1970). Processing and analysis techniques used with the NSSL weather radar system, *Preprints, 14th Conf. on Radar Meteorology* (pp. 369–374). American Meteorological Society.

Wilks, S. D. (2006). *Statistical methods in the atmospheric sciences*. Academic Press.

Wilson, J. W. (1966). *Movement and predictability of radar echoes*. National Severe Storms Laboratory.

Witten, I., Frank, E., Hall, M. A. & Pal, C. J. (2016). Data mining: practical machine learning tools and techniques.

WMO (2013, March). *The GCOS Reference Upper-Air Network* (*GRUAN*) *GUIDE*. Retrieved July 7, 2021, from https://library.wmo.int/doc_num.php?explnum_id=7196

(Received January 11, 2021, revised September 29, 2021, accepted October 18, 2021, Published online November 6, 2021)