Use of Convective Indices to Improve the Prediction of Departure Delays

Juan Nunez-Portillo, Alfonso Valenzuela, Antonio Franco, Damián Rivas

Department of Aerospace Engineering
Universidad de Sevilla
41092 Seville, Spain
avalenzuela@us.es

***Abstract*— Severe convective weather disrupts European aviation, causing flight deviations and delays. This study addresses the challenge of improving long-term flight predictability, beyond two hours, focussing on departure delays. It explores the potential of convective indices, derived from atmospheric data, as proxies for departure delays. Despite limitations, these indices are appealing due to their simplicity and widespread availability in medium-range weather forecasts. The research collects historical flight data from Europe and correlates departure delays with convective indices. Deterministic and probabilistic prediction models are developed, evaluating their performance against baseline flight plan predictions. The results reveal that using convective indices significantly enhances the prediction of departure delay, particularly in probabilistic models. Lifted, Boyden, and Bradbury indices show promise. Future work includes multi-index predictors, airport-specific indices, machine learning techniques, and the extension of this approach to other flight deviations.**

Keywords- departure delay; thunderstorms; convective indices; probabilistic modelling

I. Introduction

Severe convective weather, mainly during summer in Europe [1], is one of the most disruptive events for aviation. With a possible extension of hundreds of kilometres, a duration of up to several hours, and a dynamic behaviour [2], its occurrence is associated with damaging phenomena such as turbulence, icing, lightning, hail, and strong winds. Under these circumstances, flights are subject to four-dimensional deviations with respect to their plans. Additionally, air traffic controllers experience an increase in their workload because, in a scenario of high demand due to the summer season, the traffic flow becomes irregular and difficult to anticipate, radio-communications become more frequent, and less airspace volume is available for conflict resolution.

The magnitude of the disruption due to convective weather (e.g., only in Karlsruhe UAC convective weather caused nearly 650k delay minutes between May 2022 and July 2022 [3]) justifies the need for improving the flight predictability. In particular, this study focusses on the prediction of convection-induced deviations in the long term, beyond two hours. Unlike short-term trajectory prediction (up to two hours), based on nowcasts that use radar echoes and satellite data to provide very detailed information about storm cells, long-term trajectory prediction relies on numerical weather forecast, whose spatiotemporal resolution is too coarse if compared with the size and lifespan of convective storms. Thus, the forecast uncertainty to say when a storm initiates, its potential severity, motion, and duration, translates into the aircraft trajectory prediction.

To improve the predictability in the long term, this paper examines the use of convective indices as potential proxies for deviations. In particular, we focus on delays at departure, leaving the door open to include other four-dimensional deviations along flights in further studies.

Based on thermodynamic and kinematic variables of the atmosphere, convective indices have historically been investigated to predict the occurrence and severity of thunderstorms [4]. As environmental proxies, they are irremediably burdened with inaccuracy, and their applicability is usually limited to some geographical region and season [5]. Furthermore, the forecast of a convective area does not imply that the whole area has to be circumnavigated but that storm cells may exist within it. However, the simplicity and forecasting power of these indices, along with the fact that they are the only information available in medium-range forecasts (i.e., several hours in advance, up to several days), justify their consideration for the purpose of this work.

This paper is structured as follows. Section II reviews related work, provides background information on thunderstorms, and contains the selection of indices. Section III describes the methodology. Section IV defines the case study. Section V presents the results and the analysis. Finally, Section VI provides a summary of the findings and a proposal for future research.

II. Background

A. Related Work

The study of correlations between lateral deviations and storm metrics that could be exploited in the short term has been widely reported in the United States, mostly within the framework of the development of the Convective Weather Avoidance Model or CWAM model ([6] and subsequent works). These metrics include lightning strikes and radar-based metrics such as vertically integrated liquid (VIL) and echo tops, spatially filtered over different extensions (from the radar product resolution to 60x60 km). The most popular approach consists in analysing whether or not individual historical flights deviate under specific weather metric values encountered along their tracks. With this information, pattern classification algorithms

SESAR Innovation Days
27-30 November 2023
Seville

are implemented to build deterministic deviation prediction models, as is the case of the CWAM model.

In contrast to this flight-centric approach, there are others that, putting the storm at the centre, study its effect on large-scale traffic, for example, in terms of airspace occupancy [7]. More importantly, they seem to corroborate the findings of the first approach on the en-route airspace. In summary, the difference between the flight altitude and the echo top proved to be the best indicator of pilots' avoidance strategy in the en-route airspace [6], even though such strategy cannot be accurately predicted based only on metrics [8][9]. As for the terminal airspace, where the problem appears to be more complex due to all the constraints inherent to this environment [10], the precipitation intensity (i.e. the VIL) is the most explanatory powerful feature [11][12].

Despite the promising relationships found in the short-term context, when the prediction skill is tested with forecasted weather information instead of actual weather information, it is found that the performance severely depends on the forecasting goodness [11][12][13].

Whereas the previously mentioned works tried to find relationships between convective weather and lateral deviations, at the same time efforts were put into finding relationships between convective weather and delays. The works presented in [14][15], and those that followed them, defined indices for the impact of weather, based on the number of flights close to radar-detected storms or lightning strikes, and found significant correlations between these indices and delays. One limitation of these works is that delays are considered at national or regional level, but not for individual flights.

More recently, researchers have tried to predict the delay of individual flights using data science. A complete and thorough review of recent efforts can be found in [16]. However, these studies usually focus on departure or arrival delays, only a fraction of them explicitly include weather information, and this information is usually limited to atmospheric conditions at the airports. One example is [17], where machine learning algorithms are applied.

Also of interest is the series of works that develop a methodology to evaluate the performance of arrival operations through multiple key performance indicators such as arrival punctuality, additional time in the terminal airspace, extra distance, fuel consumption and level-off during descent [18][19][20][21][22]. By using regression techniques, the performance is assessed against impact factors that include the adverse meteorology in the corresponding airport. To be precise, the weather factor is partially calculated with different convective indices. Nevertheless, the extrapolation of the methodology to the en-route problem is not clear; for instance, where and when the weather metrics would be calculated, how to isolate the weather impact from the traffic impact, or how to relate the performance indicators to individual weather encounters along the route (because, a priori, one encounter with a severe storm could cause a delay similar to the one provoked by several encounters with different weaker storms).

Some convective indices have already been used for optimal trajectory planning [23][24] and trajectory prediction [25] in the long term. In these works, a combination of two ensemble-forecasted convective indices, one proxy for convection potential (in particular, the total totals index or the lifted index, depending on the data availability) and another one for convection initiation (i.e., the convective precipitation), is used to build an index that estimates the probability of convection. However, the ability of the convective indices to predict aircraft deviations is taken for granted, hence the need to prove their correlations not with actual storms but with historical flight deviations (since flights tend to deviate only under specific circumstances).

In view of all the above, it is a challenge to study the relationship between convective indices and the four-dimensional deviations one aircraft may experience along its entire route. This paper presents a methodology focused on departure delays but applicable to any 4D deviations at any flight phase under slight modifications.

B. *Thunderstorms and convective indices*

As a consequence of moist convection (i.e., the ascension of a wet air parcel warmer than its environment), a thunderstorm needs three conditions for its occurrence [26]: (1) sufficient moisture in the boundary layer; (2) a lifting mechanism to trigger updrafts (e.g., orography, low-level converging winds, and uneven surface temperature); and (3) thermodynamic instability (i.e., a strongly decaying temperature profile with height), responsible for the buoyancy that accelerates the air parcel upwards.

Based on their size, structure, and organisation, there are four types of thunderstorms (ordinary cell, multi-cell cluster, multi-cell line, and supercell) [2], all of which go through three phases: developing phase, mature phase, and dissipation phase. During the thunderstorm development, the three conditions mentioned in the previous paragraph lead to an ascending flow of water vapour that, in the presence of cooler air at high altitude, condensates and forms the clouds, with a normal diameter of 5–10 km [27]. The expansion of these clouds continues during the mature phase and determines the intensity of the thunderstorm. For the clouds to grow, they need to be fed with the updraft driven by the instability. However, this ascending channel is interrupted when the water held in the clouds precipitates over the same column. Without the inflow of warm moist air, the thunderstorm rapidly dissipates. A different scenario would be if, in addition to instability, there were enough vertical wind shear, that is, a change in the horizontal wind speed or direction with height. In this case, the clouds are horizontally displaced in such a way that the precipitation occurs over a column different from the one supplying moisture, thereby sustaining the thunderstorm for longer and leading to more severe events [27].

The three necessary (but not sufficient) ingredients for a thunderstorm to develop, and the two main factors that determine its severity correlate to different moments in the thunderstorm lifetime. This is to be remembered when selecting the convective indices and defining the methodology because of two reasons: first, an aircraft will rarely deviate around a thunderstorm in its premature, and therefore less severe, phase; and second, wind shear alone (i.e., without a pre-existing storm) is not a cause for detours.

For this study, only convective indices dealing with stability are considered. As an additional condition for the choice, their prediction skills for thunderstorm occurrence in Europe have been taken into account (e.g., see [4][28][29][30]), and not only their appearance in other works. The reason is that the convective activity over Europe is not the same as in the United States, where the likelihood, intensity, and extension of thunderstorms are usually higher [31]. In Europe, the atmospheric instability and moisture are limited, and the wind shear is weaker. This translates into differences in the performance of indices and their thresholds [4].

The list of selected indices and some guiding thresholds are provided in Table I. Although there are indices with more than one threshold, indicating not only the thunderstorm occurrence but also the degree of severity, note that Table I only provides one per index. In these cases, the threshold indicates severe thunderstorm in Europe.

III. Methodology

This section is divided into three parts. Section III.A gives the instructions to obtain the convective indices and departure delays. Seeking to improve today's prediction ability, section III.B proposes two approaches based on the value of a single index, a deterministic one and a probabilistic one. Lastly, section III.C contains an evaluation framework of the models that will be used later.

A. *Convective indices and departure delays*

To study the relationships between indices and departure delays, as a first step, it is necessary to identify the historical traffic of interest. In this work, such traffic is formed by the flights that, within a considered region and time period, meet the following conditions:

- the flight was not cancelled,
- the departure took place where planned, and
- the departure airport was different from the arrival airport.

TABLE I. SUMMARY OF CONVECTIVE INDICES.

Index	Definition	Comment
Convective Available Potential Energy (CAPE) [32]	$\int_{LFC}^{EL} \frac{T_{v,p} - T_v}{T_v} g dz$	In J/kg, it measures the available energy to lift an air parcel from the level of free convection (LFC) to the equilibrium level (EL). The higher the value, the more unstable the atmosphere is (CAPE > 0 J/kg means instability). Severe events are probable for 500 J/kg [30].
Lifted Index (LI) [33]	$T_{500} - T_{S\rightarrow 500,p}$	In ºC or K, it measures the temperature difference only at 500 hPa (mid-troposphere) between the environment and an air parcel that is ideally lifted from the surface. The more negative the value, the more unstable the atmosphere is (LI<0 ºC means instability). Severe events are probable for LI<–3 ºC [28].
Deep Convective Index (DCI) [34]	$T_{850} + T_{d,850} - LI$	In ºC, it combines the LI with temperature and humidity at 850 hPa (generally above the atmospheric boundary layer, which means independence of surface conditions). The higher the value, the more likely a thunderstorm will occur. Severe events are probable for DCI>30 ºC [28].
Total Totals (TT) [35]	$T_{850} + T_{d,850} - 2T_{500}$	In ºC or K, it is the sum of two components: the vertical temperature gradient and the humidity in the lower-level atmosphere. The higher the value (typically when TT>45 ºC), the more likely a thunderstorm will occur. Severe thunderstorms are possible for TT>50 ºC [36].
K Index (KI) [37]	$T_{850} + T_{d,850} - T_{500} - (T_{700} - T_{d,700})$	In ºC, it is similar to the TT, but adding the effect of humidity at 700 hPa. The higher the value (typically when KI>20 ºC), the more likely a thunderstorm will occur. Severe thunderstorms are likely for KI>30 ºC [38].
Boyden Index (BoydI) [39]	$0.1(z_{700} - z_{1000}) - T_{700} - 200$	With each magnitude in its unit, it calculates an amount proportional to the mean vertical temperature gradient between 1000 hPa and 700 hPa. The higher the value, the more likely a thunderstorm will occur (tied to a frontal passage). Increase in thunderstorm activity for BoydI>95 [38].
Bradbury Index (BradI) [40]	$\theta_{w,500} - \theta_{w,850}$	In ºC or K, it quantifies the instability of the 850–500 hPa layer. This means that the lower the difference, the stronger the instability, and the more likely the formation of thunderstorms will be. Increase in thunderstorm activity in summertime for BradI<–2 ºC [4].
Rackliff Index (RackI) [41]	$\theta_{w,900} - T_{500}$	In ºC or K, it measures the instability of the 900–500 hPa layer. The higher the value, the stronger the instability and the more likely the formation of thunderstorms will be. Increase in thunderstorm activity for RackI>30 ºC [4].
Modified Jefferson Index (JeffI) [42]	$1.6\theta_{w,850} - T_{500} - 0.5(T_{700} - T_{d,700}) - 8$	In ºC, it amends the RackI by using a standard altitude (850 hPa instead of 900 hPa), and by adding the contribution of humidity at 700 hPa (which corrects the overforecast of thunderstorms over dry areas). Increase in thunderstorm activity for JeffI>28 ºC [28].

Notation: T is the temperature in ºC; T_v is the virtual temperature in ºC; T_d is the dewpoint temperature in ºC; θ_w is the wet-bulb potential temperature in ºC; z is the geopotential height in m; and g is the gravity in m/s^2.

Subscripts: While the presence of p indicates that the magnitude corresponds to an air parcel, its absence indicates that the magnitude corresponds to the environment; a number (e.g. 500) indicates a specific pressure level in hPa; S indicates surface level; and $i \rightarrow j$ indicates that an air parcel is lifted dry-adiabatically from i-level to its condensation level, and then moist-adiabatically up to j-level if this is beyond the condensation level.

For each flight of the defined dataset, the actual take-off time (ATOT) and the estimated take-off time (ETOT) are extracted, and their difference is calculated to get the corresponding actual departure delay, $ad = ATOT - ETOT$. Furthermore, each flight is assigned one index value per considered convective index; this is computed at a certain location (where the aircraft took off) and at a certain time (at the ETOT). Hence, one ends up with a sample of departure delays and associated values of convective indices.

B. Construction of the prediction models

The main hypothesis of this work is that an improvement in the prediction of delays at take-off can be achieved by considering convective indices. Therefore, innovative prediction models can be developed considering such indices. Note that, for the sake of understanding, simplicity is preferred to sophistication in this work, so one-index models are chosen.

First, a deterministic approach is addressed. In other words, a model that provides a single prediction of the departure delay for an input value of a convective index. In this model, the sample of departure delays is disaggregated by the convective index value into several sub-samples, according to a set of intervals whose definition is given below. Then, the prediction of the departure delay is given by the median of each sub-sample. Hence, a different prediction is defined for each convective index interval, in the manner of a look-up table.

To define the set of intervals for each convective index, the following rules have been considered:

1) The boundaries of the intervals take into account the order of magnitude and, more importantly, the different thresholds of each index. See Table II.

TABLE II. THRESHOLDS USED IN THIS WORK.

Index	**Values [Ref.]**	**Index**	**Values [Ref.]**
CAPE [J/kg]	250, 500 [30]	BoydI [-]	95 [38]
LI [ºC]	−6, −3, 0 [36]	BradI [ºC]	−2 [4]
DCI [ºC]	30 [28]	RackI [ºC]	30 [4]
TT [ºC]	45, 50, 55 [36]	JeffI [ºC]	28 [28]
KI [ºC]	20, 25, 30 [38]		

2) If there is only one threshold, at least two intervals are defined on both sides.

3) If there is more than one threshold, the width between indices is used as a reference for both inner (between thresholds) and outer intervals. The number of outer intervals is at least one on each side.

4) The first and last outer intervals coincide with the first and last outer intervals that, having the reference width in 3), can be populated with a sufficiently large sample of data. After their identification, they are extended towards plus/minus infinity to include the remaining data.

5) In case an interval has an insufficient amount of data, it is merged with either the previous one or the following one, depending on the location of thresholds.

Afterwards, due to the stochasticity present in the atmosphere and in aircraft operations, a probabilistic prediction model is developed. In this model, the same intervals that have been defined for the deterministic model are considered. The difference with the deterministic model is that, instead of using the median, the complete empirical probability distribution is used as a predictor.

Although an ensemble approach with a finite number of members could be desirable, when comparing the predictive skill of probabilistic models by means of scoring functions (such as the Continuous Ranked Probability Score, CRPS), the discretisation introduced by the ensemble leads to a bias [43]. Therefore, in this work, the complete empirical probability distribution of the past departure delays is used as predictor and the creation of an ensemble predictor that minimises this bias is left for future work.

C. Evaluation framework

The total amount of data is divided into sets, a training set and a test set. The training set gathers two-thirds of the total (i.e., of the traffic days) and is used to define the models, whereas the test set gathers the remaining one-third of the total and serves to evaluate their performance. Because the split considers entire traffic days, information leakage from the training set into the test set is unlikely. In other words, it is unlikely that a flight delay is predicted by a model trained with flights affected by the same weather event.

In this paper, two studies are carried out. First, the prediction skills of three deterministic predictors are compared: the flight plan, a deterministic prediction model without disaggregation by index value, and a deterministic prediction model based on the value of a single convective index. Using the flight plan as a predictor constitutes the baseline procedure and implies predicting no delay. As for the deterministic model without considering the convective index, it predicts the departure delay as the median of the whole sample of departure delays, disregarding the convective index value of the flight considered.

To quantify the performance of each model, the selected property to be assessed is the median-unbiasedness; hence, the selected score is the median value of the prediction error:

$$MedE = \boldsymbol{median}(\{pd_i - ad_i\}), \tag{1}$$

where pd_i is the deterministically predicted departure delay for flight i, and ad_i is the actual departure delay for flight i.

The second study consists in comparing the prediction skill of the probabilistic model based on the value of a convective index to the deterministic one. Now, to quantify the performance of each model, two skill scores are used: the mean absolute error (MAE) for deterministic models and the mean CRPS (MCRPS) for probabilistic models. In symbols:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|pd_i - ad_i\right|, \quad (2)$$

$$MCRPS = \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{+\infty}[F_i(y) - H(y \geq ad_i)]^2 dy, \quad (3)$$

where N is the number of flights in the test set, $F_i(y)$ is the empirical cumulative distribution function for flight i according to its corresponding index value and evaluated at the delay y, and $H(\cdot)$ is the Heaviside function. Note that when the sample space of $F_i(y)$ is a single value, the CRPS coincides with the absolute error.

IV. Case Study

The selected case study corresponds to the departures over Europe in June and September 2019, months with high traffic levels and affected by thunderstorms [44]. More precisely, the area of interest is the intersection between the area of the European Civil Aviation Conference (ECAC) Member States and the area from 10ºW to 30ºE and 35ºN to 70ºN.

As for the historical traffic data, around 1.55 million flights are considered (see Figure 1). Their corresponding ETOTs and ATOTs are obtained from their planned and actual trajectories, respectively, stored in the Eurocontrol's R&D Data Archive [45]. Notice that the flight plan information provided by this archive corresponds to the last filed flight plan; therefore, the delay considered in this analysis correspond to delays that took place between this last filing and the actual take-off.

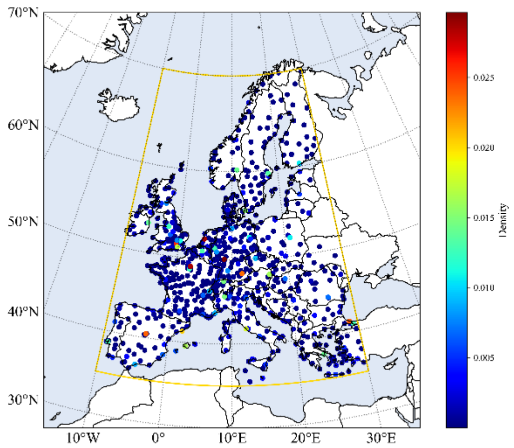

Figure 1. Density plot of the considered departures.

As for the historical weather data, the convective indices at a certain time instant and location are calculated by interpolating the hourly 3D weather information available at the ECMWF Reanalysis v5 [46]. In particular, the CAPE, TT, and KI are directly obtained from the weather dataset whereas the rest of indices are computed from the variables that appear in their definitions.

V. Results

The results of the case study are presented and discussed in the following three subsections. First, the analysis of the inputs is carried out in section V.A. Secondly, the results of the deterministic approach are presented in section V.B. Lastly, the results of the probabilistic approach are presented in section V.C.

A. *Analysis of the inputs*

Firstly, for each index, the histograms of the training and test sets are inspected, looking rather similar. This was expected because the division of the sets was made at random. However, there are differences in the shape of the curves among indices. The curves for the TT, KI, RackI and JeffI are left-skewed, whereas the curves for the BoydI and DCI are bell-shaped, and the curves for the CAPE, LI and BradI are right-skewed. Another observation is that the peak of the curves falls on the non-thunderstorm occurrence side in most cases (except for the TT and BoydI). See the example shown in Figure 2 for the LI.

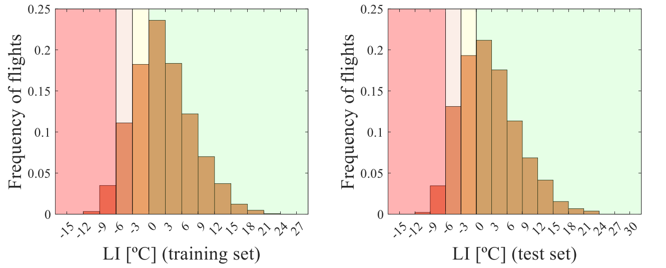

Figure 2. Histograms of the training (left) and test (right) sets for the LI.

Secondly, Table III shows, for each convective index, the percentage of flights that are under possible thunderstorm occurrence. For those indices that can determine the severity of the storm, severe events are considered. For clarity, the threshold considered for each index is included in the table. Results are shown for both the training set and the test set. On the one hand, it can be observed that, for each index, the figures are similar in both sets, training and test. On the other hand, the figures among indices are dissimilar, ranging approximately from 5% (DCI, BradI, and TT) to 45% (BoydI).

Table III. Flights under possible thunderstorm occurrence.

Index	Training set	Test set
CAPE (>500 J/kg)	10.9%	11.0%
LI (<−3 ºC)	15.0%	16.8%
DCI (>30 ºC)	5.9%	4.6%
TT (>50 ºC)	6.9%	8.4%
KI (>30 ºC)	8.6%	9.4%
BoydI (>95)	45.8%	48.0%
BradI (<−2 ºC)	6.5%	6.4%
RackI (>30 ºC)	26.3%	31.2%
JeffI (>28 ºC)	22.2%	25.6%

Thirdly, the median delay of the flight plan in the training set is 3.37 minutes; this positive value means that the flights tend to depart later than planned. If arranged by index intervals, one can analyse the behaviour of the flight plan for different atmospheric situations. For example, see Figure 3, which shows the flight plan's median delay as a function of the LI, which grows towards the side of likely (severe) thunderstorm occurrence, from 1 min (stability side) to 8 min (instability side).

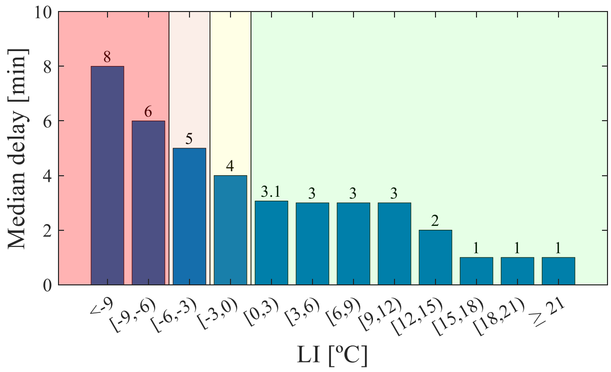

Figure 3. Flight plan's median delay as a function of LI (training set).

Finally, the whole distribution of the flight plan's delay is shown in Figure 4 as a function of the LI. The box represents 25th, 50th, and 75th percentiles, and the whiskers represent 5th and 95th percentiles. It can be seen that not only the median but also the dispersion grows towards the side of likely (severe) thunderstorm occurrence. Measured as the difference between the 95th and 5th percentiles, it goes from 33 min (stability side) to 47 min (instability side).

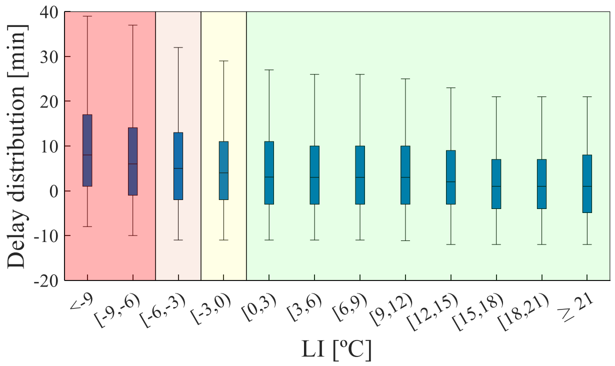

Figure 4. Flight plan's delay distribution as a function of LI (training set).

B. *Deterministic approach*

First, the median deviation is checked for the deterministic predictions of the flight plan (abbreviated as FP), the deterministic predictor without disaggregation by convective index value (abbreviated as DP_0), and deterministic predictors based on convective index values (abbreviated as DP_CI).

The prediction values of each predictor are as follows: the FP prediction is 0 minutes of delay; the DP_0 prediction is the median delay of the flight plan in the training set, i.e., 3.37 minutes; the DP_CI prediction is the median delay of the flight plan in the training set for each one of the convective index intervals. For example, for the LI, this corresponds to the values presented in Figure 3, ranging from 1 min to 8 min.

In the test set, the results are as follows. The median value of the prediction errors is 4.00 minutes for the FP, the DP_0 error is 0.63 minutes, and the DP_CI errors are given in Table IV for each convective index. Notice that the results of the DP_CI can be provided for each convective index interval; however, for the sake of comparison, results are separated just into two categories: whether the index states that the occurrence of a (severe) thunderstorm is possible or not (using the same thresholds presented in Table III). For comparison, the median errors of FP and DP_0 in these two categories (occurrence/non-occurrence) are also given in the table for each index.

The following results can be highlighted from Table IV:

- For all indices, the FP errors are larger when thunderstorms are possible.
- In all cases, the DP_0 errors are smaller than the corresponding FP errors, and these errors are larger when thunderstorms are possible.
- All DP_CI predictors outperform the FP, and the differences are larger when thunderstorms are possible.
- Most of the DP_CI predictors outperform the DP_0 predictor, except for CAPE and RackI, and the improvements are larger when thunderstorms are possible, especially for LI, BoydI and BradI.
- The DP_CI predictors based on LI, BoydI, and BradI completely cancel the median bias in both cases, when thunderstorms are possible and when they are not.

TABLE IV. MEDIAN ERRORS OF THE DETERMINISTIC PREDICTIONS (IN MINUTES).

Index	**Possible thunderstorm occurrence**			**Non-occurrence**		
	FP	***DP_0***	***DP_CI***	***FP***	***DP_0***	***DP_CI***
CAPE	5	1.633	0.075	3	−0.367	0.583
LI	5	1.633	0	3	−0.367	0
DCI	5	1.633	−1	3	−0.367	0
TT	5	1.633	1	3	−0.367	0
KI	5	1.633	−0.217	3	−0.367	0
BoydI	4	0.633	0	3	-0.367	0
BradI	5	1.633	0	3	−0.367	0
RackI	4	0.633	0.767	3	−0.367	−0.233
JeffI	4	0.633	0.133	3	−0.367	−0.300

C. *Probabilistic approach*

After the deterministic models, the probabilistic predictors based on convective index values (abbreviated as PP_CI) have been developed. Recall that, for each convective index interval, the complete empirical probability distribution is used as a predictor. For instance, for the LI, this corresponds to the distributions presented in Figure 4. As a reference, for the interval [−6, −3) ºC, the 5th, 25th, 50th, 75th, and 95th percentiles are −11, −2, 5, 13, and 32 min, respectively.

Then, the prediction skill of each model has been tested by the CRPS. Figure 5 shows the case of the LI again. In this case, the mean CRPS steadily increases from 3.6 min (stability side) to 7.9 min (instability side). Therefore, the probability distribution is harder to capture when storms are possible. A similar pattern is found for the rest of indices.

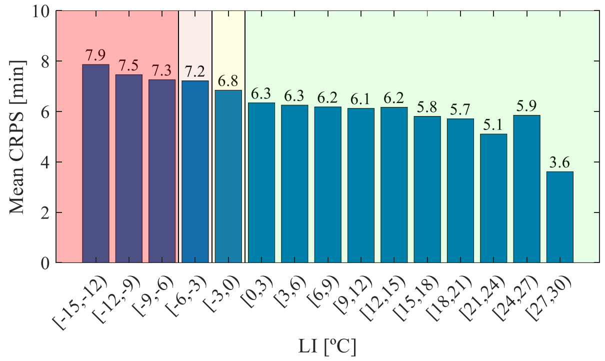

Figure 5. Performance of the probabilistic model for the LI (test set).

To compare the prediction skill of the probabilistic model to the deterministic one, the mean CRPS (MCRPS) is checked against the mean absolute error (MAE). As in the deterministic analysis, this comparison is made for each convective index and two categories: whether the index states that the occurrence of a (severe) thunderstorm is possible or not. The results are presented in Table V. It can be highlighted that:

- For all indices and both predictors, the errors are larger when thunderstorms are possible.
- All PP_CI predictors outperform the corresponding DP_CI predictors.
- The PP_CI predictor based on BoydI performs better than the other predictors, although by a small margin.

VI. Conclusions and Future Work

The general framework of this research is to improve the flight predictability several hours in advance, when the only available information about the occurrence of storms is convective indices. In this paper, the relationships between departure delays and several convective indices have been explored by way of a historical dataset over Europe. The inspection of the historical data has allowed us to observe and quantify the evolution of the departure delays as the likelihood of (severe) thunderstorm occurrence increases: they grow and become more dispersed.

Several simple prediction models have been developed and assessed, based on different convective indices and with deterministic and probabilistic approaches. It has been shown that it is possible to correct the median delay of the flight plans, and that the probabilistic approach significantly reduces the error of the prediction. In particular, Lifted, Boyden, and Bradbury indices look like the most promising ones.

Table V. Comparison of the prediction skill of the probabilistic predictors versus the deterministic predictors (in minutes).

Index	Possible thunderstorm occurrence		Non-occurrence	
	DP_CI (MAE)	***PP_CI (MCRPS)***	***DP_CI (MAE)***	***PP_CI (MCRPS)***
CAPE	9.999	7.376	8.789	6.445
LI	9.830	7.230	8.700	6.370
DCI	9.973	7.357	8.846	6.459
TT	9.701	7.173	8.846	6.484
KI	9.962	7.350	8.812	6.463
BoydI	9.300	6.808	8.544	6.234
BradI	9.505	7.013	8.870	6.487
RackI	9.480	6.960	8.691	6.360
JeffI	9.474	6.985	8.758	6.405

As future work, it is of interest the development of predictors that make use of several convective indices simultaneously, the identification of the better indices for each airport, and the refinement of the index thresholds to exploit the full potential of this weather information. All these advances can be achieved, for example, by applying machine learning techniques.

It is worth noting that the flight plans can be modified by the aircraft operators before departure because of, among other reasons, the weather at the airport. Because of these modifications, the errors and dispersions grow as a function of how far in advance the prediction is made. In this work, we have used the only openly accessible information, which is the last filed flight plan, but it would be very interesting to analyse the prediction as a function of time if the different flight plan versions are available.

Finally, the findings of this piece of research encourage us to extend this approach to other flight deviations. The immediate next step is to determine the probability of deviation to an alternative airport depending on the values of the convective indices at the destination airport. And then, to determine the spatial and temporal deviations along the route as a function of the values of the convective indices encountered along the flight.

Acknowledgment

The authors gratefully acknowledge the financial support of the Spanish Ministerio de Ciencia e Innovación through Grant PID2021-122323OB-C32.

References

[1] M. Taszarek, J. Allen, T. Púčik, K. Hoogewind, and H. Brooks, "Severe convective storms across Europe and the United States. Part II: ERA5 environments associated with lightning, large hail, severe wind, and tornadoes", *J. Clim.*, vol. 33, no. 23, pp. 10263–10286, 2020.

[2] National Oceanic and Atmospheric Administration (NOAA), "Severe weather 101: thunderstorm types," (accessed July 29, 2023), https://www.nssl.noaa.gov/education/svrwx101/thunderstorms/types/.

[3] EUROCONTROL, "Performance review report: an assessment of air traffic management in Europe during the calendar year 2022," June 2023, p. 4.

[4] R. Peppier, "A review of static stability indices and related thermodynamic parameters," ISWS Miscellaneous Publication MP-104, Oct. 1988, pp. 1, 20, 63–68.

[5] M. Tippett, C. Lepore, W. Koshak, T. Crhonis, and B. Vant-Hull, "Performance of a simple proxy for U.S. cloud-to-ground lightning," Oct. 2018, pp. 9–18.

[6] R. DeLaura and J. Evans, "An explanatory study of modeling enroute pilot convective storm flight deviation behavior," in *12th ARAM*, Atlanda, 2006.

[7] K. Kuhn, "Analysis of thunderstorm effects on aggregated aircraft trajectories," *J. aerosp. comput. inf. commun.*, vol. 5, no. 4, pp. 108–119, 2008.

[8] M. Matthews and R. DeLaura, "Assessment and interpretation of en route weather avoidance fields from the convective weather avoidance model," in *AIAA ATIO*, Fort Worth, 2010.

[9] S. Wu, C. Gooding, A. Shelley, C. Duong, and W. Johnson, "Pilot convective weather decision making in en route airspace," in *28th ICAS*, Brisbane, 2012.

[10] D. Rhoda, E. Kocab, and M. Pawlak, "Aircraft encounters with thunderstorms in enroute vs. terminal airspace above Memphis, Tennesse," in *13th Conf. on Appl. Climatol., and 10th ARAM*, Portland, 2002.

[11] S. Campbell and R. DeLaura, "Convective weather avoidance modeling in low-altitude airspace," in *AIAA MST*, Portland, 2011.

[12] M. Matthews and R. DeLaura, "Modeling convective weather avoidance of arrivals in the terminal airspace," in *ARAM Special Symp. on Weather – ATM Integration*, Washington, DC, 2011.

[13] M. Matthews and R. DeLaura, "Evaluation of enroute convective weather avoidance models based on planned and observed flight," in *14th ARAM*, Atlanta, 2010.

[14] M. Callaham et al., "Assessing NAS performance: normalizing for the effects of weather," in *4th USA/Europe ATM R&D Symp.*, Santa Fe, 2001.

[15] J. Post, J. Bonn, M. Bennett, D. Howell, and D. Knorr, "The use of lfight track and convective weather densities for natural airspace system efficiency analysis," in *21st DASC*, Irvine, 2002.

[16] L. Carvahlo et al., "On the relevance of data science for flight delay research: a systematic review," *Transp. Rev.*, vol. 41, no. 4, pp. 499–528, 2021.

[17] S. Choi, Y. Kim, S. Briceno, and D. Mavris, "Prediction of weather-induced airline delays based on machine learning algorithms," in *35th DASC*, Sacramento, 2016.

[18] A. Lemetti, T. Polishchuk, R. Sáez, and X. Prats, "Evaluation of flight efficiency for Stockholm Arlanda Airport," in *38th DASC*, San Diego, 2019.

[19] A. Lemetti, T. Polishchuk, V. Polischuk, R. Sáez, and X. Prats, "Identification of significant impact factors on arrival flight efficiency within TMA," in *9th ICRAT*, virtual conference, 2020.

[20] A. Lemetti, T. Polishchuk, and H. Hardell, "Arrival flight efficiency in numbers: what new the Covid-19 crisis is bringing to the picture?," in 10th *SIDs*, virtual conference, 2020.

[21] H. Hardell, A. Lemetti, and T. Polischuk, "Performance evaluation of the arrival operations in terminal area," in *33rd ICRAT*, Stockholm, 2022.

[22] A. Lemetti, , H. Hardell, and T. Polishchuk, "Arrival flight efficiency in pre- and post-Covid 19 pandemics," *J. Air Transp. Manag.*, vol. 107, no. 102327, 2023.

[23] D. González-Arribas et al., "Robust optimal trajectory planning under uncertain winds and convective risk," in *Air Traffic Management and Systems III* (E. N. R. Institute, ed.), (Singapore), pp. 82–103, Springer Singapore, 2019

[24] M. Soler et al., "Influence of atmospheric uncertainty, convective indicators, and cost-index on the leveled aircraft trajectory optimization problem," *Transp. Res. Part C Emerg. Technol.*, vol. 120, no. 102784, 2020.

[25] E. Andrés et al., "Probabilistic analysis of air traffic in adverse weather scenarios," in *10th ICRAT*, Tampa, 2022.

[26] C. Doswell, H. Brooks, and R. Maddox, "Flash flood forecasting: an ingredients-based methodology," *Weather and Forecasting*, vol. 11, no. 4, pp. 560–581, 1996.

[27] AU Bureau of Meteorology, "Hazardous weather phenomena: thunderstorms," (accessed Aug. 5, 2023), http://www.bom.gov.au/aviation/data/education/thunderstorms.pdf

[28] M. Kunz, "The skill of convective parameters and indices to predict isolated and severe thunderstorms," *Nat. Hazards Earth Syst. Sci.*, vol. 7, no. 2, 2007.

[29] M. Schmeits, K. Kok, and D. Vogelezang, "Probabilistic forecasting of (severe) thunderstorms in the Netherlands using model output statistics," *WAF*, vol. 20, pp. 134–148, 2005.

[30] R. Kaltenbök, G. Diendorfer, and N. Dotzek, "Evaluation of thunderstorm indices from ECMWF analysis, lighning data and severe storm reports," *Atmos. Res.*, vol. 43, pp. 381–396, 2009.

[31] H. Brooks, "Proximity sounding for severe convection for Europe and the United States from reanalysis data," *Atmos. Res.*, vol. 93, nos. 1–3, pp. 546–553, 2009.

[32] M. Moncrieff and M. Miller, "The dynamics and simulation of tropical cumulonimbus and squall lines," *Quart. J. Roy. Meteorol. Soc.*, vol. 102, pp. 373–394, 1976.

[33] J. Galway, "The lifted index as a predictor of latent instability," *Bull. Amer. Meteorol.*, vol. 37, pp. 528–529, 1956.

[34] W. Barlow, "A new index for the prediction of deep convection," in *17th Conf. on Severe Local Storms*, St. Louis, pp. 129–132, 1993.

[35] R. Miller, "Notes on analysis and severe storm forecasting procedures of the Air Force Global Weather Central," tech. rep., May, 1972.

[36] J. Grieser, "Convection parameters," (accessed Aug. 30, 2023), http://www.juergen-grieser.de/CovectionParameters/ConvectionParameters.pdf

[37] J. George, "Weather forecasting for aeronautics," Academic Press, New York, 1960.

[38] EUMETRAIN, "Stability indices," (accessed Aug. 30, 2023) https://resources.eumetrain.org/data/1/15/stability.htm.

[39] C. Boyden, "A simple instability index for use as a synoptic parameter," *Meteorol. Mag.*, vol. 92, pp. 198–210, 1963.

[40] T. Bradbury, "The use of wet-bulb potential temperature charts," *Meteorol. Mag.*, vol. 106, pp. 233–251, 1977.

[41] P. Rackliff, "Application of an instability index to regional forecasting," *Meteorol. Mag.*, vol. 91, pp. 113–120, 1962.

[42] G. Jefferson, "A further development of the stability index," *Meteorol. Mag.*, vol. 92, pp. 313–316, 1963.

[43] M. Zamo and P. Naveau, "Estimation of the continuous ranked probability score with limited information and applications to ensemble weather forecasts," *Math. Geosci.*, vol. 50, pp. 209–234, 2018.

[44] C. Walker, "CODA Digest 2019," tech. report, EUROCONTROL, April, 2020.

[45] EUROCONTROL, "Aviation data repository for research," (accessed Aug. 30, 2023) https://www.eurocontrol.int/dashboard/rnd-data-archive

[46] ECMWF, "ECMWF reanalysis v5 (ERA5)," (accessed Aug. 30, 2023) https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5