Investigating Artificial Neural Networks for Detecting Aircraft Wake Vortices in Lidar Measurements

NIKLAS WARTHA,* ANTON STEPHAN, GRIGORY ROTSHTEYN AND FRANK HOLZÄPFEL

Institut für Physik der Atmosphäre, Deutsches Zentrum für Luft- und Raumfahrt, 82234 Oberpfaffenhofen, Germany
Corresponding author (Niklas.Wartha@dlr.de)

Abstract: Convolutional Neural Networks (CNNs) are employed to identify wake vortices via their two-dimensional position and circulation strength in Light Detection and Ranging (lidar) measurement scans. A campaign at Vienna International Airport delivered data that so far has only been processed with a traditional lidar processing algorithm, namely the Radial Velocity (RV) method. Its not fully automated nature led to only a fraction of scans from the overall data set to be evaluated. Here we present ways to use CNNs for this task. A scoring algorithm engineered for verifying CNN detections has been implemented. In particular green detections (those marked as correct CNN detections by the scoring algorithm) can confidently be used for further analysis about the wake vortex encounter hazard. With this approach we end up with a significantly more processed and characterized lidar data compared to that so far delivered by the RV method.

1. Introduction

Landing aircraft are perennially prone to wake vortex encounters (WVEs) [1]. These air whirls generated by the preceding aircraft are a direct consequence of lift generation. In high-lift configuration, the most prominent wake vortices merge from trailing and wing tip vortices [2] on either side of the aircraft, generating a counter-rotating wake vortex pair. The intrusion path through the wake vortex pair dictates the impact on the aircraft, whereas the severity is determined by the strength of the vortices (typically described by their circulation value), the weight of the involved aircraft, and the aircraft's wingspan with respect to the separation of the vortices [3]. The danger of WVEs during landing becomes apparent when considering that the majority of aircraft share the same glide path towards the runway. In union with ground proximity as well as low flight speed, a pilot has limited reaction capabilities in this scenario [4, 5]. Alertness of WVEs led to the ICAO introducing first landing aircraft separations in the 70s, which nowadays are often considered exceedingly conservative, limiting airport capacity [6]. Although the COVID-19 pandemic led to a temporary decline in flight demand, Eurocontrol expects that by 2025 pre-pandemic flight demand will return [7]. Thus, by utilizing Eurocontrol predictions from 2018 we can expect roughly 1.5 million unaccommodated flights by 2045 [8] (includes a 5 year delay). Airports therefore require means to increase their capacity. One efficient approach to do so is by reducing the currently used aircraft separation standards.

A re-categorization program by Eurocontrol and the FAA, termed RECAT, therefore plans three phases for adapting current aircraft separation standards [9]: the first phase lays out six new aircraft categories taking into account the strength of generated wake vortices as well as the vulnerability of the follower aircraft. Phase two introduces pairwise static separations between individual aircraft types. Lastly, phase three foresees dynamic aircraft separations, which adjust according to prevailing weather conditions. Both phases one and two are already or soon operational at first airports. The last phase relies on real-time knowledge of wake vortices' position and strength on the glide path.

In order to reduce the lifetimes and strengths of wake vortices in ground proximity so-called plate lines have been developed and tested at Vienna International Airport (for more insight see [10]). The requirement for fast-time and automatic Light Detection and Ranging (lidar) measurement processing became evident in first analyses of the Vienna International Airport campaign. We required means to accelerate and automatize the wake vortex characterization of our enormous data set.

Thus, the determination of wake vortices is necessary for both the fast-time evaluation at airports, in order to efficiently separate aircraft during the approach and landing phase, as well as the evaluation of large measurement campaign data such as the one from Vienna International Airport.

In the current effort we analyze the usability of Artificial Neural Networks (ANNs) for characterizing the position and strength of wake vortices, ultimately to reveal methods for improving the characterization process, obtaining higher accuracies and evaluating the performance of the ANN quantifications. Section 2 first details the principles of wake vortex physics in ground proximity and then surveys in additional depth previous investigations relevant for arriving at the current effort. Section 3 gives an overview of the Vienna campaign and outlines the data sets used herein. Section 4 describes the ANN architecture as well as other relevant settings used for characterizing the wake vortices. In Section 5 we present the findings from our studies by analyzing the performance of a Convolutional Neural Network (CNN) in comparison to current lidar processing algorithms, by showcasing the advantages of our method, applying the trained CNNs to data not previously processed with a traditional method, and by suggesting future work.

2. Previous studies on wake vortex detection near the ground

Detecting wake vortices close to the ground implies the necessity to know what they behave like and what occurs to them in that scenario. In ground proximity wake vortices - starting from 1.5 initial vortex separations - are subject to ground effects, causing them to follow hyperbolic diverging trajectories [11]. Once one wingspan above the ground, the wake vortices (in this context also termed primary vortices) induce vorticity of opposite sign at the ground, forming a boundary layer [12]. An adverse pressure gradient grows and ultimately causes secondary vorticies (SV) to detach from the ground. The SV interact with the primary vorticies and cause them to rebound and simultaneously to decay in an accelerated manner [13]. Presence of a crosswind alters the decay of the wake vortices. The crosswind generates an additional boundary layer, which causes an asymmetric wake vortex decay and transport characteristics as shown in Fig. 1 (Ref. [5, 14, 15]). Under weak crosswinds, the upwind vortex may stall over the runway. Under strong crosswinds, the downwind vortex may move towards a closely-spaced parallel runway. Both of these are unfavorable situations.

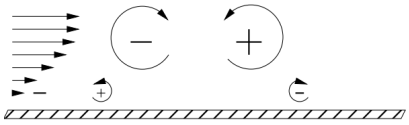


Fig. 1. Vortices in ground proximity with crosswind. Large and small circular arrows represent primary and secondary vortices, respectively (taken from [15], p.1253).

The touchdown of aircraft drastically introduces additional effects, reducing the generated lift and therefore also the shape of the wake vortices with so-called end effects. Pressure variations along the vortex cause axial flow within the core, outside the cores slower effects destabilize the vortex. Obstacles, such as the plate lines may similarly disturb the primary vortices [15]. Plate lines are a series of upright plates (with the dimensions 4.5 m height and 9 m length, as well as the separation of 20 m) which have been investigated at the Vienna campaign [10]. Figure 2 (Ref. [16]) illustrates how in the vicinity of the obstacles SV emerge and wrap around the primary vortex. Analogous physical mechanisms to those of the end effects

can be identified. Over 1000 wake vortex evolutions measured at Vienna International Airport indicate that plate lines cause the lifetimes of the vortices in a safety corridor along the final approach to reduce by 50% for the most critical ICAO separation (Medium behind Heavy) [10].

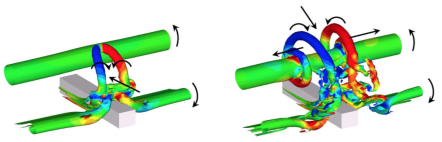


Fig. 2. Secondary vortices emerge and wrap around the primary vortex, propagate along the primary vortex, and lead to its accelerated decay (taken from [16], p.121).

Several wake vortex detection efforts to efficiently separate aircraft during the approach and landing phase have been published in the past, ranging from in-flight wake vortex determination [17] to prediction models such as the Probabilistic Two-Phase Wake Vortex Decay Model (P2P) [18]. Another method is to identify wake vortices within lidar measurements (or scans) which are recorded from a ground base. This technique was also employed at Vienna International Airport. Such set-up typically employs range-height indicator scans which measure the aerosol movement tangential to the lidar laser beam, perpendicular to the glide path of the landing aircraft (see Fig. 3). To evaluate the threat of the wake to a follower aircraft, the primary wake vortices, marked as CW (clockwise) and CCW (counterclockwise) vortices, have to quantitatively be identified via their circulation strength Γ and position (y, z) for each vortex. Wake Vortex Advisory Systems (WVASs), which obtain the position and strength of wake vortices, exist. They are based on fast-time prediction models such as the P2P. It is suggested that lidar instruments are employed to monitor these predictions, forming a safety net [19]. Lidar processing algorithms are widely studied, however for modern micro pulsed coherent lidar instruments (micro-PCDL) no automatic fast-time evaluation algorithm is available - a demanding task particularly in highly turbulent scenarios. For micro-PCDLs, a popular processing algorithm exists - the Radial Velocity (RV) method [20,21]. Assuming a Carrier-to-Noise Ratio (CNR) of -10 dB the RV method achieves root mean square errors (RMSEs) of up to 1.8 m, 0.21°, and 10.3 m^2/s for the range R, elevation angle φ, and circulation Γ, respectively. The seemingly high accuracy stems from theoretical analysis. Performance estimates for real measurement data, in particular for turbulent scenarios is still pending and expected to deviate from the aforementioned capabilities. Further downsides of the RV method include its requirement for manual input in the evaluation process and a relatively high CNR. Alternative ground-based algorithms [22–24] lack one or more of the following features: sufficient and/or quantifiable characterization accuracy, universal applicability for different lidar types, a complete automation (especially in turbulent atmosphere), the ability to process lidar measurements in fast-time, and the possibility to identify the origin of a result using context information.

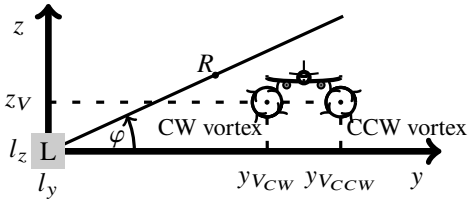


Fig. 3. Lidar measurement geometry with the aircraft flying out of the page.

Wake vortex detection and rough position estimations from lidar measurements have successfully been achieved with machine learning (ML) methods in the past [25–28]. In particular ANNs stood out as a promising method - other ML types obtained inferior performance. Employing ML for the quantitative characterization of the position and strength therefore posed an attractive path. A first study using ANNs for the aforementioned task indicated the feasibility of image processing architectures such as CNNs [29]. The trained CNN models [29] are based on archi-

tectures that have been developed for building an understanding on how hyperparameters of an ANN architecture influence the characterization capabilities. The training time on low level hardware (Intel® Core™ i7-5600U central processing unit at 2.60 GHz) was also a limiting factor. For the processing of a single lidar scan six individual scalar ANNs have been trained, two positional parameters to position each vortex and one strength parameter per vortex. Results indicate fast-time processing capabilities, with each lidar scan being processed within a tenth of a second. The custom CNNs investigated deliver positional and strength accuracies in the same order of magnitude as the traditional processing methods, nonetheless improvements are desired in future investigations. It is the effort of this work to investigate the characterization capabilities of the ANN models trained in [29], also for large campaign data sets. This shall give an insight into how ANNs may be adapted to increase their suitability for processing large measurement campaigns and being employed in WVASs.

3. Instrumentation and data sets

Vienna International Airport hosted a large scale measurement campaign in 2019 revolving around runway 16 and three Leosphere Windcube 200S (1.543-µm) micro-PCDLs alternatingly located at five different positions marked from L1 to L5 in Fig. 4. Two plate lines installed underneath the glide path towards runway 16 are marked with red dashes. About 9000 landings were measured with up to three vortex pair evolutions per overflight, each lidar position featuring slightly different elevation angle spectra, with the angle step being fixed at 0.2°. The range spectrum is universally 80 m to 530 m. LOS velocities (or radial velocities V_r) are determined by inspecting the Doppler frequency shift between an emitted and backscattered beam [30]. Lidar scans are not instantaneous captures, however the time difference between measurement initialization and finalization is considered negligible. We utilize lidar scans from this campaign to evaluate the use of ANNs for characterizing wake vortices in lidar scans.

Given that only measurement data from Vienna International Airport is employed, conclusions are restricted. Still, the use of a selection of lidar positions, scan geometries, vortex generators (aircraft types), and prevailing weather conditions throughout several seasons of the year, gives insight into the usability of ANNs for lidar measurement processing beyond the employed data. On top of this, the Vienna campaign features two kinds of scans, those with the novel plate lines erected and those with the plates flat on the ground in order to investigate the effectiveness of plate lines at mitigating wake vortices of landing aircraft. It was found that velocity fields generated differ significantly between the two plate line states. As a consequence we treat these scan categories separately. Note that it can occur that some scans have one plate line only partially erected, these scans are ignored.

This paper makes use of three different data sets (for each plate line state) emerging from the Vienna International campaign: training, validation and overall. The training and validation data set have equivalent settings and scan-case matching. Most importantly, scans part of these data sets have previously been processed with the RV method and can therefore be employed for evaluating the performance of the CNNs. However, they differ in their size. While the training data set is larger and has been used for learning features and ultimately developing the CNN models, the validation data set is employed for cross-checking the learned on independent lidar data. The overall data set emerges from all available lidar measurements at Vienna International Airport - scans must not have been processed with the RV method. Further details follow below, however since lidar measurements can include variations and perturbations from signal noise, atmospheric effects such as turbulence, SV, or erroneous data points in the velocity fields, we first describe lidar scan pre-processing applied to all lidar scans in this work prior to being fed to the CNNs.

It is seen as beneficial to mitigate confu-

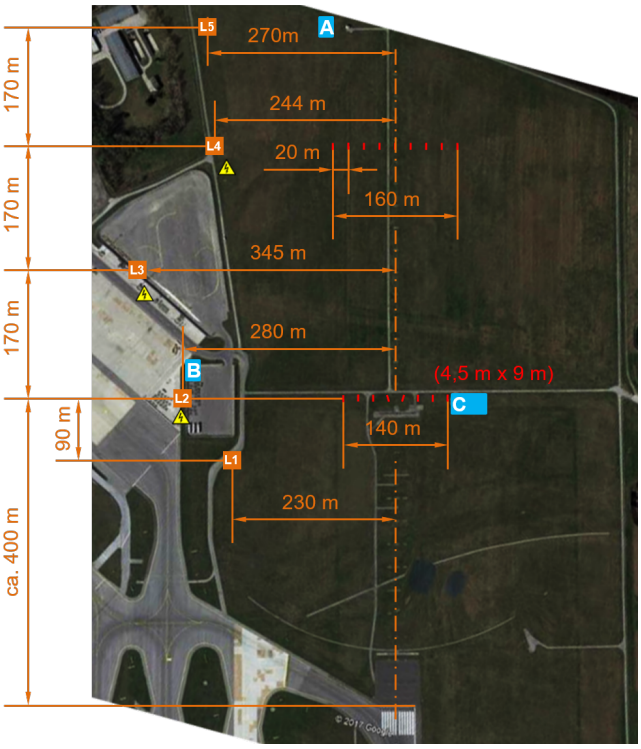


Fig. 4. Campaign set-up (red dashes = plates) (taken from [10], p.5).

sion which a crosswind can introduce to the characterization process of the ANNs. For each overflight (and the associated lidar), we identify a so-called background scan - the last lidar scan before the overflight occurred - which although being recorded a short time prior we consider to feature similar crosswind characteristics as the scan of concern. It is assumed that the wind does not alter significantly throughout the overflight. Analogous to the RV method, for each LOS of the background scan the mean LOS velocity is computed and subtracted from the associated LOS of wake vortex scan.

Boundary layers at the ground or interactions with plate lines may significantly impact vortex identification. To avoid the detrimental influence of hard target (plate) hits and high velocity gradients close to the ground, we disregard data points below an altitude of 7 m above the runway centerline and set their values artificially to zero. This altitude ensures that laser beams do not encounter a plate, while also clipping off substantial parts of the SV that have detached from the boundary layer. Vortex cores descending down to a minimum altitude of half of their initial vortex separation are not neglected with this approach.

In addition to easing the characterization capabilities of ANNs with our pre-processing, we also add generalizability by not differentiating at what lidar position the scan was recorded. We establish a universal measurement grid, spanning from the minimum to maximum elevation angle of any lidar position L1- L5, which is planted with the smaller size scans to match the LOS elevation angles. LOS not investigated for a lidar position are fed with a LOS velocity of zero. This approach not only increases generalizability of the models, but also gives access to a significantly larger training data set, which counteracts velocity field variations introduced with

the different lidar positions that all investigate slightly different phenomena along the glide path.

Lastly, since ANNs can deal better with input data that has a low variance, lidar scans are normalized such that all measurement points of the complete training data set have a LOS velocity mean of zero and a standard deviation of unity.

3.1. Training and validation data sets

The two data sets used for developing the ANNs and evaluating their performance in comparison to traditional lidar processing algorithms are the training and validation data sets. The former is directly integrated into the learning algorithm (gradient descent) of the ANNs, while the latter is used to judge the performance and thereafter make adaptations to the ANN architecture. Lidar measurements part of these data sets are chosen based on the available targets. In other words, whichever lidar scans have been processed with the RV method are also part of the ANN training and validation data sets. From all lidar scans available from Vienna International Airport those chosen to be evaluated with the RV method are low crosswind scenarios (as this represents the worst-case WVE conditions).

For evaluating the performance of the ANNs, the overflight cases and the lidar scans must be matched with one another. Models trained herein always assume two wake vortices (one pair) to be present in a lidar scan. This assumption does not hold for all cases: more than two vortices can be present in a scan when wake vortices of a previous overflight hover over the runway for long periods of time. Less than two vortices are present in a scan if they have been convected outside the measurement window of the lidar. The RV method only processed measurements from Vienna International Airport with at least one wake vortex detected and therefore 'empty' scans are already not part of the training nor validation data set. In the training and validation data set lidar scans exceeding the primary pair are shown to the ANN twice. In other words, a scan with three vortices will be treated separately for the pair and the additional vortex. Scans where the RV method only detected one vortex are treated with an imaginary wake vortex which has all its parameters (position and strength) set to zero. Knowledge of which vortices belong to which overflight can be derived from the RV method and manual analysis. Fixing unavailable targets to a pre-defined value is a common approach for missing data in ANNs, yielding the ability to learn that this value is designated for missing data

Just under 500 vortex pair evolutions were evaluated using the RV method and used for training and validation of the ANNs, with over 70% associated to the Medium ICAO weight class. Approximate initial vortex circulation strengths for each aircraft class are 650 - 750 m^2/s for Super, 300 - 550 m^2/s for Heavy, and 200 - 350 m^2/s for Medium aircraft. Roughly 41% of overflight measurements were pursued with plate lines erected.

The characterizations used as training and validation data sets are subject to uncertainties inherent to the RV method [20, 21]. Furthermore, typically large discrepancies between the RV method and ANNs occur when the latter detect a turbulent eddy rather than a wake vortex. Analysis of the estimations by the ANNs revealed that large deviations to the RV method targets typically originate from factors unrelated to the concrete ANN method. For instance, incorrect data labels/targets cannot be corrected by the ANN. Cases which result from such an error are disregarded in the analysis found herein through post-processing. Effectively this has the consequence that the validation data set is restricted to lidar scans with exactly two wake vortices. Less vortices cause either large errors with artificial labels, or more vortices cannot be explicitly assigned to the foregoing or forthcoming overflight.

To avoid previous knowledge of vortex constellations, vortex evolutions have been randomly allocated in full either in the training of validation data set.

3.2. Overall data set

With the goal of evaluating large measurement campaigns we also consider the entire

data recorded at Vienna International Airport in an overall data set. This includes lidar scans with any amount of vortices present, even none. For this data set RV method detections often do not exist and as a consequence the matching of the overflight case and scan cannot be performed as for the training and validation data set. Here lidar scans are categorized between overflights and associated to the overflight that occurred prior to each scan. This data set is supposed to show how our ANNs can perform on large data sets with limited information about the overflight and scan matching. It represents a realistic scenario which application of this method will face.

4. ANN architecture

In this effort we use the ANN architecture developed in [29] shown in Fig. 5. The design of this CNN aims for reasonable ANN training times on our hardware and using the image processing capabilities of CNNs. Complex velocity fields captured by the lidar instruments are broken down to a small number of parameters. In fact, the output shape of the CNN architecture is unity, signaling that the herein trained ANNs represent scalar architectures. It is assumed that each lidar scan contains two vortices, which are both individually characterized by their strength (circulation) and position (range and elevation angle). Thus, six separate CNNs (one for each parameter and vortex) are employed. The CNN is made up of four ConvPool blocks - a convolutional layer followed by a max pooling layer. The fourth ConvPool block features an additional max pooling layer in order to limit the number of parameters resulting from the flatten layer transformation. The activation function employed in the neuron operation is the rectified linear unit (ReLu). With ReLu it is possible to establish a nonlinear relationship between the lidar scans and vortex parameters.

Training of the CNN introduced was accomplished with the gradient descent algorithm, which is typical for supervised tasks. The Adaptive Moment Estimation (ADAM) optimizer was used in the gradient descent operation to enable accelerated and more efficient minima results in the weight space. In the case of our work, the targets or labels for the input data sets (training and validation data sets from Section 3.1) originate from the RV method. We employed the mean square error (MSE) between the CNN output and RV method target as a loss function. Following computation of the loss, the gradient to each weight is established, and subsequently the weights are updated in the opposite direction to their gradient. The algorithm speed can be adapted by using the learning rate η and an optimizer function, aimed at enabling more straightforward discovery of minima in the weight space. In the current effort we trained the CNNs for a maximum of 100 epochs (iterations of going through the training data set). If no improvement was registered in the validation data set for 30 epochs training was terminated earlier, avoiding overfitting of the models to the training data.

For this work's purpose, lidar scans can be seen as LOS velocity matrices. Hence, the channel size of an associated image is one - similar to a black and white image. The filters of the convolution layers are responsible for generating activation maps - matrices which indicate where and at what magnitude the filter's pattern is detected in the input matrix. In this work the filter size was 3×3, the padding was such that the scan size does not alter with the convolutional procedure, and the stride was set to unity. The number of filters started with 32 in the first convolution layer and doubled with each ConvPool block. The number of neurons in the dense layers was 64. The pooling layers made use of a filter size 2×2 and stride of two, reducing the scan size by half with each pooling layer.

For dense layers, the number of trainable parameters P_d depends on the number of neurons in the layer OS, and the number of input data points per scan IS for each layer (defined by index l). It is described with Eq. (1). For convolution layers, the number of trainable parameters P_c depends on the number of filters in the layer OC, the number of input channels IC, as well as the filter height and width $F_h \times F_w$. It is described with Eq.

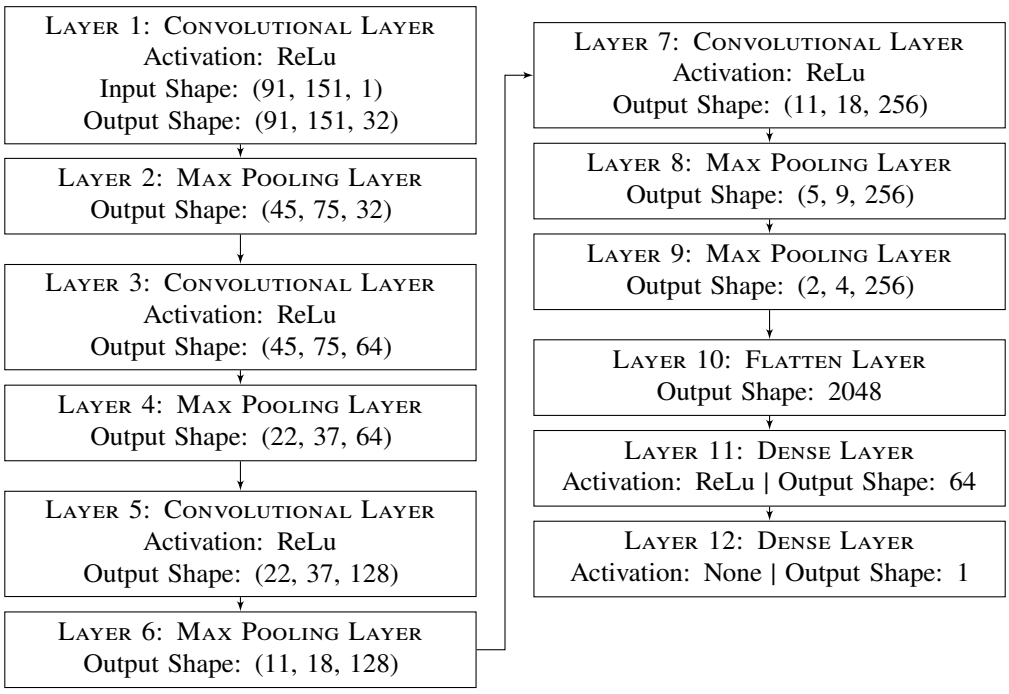


Fig. 5. Convolutional Neural Network architecture.

(2). For lidar scans of size 91×151 the total number of trainable parameters is 519041.

$$P_d^l = OS^l \left(IS^l + 1\right) \tag{1}$$

$$P_c^l = OC^l \left[(IC^l \times F_h \times F_w) + 1\right] \tag{2}$$

5. Results

With this section we aim to gradually increase the level of performance analysis, starting from qualitative discussions, to statistical hazard analysis, comparisons of quantitative metrics with the RV method, and the examination of ANN performance for large lidar measurement data sets. Both the validation data set and a scoring algorithm are used to judge the characterization capabilities of the ANNs. Whereas the validation data set offers a comparison to the traditional RV method, the scoring algorithm can be compared to the trends found with the validation data set in order to then evaluate the appropriateness of the scoring algorithm. This allows independently processing lidar scans with ANNs (with no RV method detections for verification).

5.1. Qualitative analysis

A typical overflight with corresponding detections from the CNNs as well as detections from the RV method are presented in Fig. 6. We can see individual time frames of a Heavy generator aircraft with plate lines flat on the ground. The time frames illustrate that young vortices are coherent featuring high circulation strengths. Young vortices can easily be distinguished from the background turbulence present in lidar scans even after the pre-processing which removed a constant crosswind. Although minimum radar separation behind a Heavy generator aircraft is 60 s and here the CNNs characterize the wake vortices with the highest precision, we also see satisfactory estimates in the time frames after that. Still, we can see the CNN accuracy deteriorates with time, in particular for the CW vortex. If we inspect Fig. 6(f), the progressed state of vortex erosion makes it challenging to identify where the CW vortex

is positioned in reality. Nonetheless, when the localization error grows, we observe the circulation still being estimated in the correct order of magnitude. Thus, when considering both parameters in union, WVEs are unlikely to be missed.

5.2. Hazard detection analysis

We extend our analysis by inspecting how well WVE risks are recognized. Ultimately the goal is to understand the hazard a vortex poses above the runway. For this we individually inspect the vortex strength and the position. Whether knowledge of the position or strength of a wake vortex is more critical depends on the application of such algorithm.

5.2.1. Localization

If it is of interest whether a vortex is hovering over the runway and posing a 'potential hazard' or not, a ±50 m lateral safety corridor can be defined outside which vortices are not at risk to lead to WVEs [10, 31], independent of the vortex strength. Table 1 shows whether the localization regions of the RV method and CNNs - either within or outside the corridor - match. The False-Positives rate is defined as $FPR = FP/(FP+TN)$, whereas the False-Negatives rate is given by $FNR = FN/(FN+TP)$. False-Negatives (FN) are defined as detections where a vortex was incorrectly detected outside the safety corridor. In general, False-Negatives are more common than False-Positives (FP). Note that correct vortex allocations are analogously defined using True-Negatives (TN) and True-Positives (TP). Lastly we investigate the positive predictive rate ($PPR = TP/(TP+FP)$), the probability of predicting a potential hazard when there actually is one - it is considered to be one of the most vital metrics. Achieving up to 89% is a very satisfactory result in that regard.

5.2.2. Circulation

Analogous to our localization analysis we can use statistical metrics in the circulation strength characterization to understand the abilities of the CNNs further. In Table 2 a wake vortex is considered a 'potential hazard' in case it has a circulation strength of above 100 m^2/s (as defined by Fig. 7(b)), regardless of its position above the runway. The positive predictive rate shows even better performance with this definition. We can expect the CNNs to be about 93% accurate.

Table 1. Statistics regarding localizations in a runway corridor (CW and CCW vortices). Positive/negative are vortices within/outside the corridor. The RV method is used for comparison.

Metric	Plates down	Plates up
Match (%)	86.0	66.1
FPR (%)	13.4	26.7
FNR (%)	14.5	38.5
PPR (%)	88.6	78.6

Table 2. Statistics regarding the circulation characterization (CW and CCW vortices). Positive/negative are vortices above/below 100 m^2/s. The RV method is used for comparison.

Metric	Plates down	Plates up
FPR (%)	17.6	19.9
FNR (%)	7.4	9.6
PPR (%)	93.3	93.1

5.3. Comparison to the RV method

Although WVE hazard judgment is the essential mechanism which needs to function, defining further validation metrics to analyze the CNN performance in comparison to the RV method appears helpful in understanding the dynamics of the vortices, as well as the CNN accuracy. These additional metrics are: the RMSE responsible for comparing the circulation strength as well as the individual vortex center positions of the RV method and ANNs, and ΔD which represents the Euclidean distance between the vortex center defined by the RV method and the ANNs. We also define the relative vortex circulation error and the normalized vortex circulation

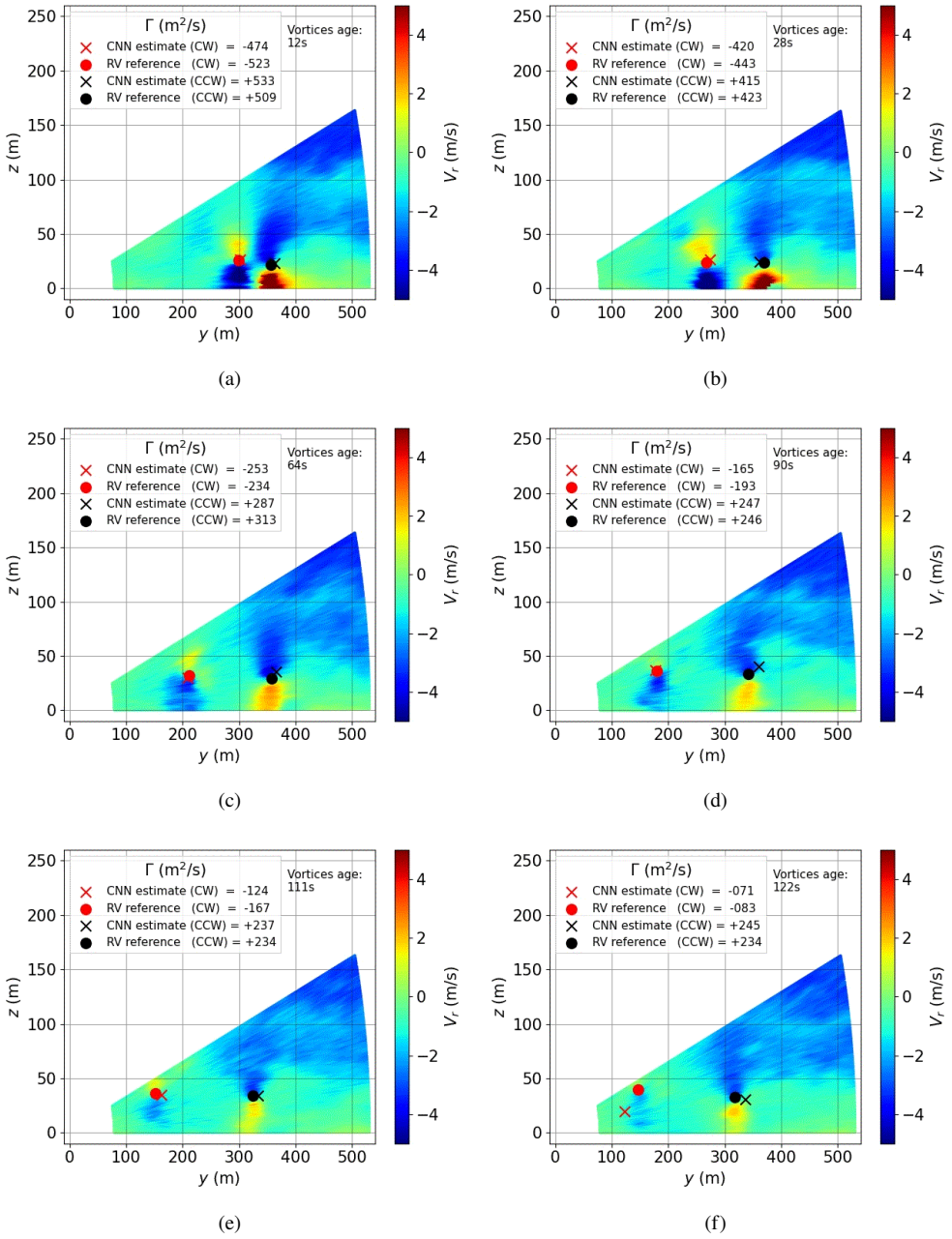


Fig. 6. Time frames with predictions and targets of a B777 overflight without plates.

error. Both make use of the absolute difference between the circulation strengths estimated by the RV method and the CNN i.e. $\Delta\Gamma = \mid \Gamma_{RV} - \Gamma_{CNN} \mid$. The relative circulation error is divided by the RV method's circulation to obtain $\Delta\Gamma/\Gamma_{RV}$. Similarly, the normalized circulation error $\Delta\Gamma^* = \Delta\Gamma/\Gamma_0$ is given by dividing the circulation error $\Delta\Gamma$ by the initial vortex circulation Γ_0 derived from the aircraft's maximum landing weight, air density and true airspeed as detailed in [10]. In rare cases Γ_0 is unavailable from this method, as an alternative mean values of that aircraft type are used - standard deviations of at most 4% promise to give representative Γ_0 values.

Table 3 shows that the CW wake vortex leads to higher errors than the CCW wake vortex for nearly all parameters, indicating that the CNN performance is generally worse for the CW vortices. Several causes can be identified which could lead to this observation, all of which are linked to the usual vortex position in the measurement window. First, the training and validation data sets we are using for the current effort contain approximately 20% more CCW vortices than CW vortices. The difference stems from the CW vortices being transported out of the lidar measurement window far more frequently than the CCW vortex (given that only lower altitudes of the atmosphere are covered at lower ranges from the lidar). Second, the focal length of the lidar could play a role in facilitating wake vortex characterization. In the case of the Vienna campaign, the focal length of the lidar instruments was set to roughly 500 m, which we can realize to be in close vicinity to the usual CCW vortex position, rather than the usual CW vortex position. The lidar focal length decides where we can find the highest CNR measurement. While it was expected that grid coarseness could also impact the characterization capabilities, this could not be confirmed. Generally the CCW vortex is placed at higher range than the CW vortex and therefore has a worse resolution. Apparently grid coarseness has little effect on the characterization quality or the effect is counteracted by the two previous points.

Plate lines evidently also affect the characterization of wake vortices (see Table 3). We affiliate the weakened accuracy capabilities when plate lines are erected with their ability to seriously dismantle the coherency of the vortices. The coherency loss may correspond to a larger variation in vortices in the lidar scans and thus complicate generalization of features further. Particularly the localization of the vortices is affected, while the circulation accuracy worsens only marginally.

Table 3 also includes the RV method error estimates for low turbulence scenarios based on two-dimensional simulated data. These stem from theoretical analysis, rather than observations or even analysis of our data. We can observe that the RV method is substantially more accurate in all characterizations. While the circulation can be estimated with the same order of magnitude by the CNN, both range and elevation angle have an error with an order of magnitude higher for the

Table 3. Errors with 8925/6520 train & 700 - 1100 validate scans (plates down/up).

Metric	Γ RMSE (m^2/s)		φ_V RMSE (°)		R_V RMSE (m)		Median ΔD (m)	
	CW	CCW	CW	CCW	CW	CCW	CW	CCW
CNN								
(Down)	44.03	35.61	2.46	1.53	35.82	35.22	11.49	12.39
(Up)	45.93	43.18	2.04	1.21	56.73	66.33	39.50	31.60
RV [20]								
(Either)	10.30		0.21		1.80		-	

CNN than the RV method. The difference in the accuracy between the RV method and the CNNs is expected. Evidently we cannot expect the CNNs to surpass the accuracy of the RV method which was used for training of the CNNs. The CNNs are especially vulnerable when the RV method has incorrectly or inaccurately identified a wake vortex in a lidar scan. In such case, the learned features of the CNNs might have to choose a compromise parameter value to satisfy both the learned features as well as the targets of the RV method. This overall leads to an increase in the herein presented errors, with their origin only partly evolving from the CNN operation itself. It should be kept in mind that lidar instruments themselves carry intrinsic measurement limitations. Given these limitations one of the future ambitions is to generate artificial lidar measurements using high fidelity aircraft landing simulations. Alternatively, the scoring algorithm introduced in Section 5.4 could be integrated as a loss function for the ANNs. Both of these methods would eliminate the dependency of the RV method and its targets, also allowing to determine the accuracy of the RV method.

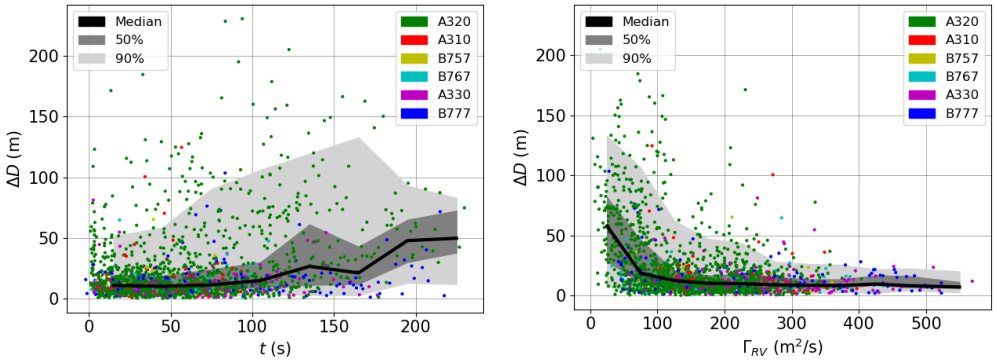
The localization error ΔD can be investigated further, by looking at how ΔD varies with vortex age and vortex strength. In Fig. 7 we have grouped overflights with the corresponding aircraft type. Note that the A320neo (A20N) is considered part of the A320 family, it features the same maximum landing weight and wingspan. The dark grey area is limited by the 25^{th} and 75^{th} percentiles (thus containing 50% of all data points), whereas the light grey area encloses the 5^{th} and 95^{th} percentiles (90% of all data points). The vortex-age-bins have dimensional widths of 30 s and circulation-bins have dimensional widths of 50 m^2/s (with the exception of one circulation-bin which spans 500 m^2/s to 600 m^2/s). In Fig. 7(a) we observe localization precision decreasing with aging vortices. Older vortices are typically weaker and thus Fig. 7(b) correspondingly indicates high ΔD at these vortex strengths. Note however that far less data points exist at older vortex age making the percentiles less reliable. We can see a circulation strength threshold of 100 m^2/s underneath which vortices are tougher to recognize. Arguably, the circulation strength regions of concern (above 100 m^2/s) are covered with higher precisions. We can also see very young incoherent vortices still rolling up, giving higher ΔD values.

Comparison to the vortex age and RV method circulation strengths is also of interest for the characterization of the circulation strengths of the CNNs (see Fig. 8). The relative circulation error (left) confirms the error increase with older vortex age and lower circulation strengths. Figure 8(c) also confirms the circulation strength threshold of 100 m^2/s. Using the normalized vortex circulation error plots (right) we can see how smaller aircraft such as the A320, generating smaller and weaker vortices, are more challenging to characterize in comparison to larger Heavy aircraft such as the B777. This occurs despite the validation data set containing far more A320 lidar scans and inspecting the same vortex age or circulation strength.

Processing times of the CNNs are significantly outperforming those of the RV method. While characterizing the position and strength of wake vortices in a single lidar scan requires manual operations and takes several seconds to obtain the results, the herein trained CNN processing time amounts to 0.16 s on our low level hardware.

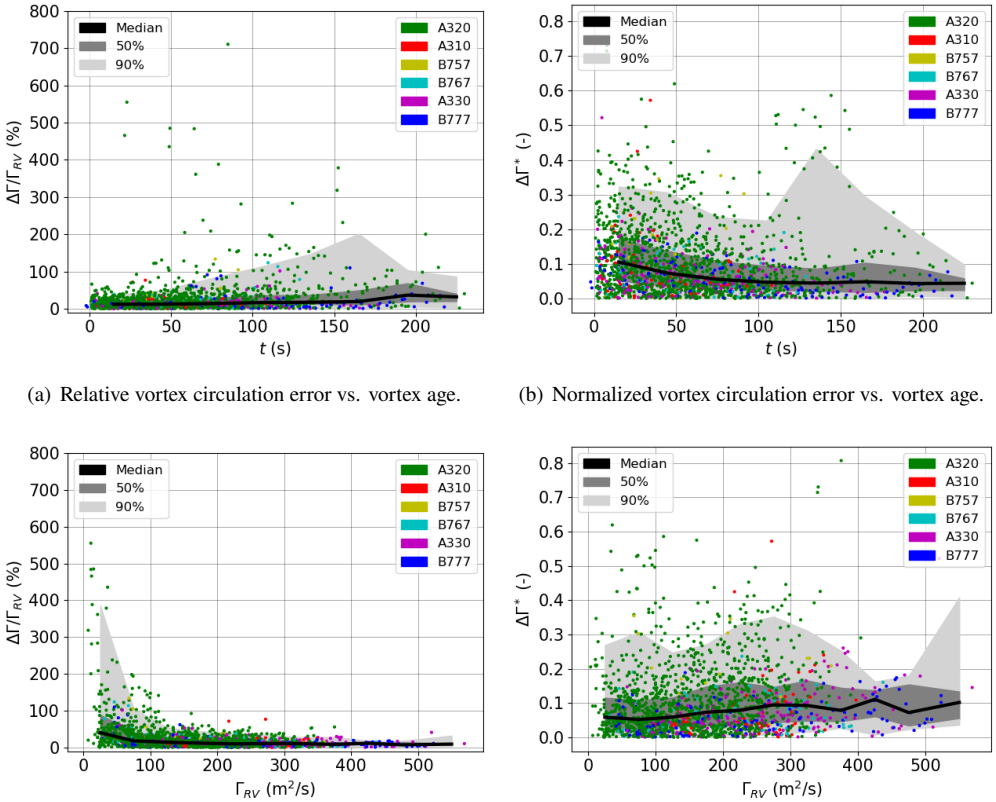
5.4. Analysis of large measurement campaign data sets using the scoring algorithm

In order to judge the ANN vortex characterization capabilities more objectively, independent from uncertainties of the RV method, a scoring method was developed, which calculates a confidence score for each individual vortex characterization of the ANN. The results of the scoring are given in terms of traffic light colors: green indicates high confidence in the characterization result, yellow indicates a limit case, red indicates that the characterization should be discarded, and grey indicates that the characterization is outside of the measurement area.



(a) Temporal evolution. (b) Circulation strength influence.

Fig. 7. CNN ΔD correlation to vortex parameters for plate down lidar scans.



(a) Relative vortex circulation error vs. vortex age. (b) Normalized vortex circulation error vs. vortex age.

(c) Relative vortex circulation error vs. RV circulation. (d) Normalized vortex circulation error vs. RV circulation.

Fig. 8. Circulation error dependent on vortex age and RV reference circulation without plates.

To calculate the scoring result, the measured flow field of each lidar scan is compared to a potential flow model derived from the wake vortex characterizations made by the ANN. The vortices are represented by the Burnham-Hallock vortex model [32], which

includes parameters for vertical and horizontal position, circulation, and core radius. The background wind is assumed to be in horizontal direction only and independent from the altitude. While the position and circulation parameters are taken directly from the ANN characterization, the horizontal wind component and the core radius are obtained by fitting the potential flow model to the measured wind field. Fig. 9 shows the radial velocity as seen by the lidar for the measured field and for the reconstructed model. The scoring result is then calculated based on the scan's median CNR value, the calculated core radius, and the normalized RMSE (cf. Fig. 10). The normalized RMSE is calculated by multiplying the RMSE between the measured and the modeled radial velocity with $b_0/\Gamma(t)$, where b_0 is the initial vortex spacing and $\Gamma(t)$ is the circulation as indicated by the ANN.

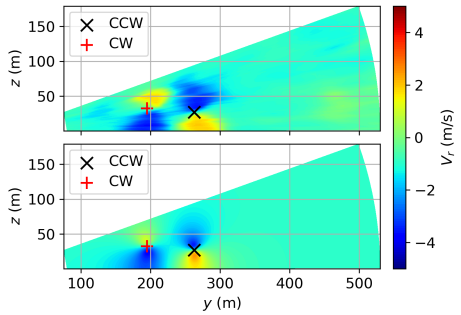


Fig. 9. Comparison between lidar measurement (top) and model of the scoring algorithm (bottom).

We have previously established that evaluating comprehensive wake vortex lidar data sets is one driver of this effort. In such a case no RV method targets are available, thus we employ the above introduced scoring algorithm to evaluate the CNN characterizations. In this section we also aim to confirm the functionality of the scoring algorithm, by comparing data trends to those found in Section 5.3. Here we consider all overflight cases encountered during the months of the measurement campaign - the overall data set. Thus, also scans with any number of vortices are analyzed. While the CNN always characterizes two vortices for each scan, even for 'empty scans', the scoring can be used to select only sensible results of the CNN.

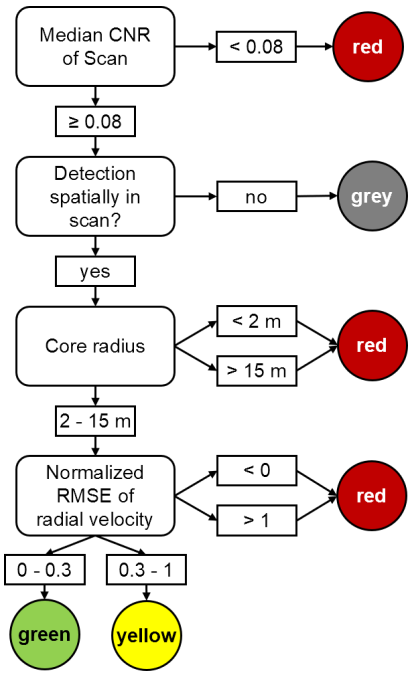


Fig. 10. Logic of the scoring algorithm.

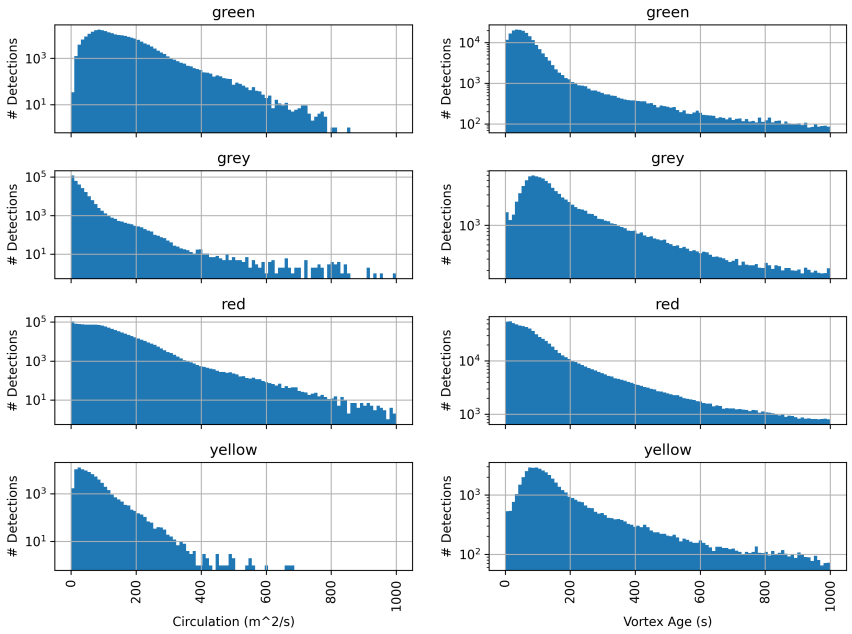
Figure 11 summarizes the scoring results of evaluating all the data from the Vienna International Airport campaign (about two million individual vortex detections) using the CNN and the scoring algorithm with respect to the circulation, vortex age and position. Figure 11(a) shows the histogram of the circulation indicated by the CNN based on the scoring result. All scoring results show distinct circulation distributions, indicating that the scoring differentiates between different phenomena. Green scoring results appear for any circulation value between 0 m^2/s and 800 m^2/s. Weak vortices appear to be more common in the overall data set than strong vortices. The drastic reduction of green detections at low circulation strength matches the identified threshold of 100 m^2/s from before. Note that vortex detections disregarded by the scoring (grey and red) do not show this characteristic. Previous model validation only evaluated scans with two vortices, low circulation errors are there-

fore associated with the red scoring. Here an additional factor comes into play. When less than two vortices are found within a scan, the CNNs aim to place the missing vortices of the pair at the origin outside of the scan, this is represented by the grey plot. It should be noted that grey scoring does not rate the CNN characterization quality, rather it labels detections as irrelevant. The green detections deplete towards higher circulation strengths, however this can be tracked back to less data points at these magnitudes. Yellow scoring of a detection urges to manually check the result. We can observe that the yellow scoring plot represent a combination of the green and red scoring plots - as expected. Additionally, yellow scoring results do not exist for high circulation values: strong vortices are easier to distinguish for the scoring algorithm.

Detection scoring results in comparison to their vortex age (Fig. 11(b)) unveils similar conclusions as previously seen for the RV method verified data set (the validation data set). Most green scoring detections occur at young vortex age, with worse detections during vortex roll-up. The drastic reduction of green detections at old vortex age is due to the lower data availability within these regions. It is interesting that the peak of yellow scoring is not an intermediate of the green and red scoring plots. From the green scoring plot's unique peak it becomes clear that coherency of the vortices is crucial for their characterization.

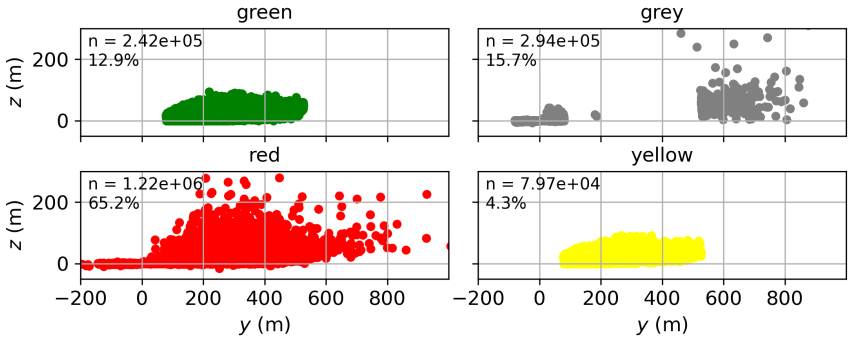
Figure 11(c) not only visualizes where green, red, yellow and grey scoring detections occur, but also how many. Since the overall data set includes also lidar scans with no vortex at all, but the CNNs always produce detections for a vortex pair, the majority of detections are red. The green detections are compact within the typical measurement window of the campaign (as defined by the universal measurement grid). The absolute number of green detections is far greater than the number of detections evaluated using the RV method, resulting in a substantial benefit already. Yellow detections are typically also within the defined universal measurement grid, whereas grey and red can be placed also outside the window. In fact, we would expect all grey detections outside the measurement window - a computational error has caused an outlier in the plot.

From this exemplary large campaign data set (the overall data set) analysis we can see that trends match those found for the validation data set as shown by Fig. 7 and Fig. 8 (which we can compare to the RV method). To substantiate the above an exemplary overflight vortex trajectory is given in Fig. 12. The figure represents RV method detections with crosses and CNN characterizations with colored dots. The color map follows that of the scoring. The position and circulation values are obtained by the CNNs, the core radius, crosswind, and mean CNR values are obtained by the scoring algorithm. We can observe that the circulation and position characterizations are fairly accurate in the majority of detections, as indicated by the green scoring and the good match with the characterizations of the RV method. Note that the scoring result is independent from the RV method. We can see how young incoherent vortices are harder to identify by the CNNs as indicated by big differences to the RV method. The scoring catches these wrong characterizations and labels them as red. At around 100 s, an outlier circulation of 100 m^2/s is recognized by both the RV method and CNN, which is labeled yellow by the scoring (indicating to double-check this result). Considering the temporal evolution of the vortex before and after this outlier, this detection could be excluded from further analysis. On the other hand, both the RV method and the CNN indicate that a vortex is present. Complex atmospheric turbulence in this scan may cause this discrepancy between the simulated wind and the real measurement. Towards older vortex age, beyond where RV method detections exist, the scoring begins to class CNN detections as yellow or red. This is expected, the vortices lose coherency. Nonetheless, the trends in circulation and position are in-line with that where the RV method was still able to detect the vortices. As a consequence, it can be said that the CNNs are able to track vortices for



(a) Vortex strength.

(b) Vortex age.



(c) Vortex position.

Fig. 11. Scoring result for all cases with respect to different parameters.

longer periods of time and they can distinguish vortices from background turbulence in a superior manner.

Throughout this work we have seen the circulation strength being characterized more precisely than the vortex position. Since the CNNs for position and circulation are trained independently, it is of interest whether there is a correlation between low ΔD and low $\Delta\Gamma^*$. Figure 13 investigates this correlation (thus only investigating RV method processed validation data). No correlation between the

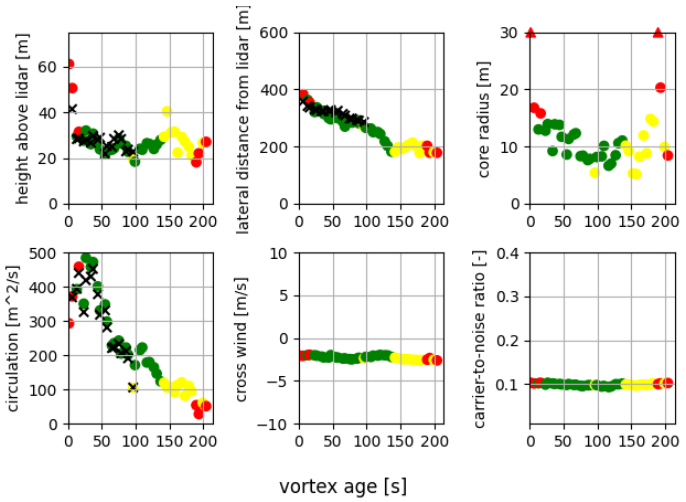


Fig. 12. B777 port vortex trajectory from lidar position L3 with plates erected. Dots represent CNN detections, crosses denote RV method detections. Triangles represent detections outside the shown plot window for clarity.

localization error and normalized circulation strength error can be identified, suggesting that the position and strength of a vortex are determined using differing features. Learned filters of the CNNs are most likely also dissimilar.

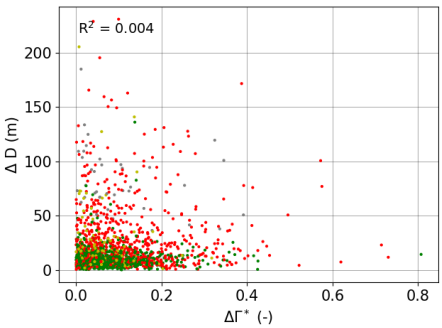


Fig. 13. Correlation between the localization and circulation strength error. Colors follow those of the scoring algorithm.

The color map of Fig. 13 also confirms much lower errors for the green scoring detections. As expected, yellow scoring detections are slightly further spread and red scoring detections correlate to the most significant errors. This result confirms the functionality of the scoring method. We can therefore employ the scoring method also for lidar scans which have not been processed by the RV method. In other words, the combination of the ANN and the scoring algorithm can act as a substitute for the cumbersome RV method.

With these results, we can make several conclusions about future implementations in this algorithm which may benefit the characterization of the landing aircraft wake vortices. First, we suggest to interconnect the CNNs for the three different parameters per vortex. It has previously been attempted to implement vector ANNs rather than scalar ANNs (one ANN with a vector output characterizing all parameters at once rather than six individual ANNs with no knowledge of one another) [29]. Although it was found that the scalar ANNs option perform slightly better, it might be wasted potential to not attempt another way of linking the ANNs. For instance, it has so far not been attempted to solely train the localization parameters at once i.e. elevation angle and range. Second, the realization that localization and circulation strength parameters are characterized based on distinguishing parameters can turn out to be a benefit. We could first use CNNs to characterize the circulation strength of the wake vortex and thereafter use these results

also as an input for the localization CNNs. Third, we believe it to be of benefit to first employ a classification ANN as done in [25], getting away from the restriction to always have to characterize two wake vortices in a lidar scan. Furthermore, usage of the temporal component in one vortex evolution could be advantageous, as illustrated in Fig. 12. By inputting the previous position and strength of the wake vortex into the algorithm for the next lidar scan, a fairly accurate first guess or limitation for the parameter space could be implemented. Lastly, the training, validation and overall data set should share the same method of how overflight cases and lidar scans are matched. Different matching can cause confusion especially in scenarios where more or less than two vortices are present in a lidar scan. Still, the current ANN architecture provides a good overview of the possibility to use CNNs for characterizing wake vortices in lidar scans. With the current effort we deepened our analysis of the CNN capabilities, showed how these could be applied to large data sets and suggested possible future work.

6. Conclusion

This work has employed Convolutional Neural Networks (CNNs) to quantitatively characterize the position and circulation of wake vortices in Light Detection and Ranging (lidar) scans from a large data set measured at Vienna International Airport. Previously only a fraction of the measurements have been processed with the Radial Velocity (RV) method, with the current effort we managed to characterize also the large remainder of lidar scans. The Burnham-Hallock vortex model has been employed in a scoring algorithm, which we show to correlate well with trends observed from the CNN versus RV method comparison. The scoring algorithm offers a validity check of the CNN characterizations. We can therefore use the CNNs in combination with the scoring algorithm to identify and characterize wake vortices in lidar measurements with no other background knowledge. Previous efforts have exclusively focused on identifying vortices with approximate positions. Although the RV method still obtains higher accuracies, it lacks full automation and the ability to process lidar scans in fast-time. These deficiencies were overcome with our CNNs.

In order to tackle the remaining downsides of the CNNs, we suggest to make use of temporal factors of the data set, as well as the different features learned in the position and circulation CNNs (detected from the lack of correlation), and we aim to implement an algorithm which first checks a lidar scan for the number of vortices, allowing their independent analysis.

Acknowledgments. The Vienna measurement campaign has received funding within the framework of the SESAR Joint Undertaking "Increased Runway and Airport Throughput" project (PJ.02 EARTH) within the European Union's Horizon 2020 research and innovation programme under grant agreement No 731781 as well as from the German Aerospace Research Center (DLR) project "Wetter und disruptive Ereignisse".

References

1. J. N. Hallock and F. Holzäpfel, "A review of recent wake vortex research for increasing airport capacity," Prog. Aerosp. Sci. **98**, 27–36 (2018).
2. A. Stephan, D. Rohlmann, F. Holzäpfel, and R. Rudnik, "Effects of detailed aircraft geometry on wake vortex dynamics during landing," J. Aircr. **56**, 974–989 (2019).
3. V. Rossow and K. James, "Overview of wake-vortex hazards during cruise," J. Aircr. **37**, 960–975 (2000).
4. J. Critchley and P. Foot, "Uk caa wake vortex database: analysis of incidents reported between 1982 and 1990," CAA Pap. **91** (1991).
5. F. Holzäpfel and M. Steen, "Aircraft wake-vortex evolution in ground proximity: analysis and parameterization," Am. Inst. Aeronaut. Astronaut. journal **45**, 218–227 (2007).
6. T. Gerz, F. Holzäpfel, and D. Darracq, "Commercial aircraft wake vortices," Prog. Aerosp. Sci. **38**, 181–208 (2002).
7. Eurocontrol, "Eurocontrol forecast update 2021-2024," (2021).
8. Eurocontrol, "Challenges of growth," Tech. rep., Eurocontrol (2018). European aviation in 2040.
9. J. Cheng, A. Hoff, J. Tittsworth, and W. A. Gallo, "The development of wake turbulence re-categorization in the united states," in *8th American Institute of Aeronautics and Astronautics Atmospheric and Space Environments Conference,* (2016), pp. 1–12.
10. F. Holzäpfel, A. Stephan, G. Rotshteyn, S. Körner, N. Wildmann, L. Oswald, T. Gerz, G. Borek, A. Floh, C. Kern, M. Kerschbaum, R. Nossal, J. Schwarzenbacher, M. Strobel, L. Strauss, C. Weiß, S. Kauczok,

C. Schiefer, H. Czekala, G. Maschwitz, and I. Smalikho, "Mitigating wake turbulence risk during final approach via plate lines," AIAA J. **59**, 4626–4641 (2021).
11. R. E. Robins, D. P. Delisi, and G. C. Greene, "Algorithm for prediction of trailing vortex evolution," J. aircraft **38**, 911–917 (2001).
12. J. Harvey and F. J. Perry, "Flowfield produced by trailing vortices in the vicinity of the ground," AIAA journal **9**, 1659–1660 (1971).
13. P. Spalart, M. K. Strelets, A. Travin, and M. Shur, "Modeling the interaction of a vortex pair with the ground," Fluid Dyn. **36**, 899–908 (2001).
14. Z. Zheng and R. L. Ash, "Study of aircraft wake vortex behavior near the ground," AIAA journal **34**, 580–589 (1996).
15. A. Stephan, F. Holzäpfel, and T. Misaka, "Aircraft wake-vortex decay in ground proximity—physical mechanisms and artificial enhancement," J. Aircr. **50**, 1250–1260 (2013).
16. A. Stephan, F. Holzäpfel, T. Misaka, R. Geisler, and R. Konrath, "Enhancement of aircraft wake vortex decay in ground proximity," CEAS Aeronaut. J. **5**, 109–125 (2014).
17. N. Fezans, J. Schwithal, and D. Fischenberg, "In-flight remote sensing and identification of gusts, turbulence, and wake vortices using a doppler lidar," CEAS Aeronaut. J. **8**, 313–333 (2017).
18. F. Holzäpfel, "Probabilistic two-phase aircraft wake-vortex model: further development and assessment," J. Aircr. **43**, 700–708 (2006).
19. F. Holzäpfel, T. Gerz, M. Frech, A. Tafferner, F. Köpp, I. N. Smalikho, S. Rahm, K.-U. Hahn, and C. Schwarz, "The wake vortex prediction and monitoring system wsvbs part i: Design," Air Traffic Control. Q. **17**, 301–322 (2009).
20. I. N. Smalikho, V. Banakh, F. Holzäpfel, and S. Rahm, "Method of radial velocities for the estimation of aircraft wake vortex parameters from data measured by coherent doppler lidar," Opt. Express **23**, A1194–A1207 (2015).
21. I. N. Smalikho and V. Banakh, "Estimation of aircraft wake vortex parameters from data measured with a 1.5-μ m coherent doppler lidar," Opt. letters **40**, 3408–3411 (2015).
22. F. Köpp, S. Rahm, and I. N. Smalikho, "Characterization of aircraft wake vortices by 2-μ m pulsed doppler lidar," J. Atmospheric Ocean. Technol. **21**, 194–206 (2004).
23. V. Banakh and I. N. Smalikho, *Coherent Doppler Wind Lidars in a Turbulent Atmosphere* (Artech House, 2013), chap. 5.
24. L. Thobois, J.-P. Cariou, V. Cappellazzo, C. Musson, and V. Treve, "Comparison and validation of wake vortex characteristics collected at different airports by different scanning lidar sensors," in *EPJ Web of Conferences,* vol. 176 (EDP Sciences, 2018), p. 06002.
25. P. Weijun, D. Yingjie, Z. Qiang, T. Jiahao, and Z. Jun, "Deep learning for aircraft wake vortex identification," in *IOP Conference Series: Materials Science and Engineering,* vol. 685 (IOP Publishing, 2019), pp. 1–8.
26. P. Weijun, D. Yingjie, Z. Qiang, W. Zhengyuan, and L. Haochen, "Research on aircraft wake vortex recognition using alexnet," Opto-Electronic Eng. **46** (2019).
27. P. Weijun, Z. Wu, and X. Zhang, "Identification of aircraft wake vortex based on svm," Math. Probl. Eng. **2020** (2020).
28. N. Baranov and B. Resnick, "Wake vortex detection by convolutional neural networks," Int. J. Electr. Eng. Comput. Sci. (EEACS) **3**, 92–97 (2021).
29. N. Wartha, A. Stephan, F. Holzäpfel, and G. Rotshteyn, "Characterizing aircraft wake vortex position and strength using lidar measurements processed with artificial neural networks," Opt. Express **30**, 13197–13225 (2022).
30. N. N. Ahmad and F. Proctor, "Review of idealized aircraft wake vortex models," in *52nd Aerospace Sciences Meeting,* (2014), pp. 1–28.
31. T. Gurke and H. Lafferton, "The development of the wake vortices warning system for frankfurt airport: Theory and implementation," Air Traffic Control. Q. **5**, 3–29 (1997).
32. D. C. Burnham and J. N. Hallock, "Chicago monostatic acoustic vortex sensing system," (1982). Report no. DOT-TSC-FAA-79-103.IV.