

Bivariate Probabilistic Multimodel Ensemble Wake-Vortex Prediction employing Bayesian Model Averaging

Stephan Körner*, Frank Holzäpfel† and Ingo Sölch ‡

Deutsches Zentrum für Luft- und Raumfahrt, Oberpfaffenhofen, Germany

This paper investigates the capability of the Multimodel Ensemble approach to improve the deterministic forecast of wake vortex behavior and to produce reliable vortex habitation areas. Therefore the models D2P, APA 3.2, APA 3.4, APA 3.8 and TDP 2.1 were provided within the framework of a NASA-DLR cooperation. In a previous study the Bayesian Model Averaging (BMA) approach, that computes the ensemble forecast as a weighted sum of PDFs, turned out to be promising. For this reason the focus of this paper lies on the further development and assessment of this method. While the previously presented methods did not take into account that the error increases temporally, the new approach considers growing uncertainties. In addition, combined confidence areas for the vertical and lateral vortex position are derived from bivariate probability density distributions. For training and evaluation wake vortex campaigns accomplished by NASA (MEM95, DFW97, DEN03, MEM13) and DLR (WakeMUC, WakeFRA, WakeOP) are employed.

Nomenclature

b	vortex spacing
f_i	forecast
$g_i(y f_i)$	probability density distribution of forecast
L	likelihood function
$p(y)$	probability of forecast y
P_{cut}	probability that corresponds to specific confidence level
s	relative skill improvement
t	time
w_i	weight
y	lateral vortex position
y^T	training data
z	vertical vortex position
$\hat{z}_{i,t}^{(j)}$	unobserved quantity
<i>Greek</i>	
α	risk level
Γ	vortex strength in terms of circulation
μ	distribution mean
Ω_{cut}	confidence contour
σ	standard deviation

Abbreviations

APA	AVOSS (Aircraft Vortex Spacing System) Prediction Algorithm
BMA	Bayesian Model Averaging
DEA	Direct Ensemble Average
D2P	Deterministic Two-Phase wake vortex model
hr	high resolution

*Research Scientist, Institute of Atmospheric Physics, DLR, Oberpfaffenhofen, Germany.

†Senior Research Scientist, Institute of Atmospheric Physics, DLR, Oberpfaffenhofen, Germany.

‡Research Scientist, Institute of Atmospheric Physics, DLR, Oberpfaffenhofen, Germany (until 08/2017)

IGE	In-Ground-Effect
lidar	light detection and ranging
lr	low resolution
MME	Multimodel Ensemble
mr	medium resolution
NGE	Near-Ground-Effect
OGE	Off-Ground-Effect
PDF	Probability Density Function
P2P	Probabilistic Two-Phase wake vortex model
REA	Reliability Ensemble Average
RMS	Root Mean Square
RMSE	Root Mean Square Error
TDP	TASS (Terminal Area Simulation System) Driven Algorithms for Wake Prediction

Subscript

<i>i</i>	model number
<i>I</i>	number of models
<i>j</i>	grid point in vertical direction
(<i>j</i>)	iteration number
<i>k</i>	grid point in lateral direction
<i>m</i>	number of entries in 1D array
<i>n</i>	number of observations
<i>o</i>	observation
<i>p</i>	parameter
0	initial value

Superscript

*	normalized quantity
---	---------------------

I. Introduction

Wake vortices [1], generated at the wings of aircraft as a response to lift, can induce a potentially hazardous rolling moment to any follower. As global air traffic is expected to double in the next 15 years [2] wake vortex forecasts may not only increase the safety in the terminal area but also in-flight [3]. Various fast-time wake vortex models have been developed [4, 5, 6, 7, 8, 9, 10, 11] to achieve a safety gain and potential airport capacity benefits. Depending on the ambient conditions, the forecast quality of the wake vortex models varies due to different underlying physics and due to the data that was available for model development. If the conditions under which one model outperforms the others were known we could increase the prediction skill easily by always choosing the most suitable model. However, varying model performance and measurement data quality for different campaigns as well as inaccurate initial conditions complicate the identification of a best model for each forecast [12]. That is why the combination of various structurally independent models in a Multimodel Ensemble (MME) can improve the forecast skill [12]. In addition, the model spread can be utilized to generate probabilistic envelopes, which take into account uncertainties of the model input parameters and related observed variability of the vortices.

These advantages shall be exploited in this paper by fusing the models APA 3.2, APA 3.4 [5], TDP 2.1 [8, 13] and the newly available APA 3.8 [14] from NASA with the D2P-model from DLR [6, 7, 9]. All models base their forecasts on wind, temperature and turbulence profiles as well as on aircraft data such as wingspan, velocity and altitude. The meteorological data are usually given as 10-minute averages.

In a previous study [15] the evaluation of suitable MME methods demonstrated that the ensemble approach can indeed improve wake vortex forecasts. The Bayesian Model Averaging [16] turned out to achieve better results for our application while being more robust than the Reliability Ensemble Averaging [17, 18] and is thus chosen to be further refined. Compared to Monte-Carlo Simulations the probabilistic BMA-forecast could reach a much higher reliability in reproducing the vortices measured by lidar as demonstrated in Körner et al. [19].

The goal of this paper is not only to further improve the deterministic forecast but also to generate reliable combined vortex habitation areas in space and time based on a vortex-age dependent training and bivariate Probability Density Functions (PDFs). For training purposes and for model evaluation lidar measurements from DLR field experiments at Munich (WakeMUC) [20], Frankfurt (WakeFRA)

[21] and special airport Oberpfaffenhofen (WakeOP) [20] are employed. Further, the dataset provided by NASA has been extended compared to previous studies with measurements from the Dallas (DFW97) [22], Denver (DEN03) [23] and Memphis (MEM95, MEM13) [24, 25] campaigns being utilized. While the DLR data comprises landings In-Ground-Effect (IGE) and Near-Ground-Effect (NGE), the NASA measurements consist to a good part of Off-Ground-Effect (OGE) data. In the following all data are treated non-dimensionally, with the vortex strength Γ normalized by the initial circulation Γ_0 , and the vertical position z as well as the lateral position y normalized by the initial vortex spacing b_0 . Furthermore, time t is normalized by t_0 , which constitutes the time that a vortex needs to descend one b_0 . The normalized time is also termed *vortex age*.

A. Measurement Campaigns

An overview of the vortex generation height z_0 and the initial vortex separation b_0 for the different campaigns is presented in Figure 1. A distribution of the meteorological conditions encountered during the measurements is given in Figure 2. Firstly, the figures demonstrate that the measured traffic mix strongly differs between the campaigns and that a multitude of conditions is discovered at different measurement sites. In comparison to other airports a high percentage of landings take place with tailwind in Memphis. Further, the majority of the measurements at Memphis airport have been accomplished over forest south of the airport which affects wake vortex behavior [25]. In order not to mix different measurement strategies and qualities we decided to split the data into a DLR and NASA dataset in this paper.

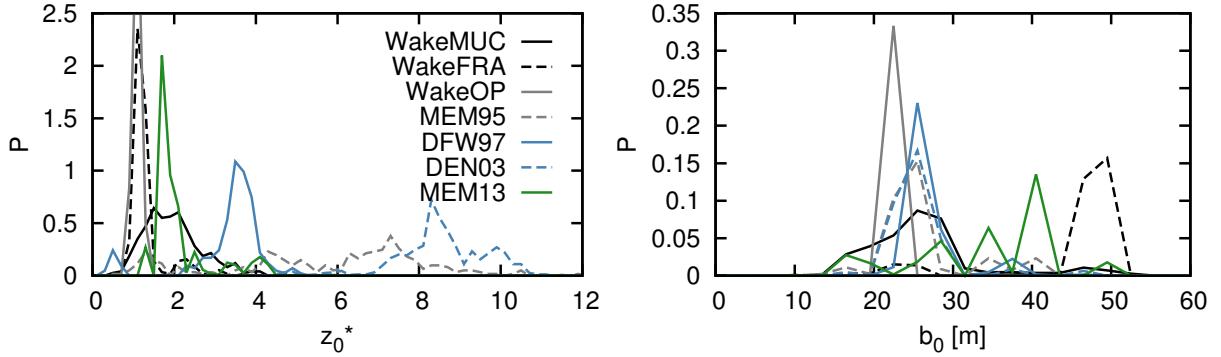


Fig. 1: Distribution of the normalized vortex generation height and initial vortex spacing for the individual wake vortex field measurements.

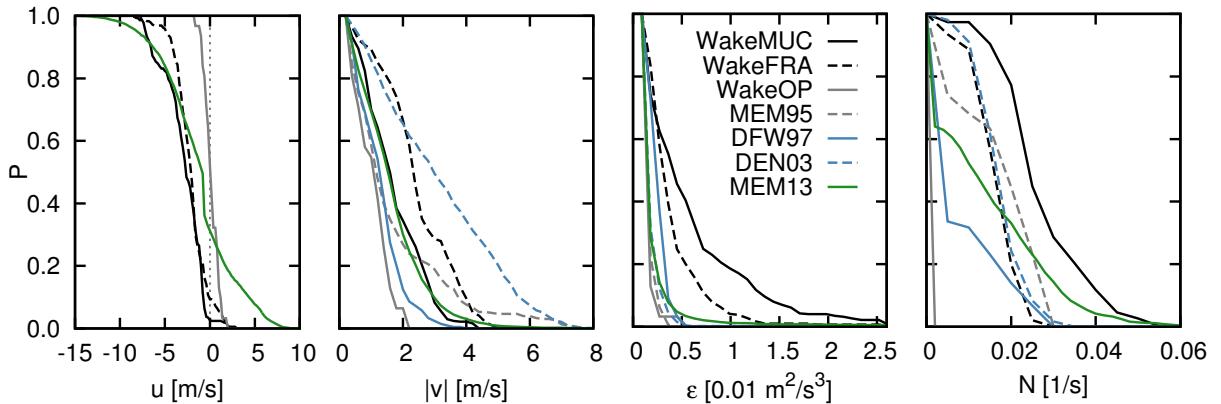


Fig. 2: Distribution of the meteorological conditions for the individual wake vortex field measurements, averaged over the vertical profile.

B. APA 3.8

In contrast to earlier versions APA 3.8 includes more realistic stratification effects, such as detrainment and entrainment from the vortex cell during its descent [14] as well as the effect of baroclinic vorticity on vortex descent speed. Further, the modeling of the rebound has been improved according to the approach of Holzapfel and Steen [9]. In addition, the effect of the ground on the strength of the secondary vorticity, triggering vortex rebound, has been investigated and implemented by means of a ground effect factor [14].

The other models that are utilized for the here presented ensemble have already been briefly discussed in Körner et al. [19].

II. Bayesian Model Averaging (BMA)

Model forecasts are associated with uncertainties with respect to initial conditions and model physics. Further uncertainty is ascribed to the assessment of the model performance against measurement data which itself is subject to measurement uncertainties. Calculating deviation uncertainty allowances is one way to handle them. Another way is to include the uncertainty *a priori*, which constitutes the essence of the Bayesian Model Averaging [16]. This approach quantifies the prediction uncertainties by dressing each model forecast with a Probability Density Function (PDF). An ensemble forecast can then be generated by weighting the PDFs according to the respective model performance and summing them up. The BMA calculates the ensemble forecast based on the probability w_i that one of the forecasts f_1, \dots, f_I of model i is best, and on the probability $g_i(y|f_i)$ (represented by a PDF) that a distinct accuracy level is achieved if it is best.

$$p(y|f_1, \dots, f_I) = \sum_{i=1}^I w_i g_i(y|f_i), \quad (1)$$

Conclusively, the ensemble constitutes a weighted sum of PDFs [16] where the individual model PDFs $g_i(y|f_i)$ are centered at the respective deterministic model forecast. From the resulting ensemble PDF both a deterministic forecast (= ensemble PDF mean) and a probabilistic envelope (= percentiles for a given probability level) can be determined. The percentiles are computed by integrating the PDF equally from the tails. The BMA has already been applied to wake vortex forecasts previously and is described in more detail in Körner et al. [19].

A. Maximum-likelihood estimation of ensemble parameters

To determine w_i and the standard deviation σ of the PDF g_i it is useful to employ a maximum likelihood method [26] which combines various PDFs, such that their sum matches the overall PDF best. Such an approach is also called finite mixture modeling [27]. The maximum-likelihood estimation approach is based on the idea that only a random sample from an unknown population is available [26]. The goal of this method is to estimate the parameters of the PDF of a population that has most likely generated the sample [26]. In the case of the ensemble forecast the population is given by the model errors for a training sample and the likelihood is determined by

$$\log(L(w_i, f_i, \sigma)) = \sum_{o=1}^n \log \left(\sum_{i=1}^I w_i \cdot g_i(y_o|f_{i,o}, \sigma) \right). \quad (2)$$

The logarithm is applied such that the likelihood function can be expressed by sums instead of products to simplify the calculation. As an analytical maximization is not possible here the Expectation-Maximization (EM) algorithm [28, 29] is utilized to find the optimal ensemble properties from the training dataset [16]. It determines the parameters of a weighted sum of PDFs that have most likely produced the available distribution iteratively. Its application is described in more detail in Raftery et al. [16]. In Körner et al. [19] we showed that the error distributions of the employed wake vortex models are leptokurtic. For this reason four Gaussian distributions are utilized as baseline PDFs to model the correct error distribution based on finite mixture modeling [19, 27].

B. Temporal development of model uncertainty

In the original approach [16] the uncertainty is assumed to be constant over time. However, Figure 3 clearly indicates that the error is increasing with advancing vortex age, regardless of which model and parameter is evaluated. While the steepest error increase occurs below one t_0 in most cases, the growth progresses more slowly afterwards. Qualitatively, the error development is similar among all models, featuring an increase for all parameters. Nevertheless, the gradients of the fits differ. In this plot the error at $t \approx 0$ represents the initial condition uncertainty and is as expected the same for all models.

Given these results, constant standard deviations are a simplification, leading to conservative envelopes for early and to non-conservative bounds for later vortex ages. Note, that in the approach with constant parameters the impact of the early conservative and the late non-conservative envelopes may compensate each other such that on paper the probabilistic results look good on average.

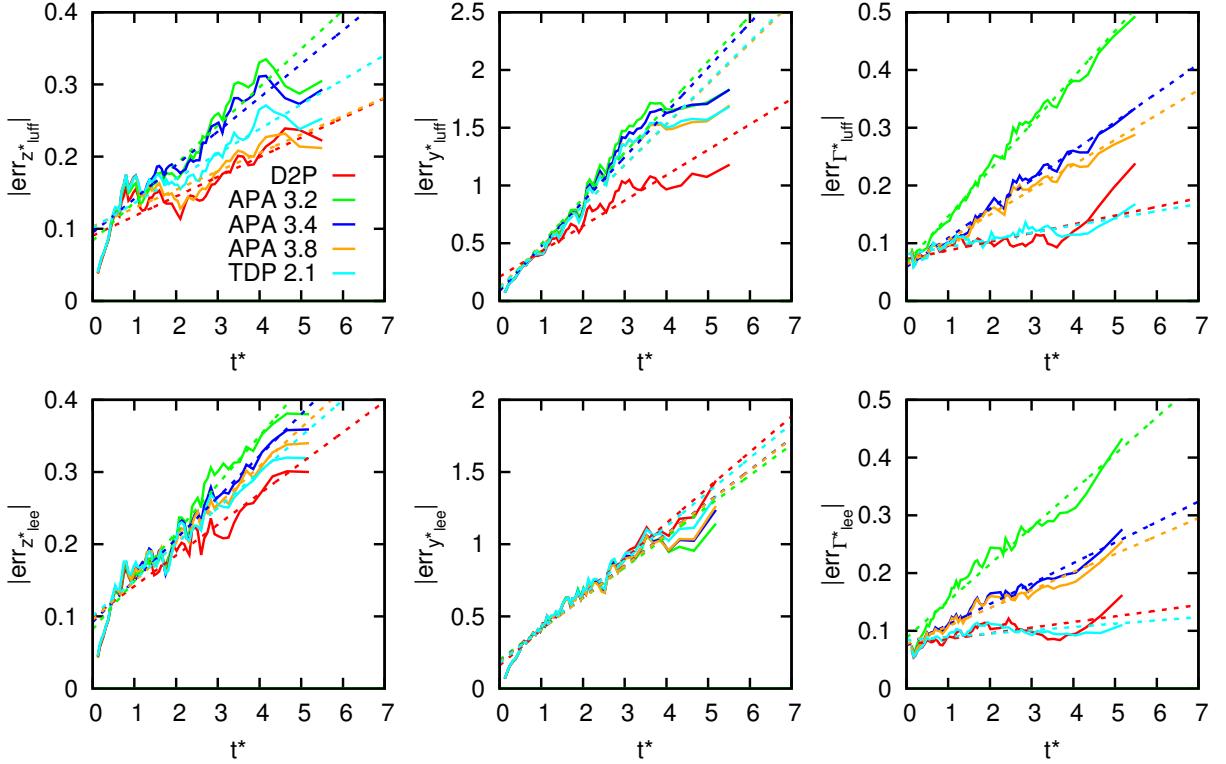


Fig. 3: Development of the model uncertainty in terms of error magnitude against vortex age for the WakeMUC, WakeFRA and WakeOP training dataset that consists of 335 landings.

The temporal increase in model error can be considered by applying the maximum likelihood method described before to vortex age bins. To avoid overfitting of the ensemble to the training data both the weights and the standard deviations are approximated linearly instead of applying interpolation to each vortex age. Due to the dependence of w_i and σ (see Körner et al. [19]), both parameters must be treated vortex-age-dependently.

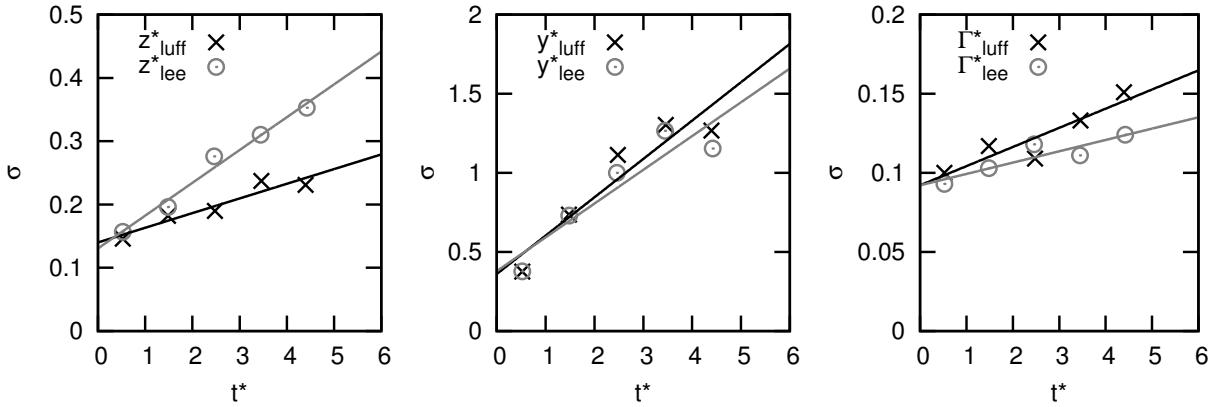


Fig. 4: Development of σ if computed by maximum likelihood based on the test dataset, consisting of 335 landings from WakeMUC, WakeFRA and WakeOP.

Figure 4 displays the standard deviations derived in this manner for vortex ages below $5 t_0$, for which the data coverage is sufficient. While σ clearly increases linearly for z^* and Γ^* it appears that for y^* a square fit would be more suitable for this dataset. However, a linear approximation is expected to be the best compromise as the order of the best fit deviates between different datasets. It is worth to mention that $\sigma_{z^*,luff}$ could be significantly reduced by implementing APA 3.8 in the ensemble, indicating that this model exhibits increased skill compared to its earlier versions. This trend can also be noted for σ_{y^*} , but is missing for $\sigma_{z^*,lee}$ and Γ^* .

Figure 5 depicts the corresponding model weights. While a clear trend can be identified for the

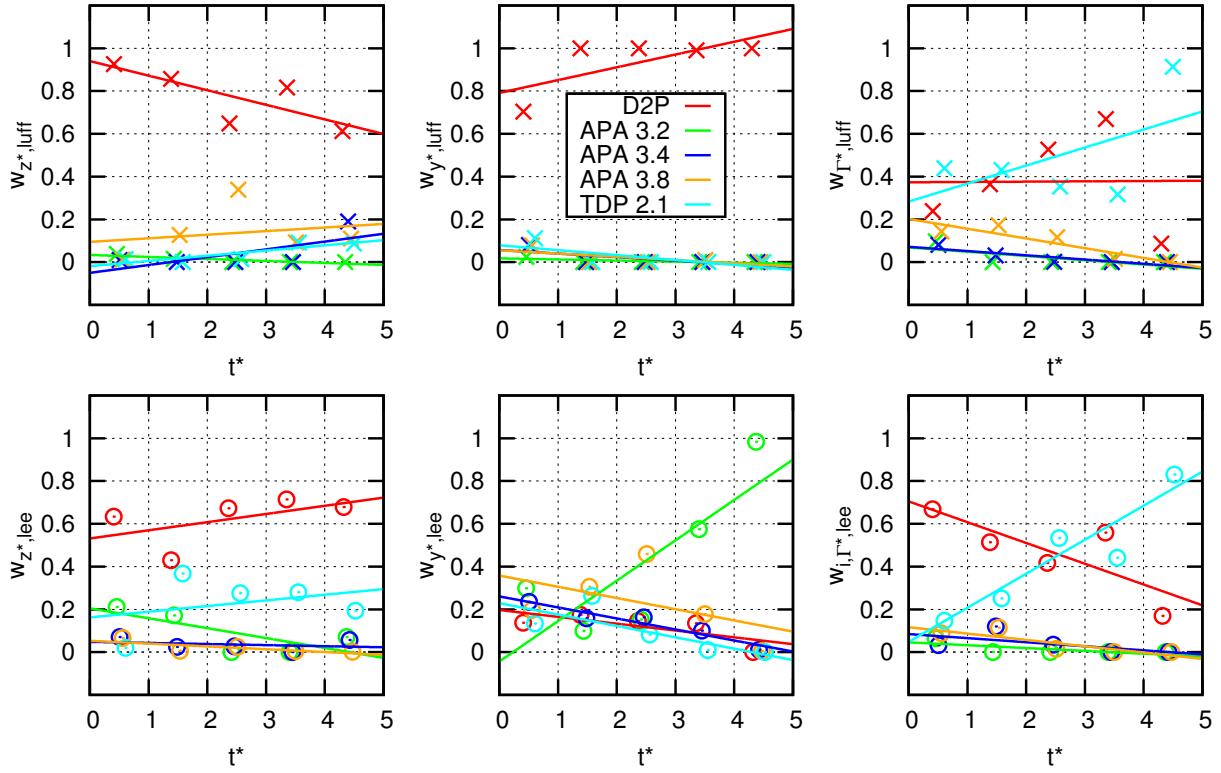


Fig. 5: Development of w_i if computed by maximum likelihood based on the test dataset, consisting of 335 landings from WakeMUC, WakeFRA and WakeOP.

z^* weight of D2P, it remains less clear for the other models and parameters. Nevertheless, for the sake of simplicity and robustness a linear approximation is employed again. Note, that the constraints $\sigma_i w_i > 0$ and $\sum_i w_i = 1$ must be satisfied when applying the parameters eventually. Note that also the shapes of the error distributions evolve with vortex age. Both gaussian and leptokurtic PDFs may occur during different stages of wake vortex lifetime. We observe that the PDFs for early vortex ages are more leptokurtic while they turn more gaussian towards the end of the vortex lifetime. As a consequence the approximation of the model deviation PDFs by the EM algorithm must be applied vortex-age-binned as well. The trend of the thereof derived parameters is then approximated linearly again.

In addition, the training is applied to two different initial altitude ranges as both the model uncertainty and performance varies in IGE/NGE and OGE. Thus the first interval ranges from 0 to 2.5 z_0 while the second comprises all initial altitudes above.

C. Performance

Before the ensemble is applied the dataset is split into a training and a test sample. To provide realistic operational conditions the training sample contains landings from the first part of each of the campaigns whereas the test sample consists of the second part. Fig. 6 illustrates a BMA forecast with both the deterministic mean and the probabilistic envelopes for the 90, 95 and 99% uncertainty level. As the ensemble parameters have been computed separately for luff and lee, the envelopes may differ for both vortices. While the bounds of the lee vortex are wider compared to the luff vortex for the z^* forecast, it is vice versa for the y^* prediction. This can be explained by the crosswind that largely affects the rebound of the lee vortex (strong crosswind, high rebound) [9, 20]. For this reason uncertainties in the wind measurements are directly transferred to the forecast of the vertical position. In contrast, the rebound of the luff vortex is suppressed with increasing crosswind speeds [9] which decreases its sensitivity. Due to its lower rebound the vortex ground interaction regarding lateral divergence is stronger. At this point additional uncertainty about the physics seems to be introduced, such that the uncertainty envelope for the lateral position of the luff vortex is wider.

After being trained with a training sample consisting of meteorological input data, aircraft data and lidar measurements, the ensemble is applied to a set of test data for evaluation purposes. Table 1 lists the median RMS deviations between deterministic prediction and measurement data for the model and its members if applied to the DLR test dataset (359 landings). In addition to the BMA the Direct Ensemble Average (DEA), a simple arithmetic average of the forecasts, is employed as a baseline. Likewise Table

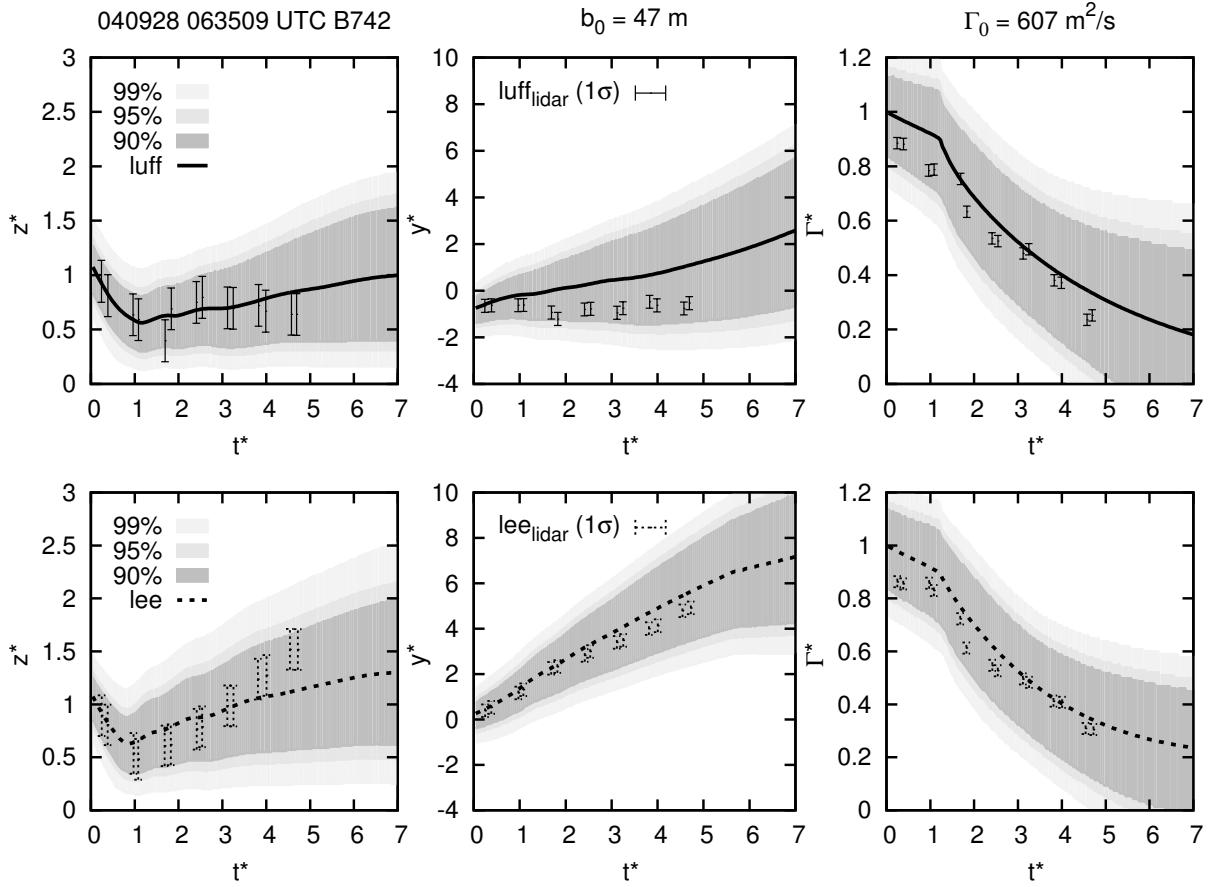


Fig. 6: BMA forecast with temporal dependency of σ and w_i . The predictions are given separately for each parameter.

2 lists the performance for the quality controled NASA dataset (1572 landings). The training data comprised 335 (DLR) and 2905 (NASA) landings, respectively.

The results indicate that the BMA method is indeed capable of achieving the smallest RMS deviation for the y and Γ forecasts if applied to the DLR data. Further, the DEA can only enhance the performance for the z_{luff} -forecast, where the BMA reaches the same RMSE as D2P, the best model in this case. For all other parameters the BMA is superior, although it only achieves the second best performance for the z_{lee} -forecast. It is interesting that the ensemble is capable of improving the forecast of the lateral position for this dataset. The fact that the lateral transport is governed by the crosswind to a large part would imply that the model forecasts do not deviate much from each other. However, it seems that in ground proximity the differences in parametrization are large enough to achieve a widening of the solution space.

For the NASA data the BMA outperforms the DEA with exception of the z_{luff} -forecast, where it achieve the same RMSE. From the results for the DLR data one could conclude that D2P would be the best choice if a single model had to be picked. However, it turns out that the NASA models are superior for the NASA dataset. As a consequence, the BMA reaches smaller RMS deviations than D2P for all parameters. Further, the BMA accomplishes the smallest RMS deviations for the Γ - and z_{lee} -forecast compared to the respective best models. By contrast the DEA does not achieve smaller RMS deviations than the best individual model for any parameter.

To evaluate the skill improvement by the ensemble in percent for all model parameters rather than for a single one, we introduce the relative skill s . It calculates as the average RMSE relation between ensemble and model i for all parameters p . Negative skill values indicate that the individual model performs worse than the ensemble with regard to the RMSE of all parameters.

$$s_i = \left(\frac{1}{6} \sum_{p=1}^6 \frac{\text{RMSE}_{\text{BMA},p}}{\text{RMSE}_{i,p}} - 1 \right) \cdot 100 \quad (3)$$

Table 3 lists the relative skill computed in this manner. It reveals that skill improvements of 4.3% for the DLR and 3.1% for the NASA data in comparison to the respective best model can be reached. D2P

Table 1: Median RMS deviations between deterministic prediction and measurement data for the ensemble and its members for the WakeMUC, WakeFRA, WakeOP test sample (359 landings).

model	RMSE Γ^*		RMSE y^*		RMSE z^*	
	luff	lee	luff	lee	luff	lee
BMA	0.099 ^a	0.094 ^a	0.588 ^a	0.510 ^a	0.161	0.180
DEA	0.174	0.146	0.819	0.520	0.187	0.178 ^a
TDP 2.1	0.108	0.102	0.788	0.542	0.198	0.196
APA 3.8	0.154	0.133	0.799	0.525	0.169	0.202
APA 3.4	0.158	0.142	0.814	0.525	0.205	0.209
APA 3.2	0.248	0.218	0.905	0.513	0.208	0.223
D2P	0.115	0.104	0.599	0.529	0.156 ^a	0.180

^a smallest RMSE.

Table 2: Median RMS deviations between deterministic prediction and measurement data for the ensemble and its members for the DFW97, MEM95, DEN06, MEM13 test sample (1572 landings).

model	RMSE Γ^*		RMSE y^*		RMSE z^*	
	luff	lee	luff	lee	luff	lee
BMA	0.166 ^a	0.164 ^a	0.999	1.002	0.228	0.227 ^a
DEA	0.178	0.174	1.020	1.022	0.228	0.231
TDP 2.1	0.213	0.207	1.025	1.028	0.227	0.229
APA 3.8	0.176	0.173	1.020	1.029	0.232	0.229
APA 3.4	0.172	0.169	1.008	1.017	0.244	0.245
APA 3.2	0.170	0.168	0.993 ^a	0.992 ^a	0.247	0.249
D2P	0.195	0.191	1.085	1.044	0.223 ^a	0.237

^a smallest RMSE.

is even outperformed by 7% for the NASA data, although it would have been chosen as the best model based on its performance for the DLR data. This underlines the difficulty to identify a single model that is suitable for all kinds of conditions. Further, the results indicate that overall improvements of over 30% are possible (APA 3.2 for the DLR dataset). Compared to the DEA approach the BMA is by 19.1% (DLR data) and 3.0% (NASA data) superior. The large difference between those numbers might be due to the model performance not varying that much amongst each other in the NASA data with the RMS deviations being more alike. To give proof for the robustness of the ensemble approach the DLR training and test dataset have been switched (DLR1), or newly generated by mixing the dataset (DLR2 and DLR3) in additional runs. It turns out that the average RMSE improvement compared to D2P in these three runs is between 2.8% and 4.8% with an average of 4.1%.

Table 3: Skill improvement for the DLR (359 landings) and NASA (1572 landings) test sample.

model	relative skill s [%]				
	DLR	NASA	DLR1	DLR2	DLR3
BMA	0.0	0.0	0.0	0.0	0.0
DEA	-19.1	-3.0	-	-	-
TDP 2.1	-12.5	-8.2	-12.3	-10.3	-11.7
APA 3.8	-18.4	-3.1	-17.9	-17.5	-18.2
APA 3.4	-22.9	-3.8	-22.3	-20.1	-22.2
APA 3.2	-32.5	-3.3	-29.5	-28.9	-30.9
D2P	-4.3	-7.0	-4.8	-2.8	-4.6

In practice the position of the vortices is considered as a coordinate in the z - y -space. For this reason we evaluate the RMS deviation of the position, including both luff and lee, for each model i according to

$$\text{RMSE}_{i,(z,y)^*} = \sqrt{(z_i^* - z_o^*)^2 + (y_i^* - y_o^*)^2}. \quad (4)$$

Figure 7 illustrates the RMS deviation of the predicted from the measured vortex position against the RMSE of Γ for both the DLR and NASA dataset. On the one hand the results demonstrate the large difference between the DLR and NASA dataset, on the other hand the increased model performance differences for the DLR data are revealed. The skill improvement of the position prediction in the DLR data amounts to 5.9% if compared to the best performing model D2P. This can mainly be accounted to the improved y -forecast (see Table 1). In contrast, the position prediction for the NASA data is by 0.1% inferior to the forecast of APA 3.2. However, with D2P as reference the skill enhancement amounts to 8.4%. The Γ -forecast could be improved by 11.3% (DLR data) and 2.3% (NASA data) with respect to the best performing model. With D2P as reference the Γ -RMSE can be reduced by 13.5% in the NASA data.

In a next step the ensemble size is reduced by removing the worst performing models for the DLR test dataset. First APA 3.2 and APA 3.4 are removed (BMA_{1,4,5}), then TDP 2.1 is extracted additionally (BMA_{1,5}). In both cases the RMS deviations for both position and circulation are increased in comparison to the run with the full ensemble (see Figure 7). This reveals that even the models with the least skill may improve the ensemble forecast. If the same strategy is applied to the NASA test dataset the RMSE increase turns out to be less significant, which might be due to the skill enhancements being smaller than in the DLR data anyway.

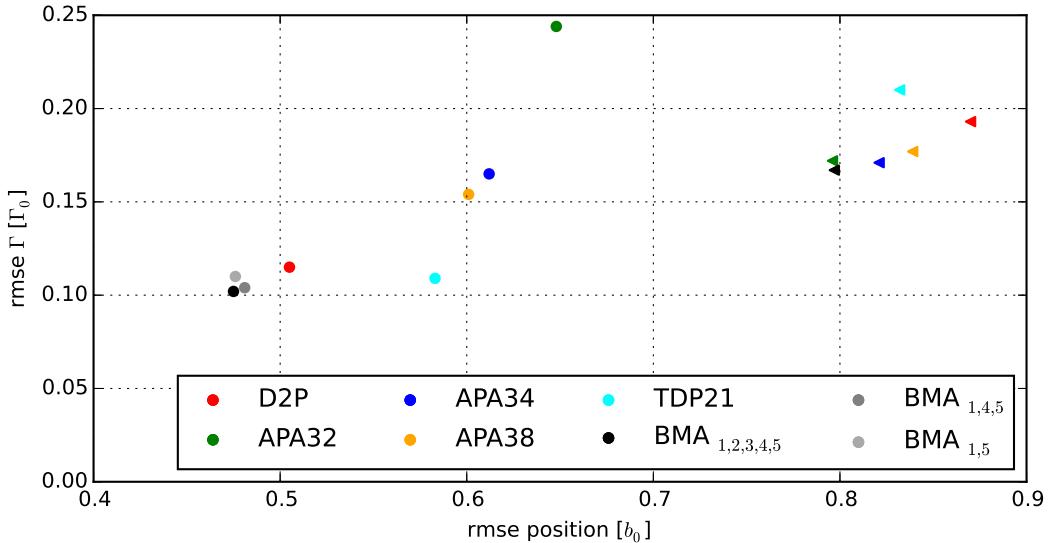


Fig. 7: Scoring of the ensemble if applied to both the DLR and NASA test data. Further, the ensemble size is reduced for the DLR data in to two further runs. Dots denote results for DLR data, triangles for NASA data.

While the importance of the improvement of the deterministic forecast performance is undoubtedly, it is the probabilistic forecast skill that matters in practice. Because the natural variability of wake vortices and the stochastic nature of the atmosphere cannot be predicted by deterministic means, it is required to provide uncertainty envelopes for the projections. Their good performance has already been demonstrated in Körner et al. [19], with the only difference that the approach in this paper achieves excellent results not only temporally averaged but for any vortex age. Therefore the focus of this paper shall rather be laid on the development of an approach to fuse the probabilistic envelopes of the z - and y -forecast to generate vortex habitation areas.

III. Combined Confidence Areas for Vortex Position

The so far introduced envelopes represent the probability of encountering a single vortex in z - or y -direction individually. However, in practice a confidence area that quantifies the probability to encounter any of the two vortices in a 2-dimensional prediction plane at a specific time is required. Previously, such confidence areas have been generated by offline Monte-Carlo Simulations and fitted to elliptic shape in the airborne version of P2P [30]. However, the BMA output consists of PDFs, such that it is straightforward

to compute bivariate distributions here instead. Then, nominal forecast probabilities may be defined to minimize the risk of a vortex encounter below a certain threshold (see Subsection A). As for the one-dimensional envelopes [19], we expect a good agreement between nominal and observed forecast probability with this approach.

In a first step the joint probabilities for z_{luff}^* and y_{luff}^* , as well as for z_{lee}^* and y_{lee}^* are calculated. In the design of the one-dimensional BMA PDFs, the lateral and vertical vortex position are assumed to be independent and, therefore, the probability to find the luff or lee vortex at a certain position is given by [31]

$$P_{luff} = P(z_{luff}^* \cap y_{luff}^*) = P(z_{luff}^*) \cdot P(y_{luff}^*) \quad (5)$$

and

$$P_{lee} = P(z_{lee}^* \cap y_{lee}^*) = P(z_{lee}^*) \cdot P(y_{lee}^*). \quad (6)$$

In general, z_{luff}^* and y_{luff}^* maybe dependent, for example in the case of strong wind shear. Equations (5) and (6) result in separate joint (bivariate) PDFs for the luff and lee vortex as depicted in Figure 8a and Figure 8b. The normalizations $\iint P_{luff} dz^* dy^* = 1$ and $\iint P_{lee} dz^* dy^* = 1$ remain. In a next step the final PDF to encounter any of the two vortices at a time is generated by [31]

$$P = P(\text{luff} \cup \text{lee}) = (P_{luff} + P_{lee}) / 2. \quad (7)$$

This can only be applied to mutually exclusive events ("or"-conjunction) and thus neglects the possibility that both vortices can be encountered at the same time. Again $\iint P dz^* dy^* = 1$ is guaranteed. Figure 9 illustrates the overall joint PDF that now exhibits two peaks. The contours are tapered between both vortices and may even be separated at lower confidence levels. It is striking that the confidence in the prediction of the vertical position is much higher than in the lateral position, which can be attributed to a large part to the uncertainty in the wind measurements.

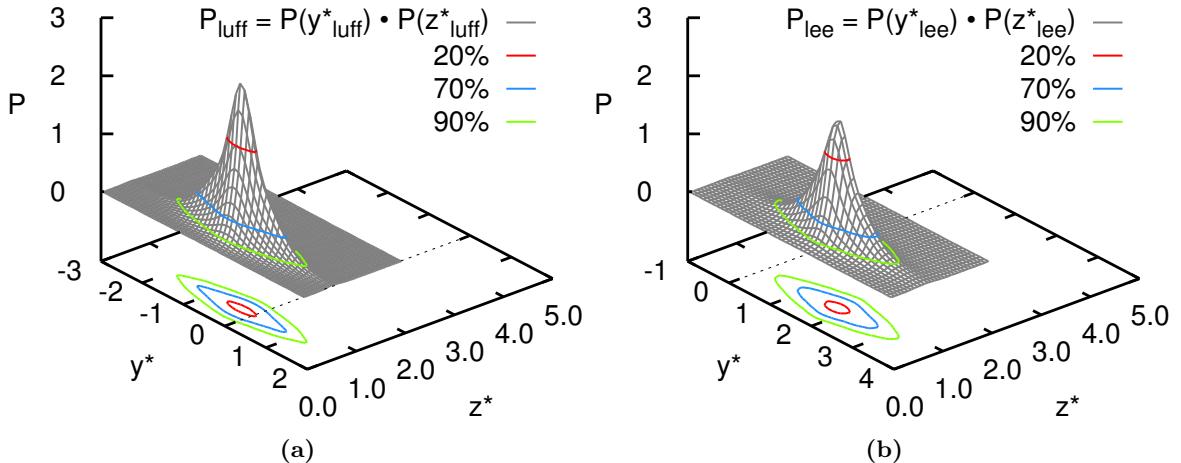


Fig. 8: Joint PDFs for the luff (a), and the lee vortex (b) for vortices generated in IGE.

A. Probability levels

To define the vortex habitation areas from the overall joint PDF a 2-dimensional area Ω_{cut} is searched which consistently minimizes the risk of a vortex encounter below a certain threshold α (usually 1%). Therefore only regions of P are considered with a probability greater than a limiting level P_{cut} (see Figure 10). Formally, P_{cut} can be derived from

$$\iint_{\Omega_{cut}} P(z^*, y^*) dz^* dy^* = (1 - \alpha) \quad \text{with} \quad \Omega_{cut} = \{z^*, y^* \mid P(z^*, y^*) \approx P_{cut}\}. \quad (8)$$

In order to calculate P_{cut} efficiently, the bivariate distribution is discretized on a 2-dimensional grid with spacing Δy^* and Δz^* . Defining a linear array of probability levels the distribution is integrated for decreasing probabilities, until Equation (8) is fulfilled. Then P_{cut} defines that iso-surface of the overall joint PDF, which encloses the area for which the probability to encounter one of the vortices is $(1 - \alpha)$ (see Figure 9).

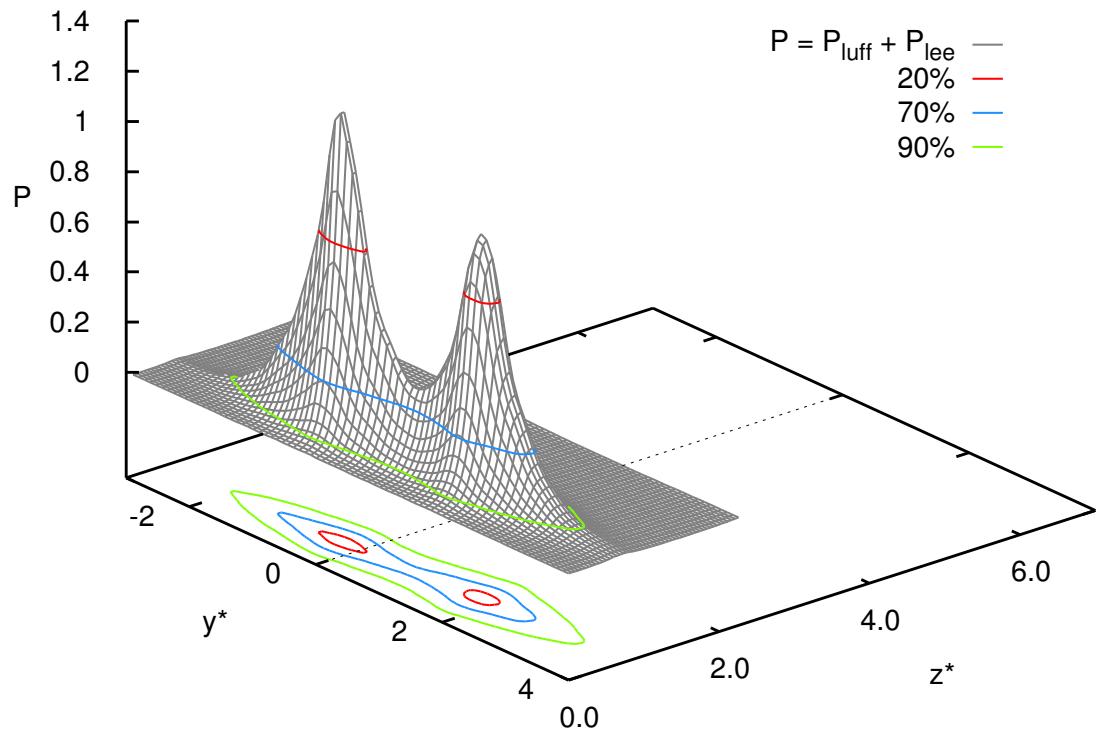


Fig. 9: Overall joint PDF, calculated as the sum of the joint PDFs of the luff and lee vortex.

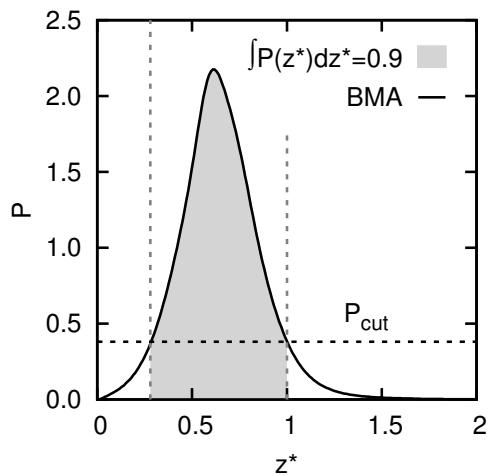


Fig. 10: Principle of confidence interval calculation for 1D example for a probability of 90%.

B. Results

The confidence interval for the discussed example is depicted in Figure 11 for $t^* = 1.3$. In this case the corresponding lidar measurements both lie within the 20% envelope close to its borders and thus also clearly within the 70% and 90% bounds. The shape of the confidence intervals is both non-elliptic and asymmetrical with less uncertainty being contributed by the luff vortex forecast. Note, that this is especially interesting because the luff vortex is the operationally more relevant vortex, hovering in the glide path and endangering following aircraft.

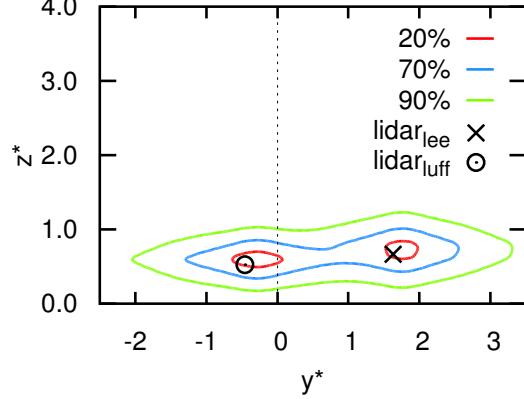


Fig. 11: Vortex habitation areas for the 20%, 70% and 95% confidence levels.

The effectiveness of the approach is demonstrated by the reliability plot in Figure 12a that depicts the percentage of measurements that lie inside of the vortex habitation area for a given confidence level. The performance if the ensemble is applied to the training data shall demonstrate how well the training performs itself. For the DLR data the deviation from the optimum (diagonal line) does not exceed 2% for both the training and the test data. While the performance for the 99%, 95% and 90% envelope achieves equally good results for the NASA data, the deviations can be as high as 5% for probabilities below 90%. This trend cannot simply be accounted to the difference between test and training data as it occurs in both datasets. However, Figure 12b indicates that the confidence areas derived from the joint probabilities for the luff and lee vortex individually show very good accordance with the measured probability, which suggests that the observed deviation from the optimum is introduced when adding the probabilities according to Equation (7). Although independence between the random variables that are to be joined is only required for the combination of z and y according to Equation (5) and (6), the linear error correlations in Table 4 suggest a connection. While the correlation between the errors of z_{luff}^* and y_{luff}^* as well as between z_{lee}^* and y_{lee}^* are small, the errors of the position coordinates between luff and lee are significant in the NASA dataset only. Nevertheless, this discrepancy cannot fully be resolved. In this context it is important to note that firstly in practice probability levels below 90% are not relevant, and secondly that the discussed deviations are towards the conservative direction. To produce the here discussed results a PDF resolution of $\Delta z^* = 0.01$, $\Delta y^* = 0.05$ and $\Delta \Gamma^* = 0.01$ was applied. However, envelopes with similar accuracy could also be achieved with half or quarter the resolution, given envelopes that are more angular shaped are acceptable.

Table 4: Correlations between the ensemble errors for y and z and the luff and lee position.

dataset	z_{luff}^* and y_{luff}^*	z_{lee}^* and y_{lee}^*	$(z, y)_{\text{luff}}^*$ and $(z, y)_{\text{lee}}$
DLR	0.17	0.04	-0.002
NASA	0.05	-0.02	0.71

C. Runtime

The drawback of this approach is the increased runtime due to the described integration process. However, decreasing the number of confidence levels (1 instead of 5), with the 99% being the most relevant for application, can reduce the simulation time. Although the BMA code has been runtime optimized in parts there is still room for improvement concerning in- and output operations. Table 5 lists the runtime of the individual models and the whole ensemble for different PDF resolutions if only the 99% envelopes are computed on a local Linux PC on a single core with 3.3 GHz with 10.3 GB of usable memory. The

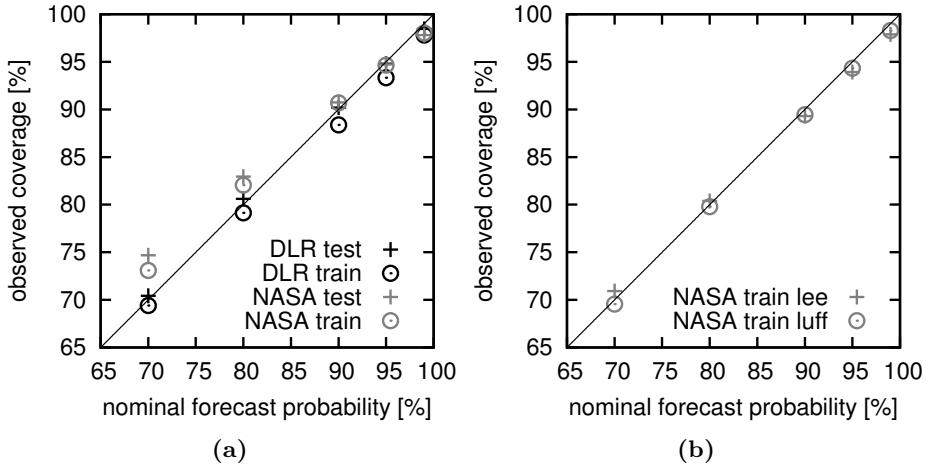


Fig. 12: Reliability of the combined confidence areas for the training and the test dataset (a) for both the DLR and NASA data as well as the reliability for the NASA data prior to merging of the luff and lee forecast (b).

runtimes indicate that the processing of the forecasts, and in particular the generation of the joint distributions, takes the largest part of the total time rather than the individual model predictions. For the in this paper applied high resolution (hr) ($\Delta z^*=0.01$, $\Delta y^*=0.05$, $\Delta \Gamma^*=0.01$) the ensemble requires 4.1 s for the prediction of an A340 landing. If the resolution is decreased to medium resolution (mr) ($\Delta z^*=0.02$, $\Delta y^*=0.1$, $\Delta \Gamma^*=0.02$) the runtime is enhanced to 2.6 s. With a further decrease to low resolution (lr) ($\Delta z^*=0.04$, $\Delta y^*=0.2$, $\Delta \Gamma^*=0.04$) a runtime of 1.9 s can be achieved.

Table 5: Runtime of the ensemble and its members for a single WakeFRA case (A340 aircraft) in IGE.

model	runtime [s]
TDP 2.1	0.009
APA 3.8	0.007
APA 3.4	0.004
APA 3.2	0.004
D2P	0.006
BMA (hr)	4.081
BMA (mr)	2.621
BMA (lr)	1.905

IV. Conclusion

This paper investigates the capability of a Multimodel Ensemble (MME) to improve the deterministic forecast of wake vortex behavior and to generate reliable confidence areas for the wake vortex position. The ensemble is based on wake vortex models that have been exchanged in the context of a NASA-DLR cooperation and comprise D2P, APA 3.2, APA 3.4, APA 3.8 and TDP 2.1. In a previous study [15], two different MME approaches have been applied which comprised the Reliability Ensemble Averaging (REA) [17] and the Bayesian Model Averaging (BMA) [16]. As the latter turned out to be superior for this application, this paper concentrates on the further development and assessment of the BMA method. This approach computes the ensemble forecast as a weighted sum of PDFs, with the weights and standard deviations being determined by maximum likelihood estimation in a training phase [16]. While the previously presented approaches did not take into account that the error increases temporally, the new approach considers the growing uncertainty. For training and evaluation wake vortex campaigns accomplished by NASA (DFW97, MEM95, DEN03, MEM13) and DLR (WakeMUC, WakeFRA, WakeOP) are employed. Both the training dataset and the sample for evaluation are of roughly equal size.

It turns out that the ensemble approach may improve the deterministic forecast, in comparison to

the respective best model, by 4.3% if evaluated for the DLR and 3.1% for the NASA test datasets. If not the best model but D2P is taken as reference for the NASA data, the improvement even amounts 7%. We also show that these results are robust as the average improvement for 3 differently composed DLR test datasets are on average 4.1%. Furthermore, it is revealed that the Direct Ensemble Average (DEA) as simple arithmetic ensemble average is inferior to the BMA. If the position rather than the vertical and lateral position individually are evaluated, the skill improvement relative to the best model amounts to 5.9% for the DLR data, while no improvement can be achieved in the NASA data. Further, we reduce the ensemble size in two steps. The results indicate that indeed even the models with the least forecast accuracy among the ensemble may add skill.

It is demonstrated that combined confidence areas of arbitrary shape for the vortex position can be generated that quantify the probability to encounter one of the two vortices. The contours of these areas are derived from bivariate PDFs that are generated by combining the z_{luff}^* and y_{luff}^* as well as the z_{lee}^* and y_{lee}^* forecast with each other in a first step. In a second step the bivariate PDFs of the luff and lee vortex are merged. The probabilistic assessment shows that excellent probabilistic reliabilities can be achieved by this approach. Compared to simple elliptic bounds, the shapes achieved by the presented approach are more accurate and thus allow the probabilistic predictions to be less overconservative. At this point the runtime is yet too large for operational use which is mainly due to the 5 probability levels that are evaluated in this study. Reducing them to one level enhances the simulation time to 4.1 s for high, to 2.6 s for medium and to 1.9 s for low PDF resolution. Nevertheless, there is further room for runtime optimization.

For the future further value could be added by deriving the uncertainties of the forecasts not only from lidar measurements, but also from the initial condition uncertainties. This would open the door to onboard wake vortex forecasts as tested in the WEAA [30] campaigns.

Acknowledgments

We thank Nash'at Ahmad, Randal VanValkenburg and Fanny Limon Duparcmeur from the NASA Langley Research Center for their continuous support and the provision of the NASA wake vortex models and measurements.

References

- [1] Gerz, T., Holzäpfel, F., and Darracq, D., "Commercial aircraft wake vortices," *Progress in Aerospace Sciences*, Vol. 38, No. 3, 2002, pp. 181–208.
- [2] Airbus S.A.S., "Flying by Numbers," research report D14029463 issue 2, Airbus, Blagnac Cedex, France, 2015.
- [3] Bauer, T., Vechtel, D., Abdelmoula, F., and Immisch, T., "In-Flight Wake Encounter Prediction with the Wake Encounter Avoidance and Advisory System," *AIAA Atmospheric and Space Environments Conference*, June 2014, doi:10.2514/6.2014-2333.
- [4] Sarpkaya, T. et al., "Wake-Vortex Eddy-Dissipation Model Predictions Compared with Observations," *Journal of Aircraft*, Vol. 38, 2001, pp. 687–692.
- [5] Robins, R. E. and Delisi, D. P., "NWRA AVOSS Wake Vortex Prediction Algorithm Version 3.1.1," research report NASA/CR-2002-211746, Northwest Research Associates, Inc., Bellevue, Washington, USA, 2002.
- [6] Holzäpfel, F., "Probabilistic Two-Phase Aircraft Wake-Vortex Decay and Transport Model," *Journal of Aircraft*, Vol. 40, 2003, pp. 323–331.
- [7] Holzäpfel, F., "Probabilistic Two-Phase Aircraft Wake-Vortex Model: Further Development and Assessment," *AIAA Journal*, Vol. 43, No.3, 2006, pp. 700–708.
- [8] Proctor, F. H., Hamilton, D. W., and Switzer, G. F., "TASS driven algorithms for wake prediction," *44th AIAA Aerospace Sciences Meeting and Exhibit, AIAA*, , No. 2006-1073, 2006.
- [9] Holzäpfel, F. and Steen, M., "Aircraft Wake-Vortex Evolution in Ground Proximity: Analysis and Parameterization," *AIAA Journal*, Vol. 45, 2007, pp. 218–227.
- [10] De Visscher, I. et al., "Aircraft Vortices in Stably stratified and Weakly Turbulent Atmospheres: Simulation and Modeling," *Journal of Aircraft*, Vol. 51, No. 3, 2013, pp. 551–566.
- [11] De Visscher, I. et al., "Fast Time Modeling of Ground Effects on Wake Vortex Transport and Decay," *Journal of Aircraft*, Vol. 50, No. 5, 2013, pp. 1514–1525.
- [12] Hagedorn, R. et al., "The rationale behind the success of multi-model ensembles in seasonal forecasting - I. Basic concept," *Tellus*, Vol. 57 A, 2005, pp. 219–233.
- [13] Proctor, F. H. and Hamilton, D. W., "Evaluation of Fast-Time Wake Vortex Prediction Models," *AIAA paper*, Vol. 2009-344, 2009.
- [14] Delisi, D. P., Robins, R., and Pruis, M. J., "APA3. 8 Fast-Time, Numerical Wake Model Description and Validation," *8th AIAA Atmospheric and Space Environments Conference*, 2016, p. 3437.
- [15] Körner, S., Ahmad, N., Holzäpfel, F., and VanValkenburg, R. L., "Multi-Model Ensemble Wake Vortex Prediction," *7th AIAA Atmospheric and Space Environments Conference*, 2015, p. 3173.
- [16] Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M., "Using Bayesian Model Averaging to Calibrate Forecast Ensembles," *Monthly Weather Review*, Vol. 133, 2005, pp. 1155–1174.

- [17] Giorgi, F. and Mearns, L. O., "Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the "Reliability Ensemble Averaging" (REA) method," *Journal of Climate*, Vol. 15, 2002, pp. 1141–1158.
- [18] Körner, S. and Holzäpfel, F., "Multi-model ensemble wake vortex prediction," *Aircraft Engineering and Aerospace Technology*, Vol. 88, No. 2, 2016, pp. 331–340, doi:10.1108/AEAT-02-2015-0068.
- [19] Körner, S., Nash'at, N. A., Holzäpfel, F., and VanValkenburg, R. L., "Multimodel Ensemble Methods for Prediction of Wake-Vortex Transport and Decay," *Journal of Aircraft*, 2017.
- [20] Holzäpfel, F., Stephan, A., Tchipev, N., Heel, T., Körner, S., and Misaka, T., "Impact of wind and obstacles on wake vortex evolution in ground proximity," *6th AIAA Atmospheric and Space Environments Conference, Atlanta*, No. 2470, 2014.
- [21] Frech, M. and Holzäpfel, F., "Skill of an Aircraft Wake-Vortex Model using Weather Prediction and Observation," *Journal of Aircraft*, Vol. 45, 2008, pp. 461–470.
- [22] Dasey, T. J., Cole, R. E., Heinrichs, R., Matthews, M., and Perras, G., "Aircraft vortex spacing system (AVOSS) initial 1997 system deployment at Dallas/Ft. Worth (DFW) Airport," Tech. Rep. L-3, NASA/A-1, Massachusetts Inst. of Tech.; Lincoln Lab.; Lexington, MA United States, 1998.
- [23] Dougherty, R. P., Wang, F. Y., Booth, E. R., Watts, M. E., Fenichel, N., and D'Errico, R. E., "Aircraft wake vortex measurements at Denver International Airport," *AIAA Paper*, Vol. 2880, No. 10, 2004.
- [24] S.D.Campbell, T.J.Dasey, R.E.Freehart, R.M.Heinrichs, M.P.Matthewsa, G.H.Perras, and G.S.Rowe, "Wake Vortex Field Measurement Program at Memphis, TN Data Guide," research report NASA/L-2. 1997, Lincoln Laboratory, Massachusetts Institute of Technology, 1997.
- [25] Pruis, M. J., Delisi, D. P., Jacob, D., and Lai, D., "Summary of NASA Wake and Weather Data Collection at Memphis International Airport: 2013-2015," *8th AIAA Atmospheric and Space Environments Conference*, 2016, p. 3274.
- [26] Myung, I. J., "Tutorial on maximum likelihood estimation," *Journal of mathematical Psychology*, Vol. 47, No. 1, 2003, pp. 90–100.
- [27] McLachlan, G. and Peel, D., *Finite mixture models*, John Wiley & Sons, 2004.
- [28] Dempster, A. P., Laird, N. M., and Rubin, D. B., "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the royal statistical society. Series B (methodological)*, 1977, pp. 1–38.
- [29] Moon, T. K., "The expectation-maximization algorithm," *Signal processing magazine, IEEE*, Vol. 13, No. 6, 1996, pp. 47–60.
- [30] Sölc, I., Holzäpfel, F., Abdelmoula, F., and Vechtel, D., "Performance of Onboard Wake–Vortex Prediction Systems Employing Various Meteorological Data Sources," *Journal of Aircraft*, 2016, doi:10.2514/1.C033732.
- [31] Tschirk, W., *Statistik: Klassisch oder Bayes*, Springer, 2014.