

Comparison and validation of two high-resolution weather forecast models at Frankfurt Airport

KLAUS DENGLER^{1*}, CHRISTIAN KEIL¹, MICHAEL FRECH², THOMAS GERZ¹ and KIRSTIN KOBER¹

¹Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

²Deutscher Wetterdienst, Meteorologisches Observatorium Hohenpeissenberg, Hohenpeissenberg, Germany

(Manuscript received December 15, 2008; in revised form June 5, 2009; accepted July 2, 2009)

Abstract

In recent years the ‘Nowcasting Wake Vortex Impact Variables’ model NOWVIV has been developed at the Deutsches Zentrum für Luft- und Raumfahrt, DLR, to forecast weather parameters in airport environments. The German Meteorological Service, DWD, employs his COSMO-DE model (Consortium for Small scale Modelling-DEutschland) for operational forecasts in Germany. A systematic comparison of model output from NOWVIV and a derivative of COSMO-DE, named COSMO-FRA, is presented. Both models are centred at Frankfurt Airport with horizontal resolutions of 2.1 km and 2.8 km, respectively. In the DLR Project *Wetter & Fliegen* the COSMO-FRA model will replace the NOWVIV model and become a key component in the future rapid update cycle for adverse weather predictions at the airports of Frankfurt and Munich. The forecast vertical profiles of runway crosswind, head/tail wind, temperature, and turbulent kinetic energy (TKE) are validated against Wind and Temperature Radar (WTR/RASS) measurements operated by the German Air Navigation Service Provider, DFS, within a 40 day period during fall 2004 and a 60 day period during winter 2007. Model and WTR output is provided every 10 minutes. In general it was found that the predictions of both models yield similar skills based on the root mean square error (RMSE) and mean bias statistics of the crosswind as well as the false alarm-rate (FAR) statistics in forecasting different crosswind thresholds. The RMSE of crosswind between ground and 1600 m altitude ranges between 2.2 and 3.0 ms⁻¹ for NOWVIV in fall and winter. For COSMO-FRA this error ranges between 2.0 and 2.5 ms⁻¹ during winter and 2.5 and 3.5 ms⁻¹ during fall, respectively. The FAR for exceeding a crosswind threshold of 3 ms⁻¹ is about 23 % in fall and 17 % in winter for both models.

Zusammenfassung

In den vergangenen Jahren wurde am DLR das ‘Nowcasting Wake Vortex Impact Variables’ Modell NOWVIV mit dem MM5 als dynamischen Kern zur Vorhersage meteorologischer Parameter im Flughafenbereich entwickelt. Der Deutsche Wetterdienst, DWD, nutzt sein COSMO-DE Modell (Consortium for Small scale Modelling – DEutschland) für operationelle Vorhersagen in Deutschland. In der vorliegenden Studie wird ein systematischer Vergleich von Prognosen des NOWVIV und des COSMO-FRA Modells, welches auf dem operationellen COSMO-DE Modell des Deutschen Wetterdienstes basiert, präsentiert. Beide Modelle wurden an die Umgebung des Flughafens Frankfurt angepasst und mit einer horizontalen Auflösung von 2,1 km (NOWVIV) und 2,8 km (COSMO-FRA) betrieben. Im DLR-Projekt *Wetter & Fliegen* wird das COSMO-FRA Modell das NOWVIV Modell ersetzen und eine Schlüsselrolle im ‘Rapid Update Cycle’ zur Vorhersage widriger Wetterverhältnisse an den Flughäfen Frankfurt und München spielen. Vorhersagen vertikaler Profile des die Landebahnen querenden und parallelen Windes, der Temperatur und der Turbulenten Kinetischen Energie (TKE) wurden mit Messungen eines Wind-Temperatur-Radars (WTR/RASS) der Deutschen Flugsicherung (DFS) für eine 40-tägige Periode im Herbst 2004 bzw. einer 60 Tage Periode im Winter 2007 verglichen. Vorhersage- und Messdaten liegen in 10 Minuten Auflösung vor. Generell wurde festgestellt, dass die Prognosen beider Modelle ähnliche Qualität haben. Dieses Ergebnis basiert auf einer Analyse des mittleren quadratischen Fehlers und der Häufigkeit schwellenwertabhängiger Fehlprognosen (False Alarm Ratio, FAR) des Querwindes. Der mittlere quadratische Fehler der Querwindvorhersagen zwischen dem Boden und 1600 m liegt zwischen 2,0 und 3,0 ms⁻¹ für NOWVIV im Herbst und Winter. Für COSMO-FRA liegt dieser Fehler zwischen 2,0 und 2,5 ms⁻¹ im Winter und 2,5 bis 3,5 ms⁻¹ im Herbst. Für beide Modelle beträgt FAR für die Vorhersage von Querwinden von mindestens 3 ms⁻¹ etwa 23 % im Herbst und 17 % im Winter.

1 Introduction

The demand for efficient, safe, and environmentally sustainable air traffic is steadily increasing. Major airports

already today operate at their capacity limits. With increasing demand the air transport system becomes more vulnerable to distortions of all kinds (EUROCONTROL, 2004). One of the major contributors to incidents, accidents, and delays in air traffic are adverse weather conditions, also en-route but especially at and around busy airports. Detailed studies of the impact of weather upon

*Corresponding author: Klaus Dengler, Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, 82234 Weßling, Germany, e-mail: Klaus.Dengler@dlr.de

aviation show that there is a need for improved weather forecast, see THEUSNER and RÖHNER (in preparation) for a comprehensive assessment.

An accurate forecast of wind, turbulence and temperature along the glide paths of an airport is also required to predict the transport and decay of aircraft wake vortices. Since the introduction of Boeing's B-747 aircraft in the 1970's, fixed mass-based separations between aircraft pairs during approach and departure have been introduced by the International Civil Aviation Organization (ICAO) to avoid hazardous wake encounters. Because these safe separations are considered to be over conservative (FRECH and ZINNER 2004), research has been conducted towards a wake vortex advisory system for optimal aircraft spacing (see, e.g., GERZ et al., 2005). The atmosphere in terms of wind speed, wind direction, turbulent kinetic energy, eddy dissipation rate, and vertical stability of air surrounding the vortices affects their horizontal and vertical displacement as well as their decay (HOLZÄPFEL and GERZ, 1999; HOLZÄPFEL et al., 2000; HOLZÄPFEL et al., 2003).

For good predictions a high-resolution weather forecast model should be appropriate which takes into account the orographic and land use characteristics at and around the airport in order to correctly balance the levels of energy, driven by turbulence, surface friction, sensible and latent heat of the air masses in the atmospheric boundary layer. The Deutsches Zentrum für Luft- und Raumfahrt, DLR, has developed the 'Nowcasting Wake Vortex Impact Variables' model NOWVIV (GERZ et al., 2005) to forecast wake-vortex affecting weather parameters in airport environments. The German Meteorological Service, DWD, employs his COSMO-DE model (CONsortium for Small scale Modelling-DEutschland) for operational forecasts in Germany. For future applications at the airports of Frankfurt and Munich, DLR plans to use a derivate of COSMO-DE, named COSMO-FRA, as a key element in the future rapid update cycle for adverse weather predictions.

It is the aim of this paper to present a systematic comparison of the two numerical weather prediction models, NOWVIV and COSMO-FRA. NOWVIV is based on the Mesoscale Model 5, MM5 of Pennsylvania State University and National Center for Atmospheric Research and has been developed to provide temporally (each 10 minutes) and spatially (2.1 km horizontally and 16 to 75 m vertically) highly resolved forecasts of profiles of the required atmospheric variables around the airport of Frankfurt, especially along the two glide paths of landing aircraft. COSMO-FRA is derived from the operational weather forecast model COSMO-DE of DWD, centred at Frankfurt Airport with a horizontal resolution of 2.8 km and an increased vertical resolution varying from 16 to 90 m. The output frequency is also 10 minutes. The model validation and comparison will use data of two measuring campaigns at Frankfurt Airport during fall 2004 and winter 2007. Such a model com-

parison including the COSMO-DE model has not been done before (A. SEIFERT, DWD, personal communication 2006). The measurement data stem mostly from the wind and temperature radar WTR operated by the German Air Navigation Service Provider DFS (KONOPKA and FISCHER, 2005) at the South-West end of the parallel runways of Frankfurt Airport.

The wind component blowing perpendicular to a runway, the crosswind, is a key parameter for airport operations because crosswind transports wake vortices out of the flight path of follower aircraft and crosswind strength, thus, largely determines the aircraft separations during take-off and landing. Especially, the exceedance of certain crosswind thresholds is a required quantity to forecast, followed by the amount of turbulent kinetic energy and temperature profiles. Therefore, the model comparison is based upon several error statistics measures as root-mean-square error, mean bias, false alarm rates and probability of detection.

The setup of the models and the instrumentation used at the airport are described in section 2. In section 3 the statistical analysis of wind, turbulence kinetic energy and virtual potential temperature is presented followed by a case study in section 4. Finally the results are discussed in section 5.

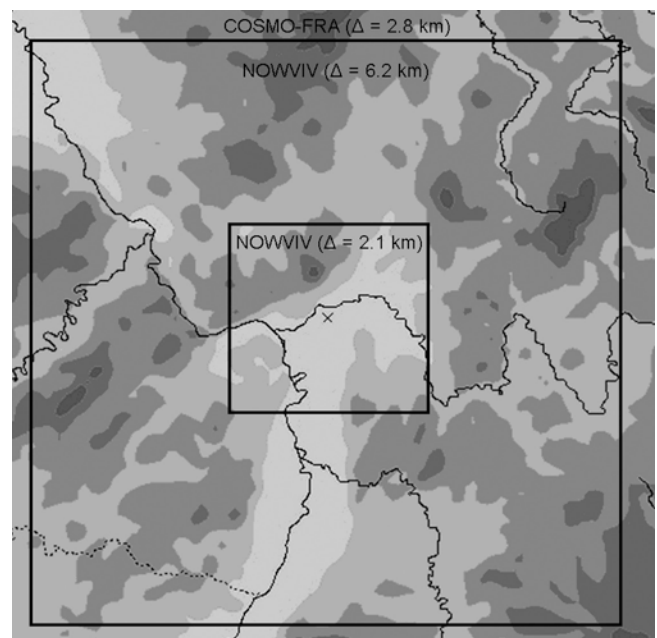


Figure 1: Domain of high resolution models COSMO-FRA and NOWVIV centred at Frankfurt Airport. Vertical profiles at Frankfurt Airport of the COSMO-FRA (2.8 km) and the nested NOWVIV (2.1 km) are used for intercomparison. COSMO-FRA topography is given in gray shading, rivers in black.

2 Model description and measurement setup

a) NOWVIV model:

The core of NOWVIV constitutes the non-hydrostatic meso-scale model 5th-generation, MM5, of Pennsylvania State University and National Center for Atmospheric Research (GRELL et al., 1994). The numerics and physics packages used have been validated successfully in earlier applications (GRELL et al., 2000; FRECH et al., 2007; GERZ et al., 2007). For turbulence a 2nd-order level-3 turbulence closure scheme is used with a prognostic equation for the turbulent kinetic energy TKE (BURK and THOMPSON, 1989). The roughness length used is 0.10 m (grass). NOWVIV has successfully been applied to predict the environmental parameters responsible for wake vortex transport and decay during several field campaigns (WakeOP 2001, WakeTOUL 2002, Wirbelschlepp 2004, 2007, C-Wake 2004, CRE-DOS 2007).

For the current application NOWVIV was centred at Frankfurt Airport using two nested domains with sizes of 250 km x 250 km (horizontal resolution 6.2 km) and 90 km x 90 km (horizontal resolution 2.1 km) (Fig. 1). The model employed 60 vertical levels. In the altitude range of interest ($z < 1600$ m above ground), the atmosphere was resolved by 33 levels with a vertical resolution increasing from 16 to 75 m. Detailed topography, land-use and soil-type data at a resolution of 1 km (MASSON et al., 2003) were used. The main forecast run of COSMO-EU (horizontal resolution 7 km) supplied the initial and hourly boundary data for NOWVIV. The 24-hour forecast run started at 00 UTC. From the model runs vertical profiles of wind, temperature, and TKE are extracted with an output frequency of 10 min at those grid points which are closest to the glide paths, see Fig. 2.

FRECH et al. (2007) compared a 30-yr surface wind climatology with the results of a 1-yr NOWVIV simulation of daily weather in the Frankfurt Airport area and found that the main climate features were recovered by the model. Validating a 40-day subset of the model output with measurements by a SODAR/RASS (sound detection and ranging/radio acoustic sounding system, see below), RMS errors between 1.5 ms^{-1} at the surface and 2 ms^{-1} at 300 m were found.

The profiles forecast by NOWVIV were fed into the probabilistic two-phase model P2P (HOLZÄPFEL, 2003) to predict the transport and decay of the aircraft wake vortices. This method provided wake vortex predictions along the glide path at altitudes between 100 and 1600 m (GERZ et al., 2007). The skill of P2P has been described in FRECH and HOLZÄPFEL (2008). They used the data of the measuring campaign in fall 2004 and found that the best wake vortex forecasts were found for wakes evolving in ground effect.

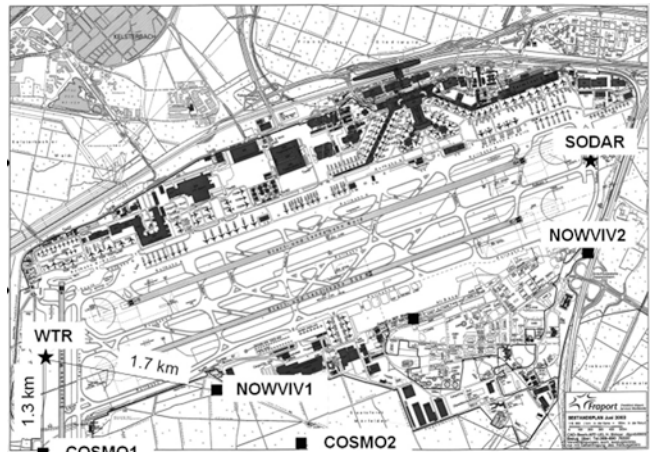


Figure 2: Instrumentation set-up during Frankfurt campaign 2006/2007 and model grid points used for the analysis. Grid points NOWVIV1 (2) and COSMO1 (2) were used for model validation with WTR/RASS (SODAR/RASS) data.

b) COSMO model:

The non-hydrostatic, fully compressible COSMO model has jointly been developed by the Consortium for Small Scale Modelling and is operationally used by several European weather services. The COSMO-DE model is the high resolution model of DWD using a horizontal mesh size of 2.8 km covering an area of roughly $1200 \times 1300 \text{ km}^2$ in Central Europe (STEPPELER et al., 2003; STEPHAN et al., 2008). For the present application the domain has been modified and is centred at Frankfurt Airport encompassing $280 \text{ km} \times 280 \text{ km}$ (see Fig. 2). Since the altitude range of interest constitutes the boundary layer, the vertical resolution there is increased amounting to 16 to 90 m corresponding to 19 levels below 1600 m (M. BALDAUF, DWD, personal communication 2006). This set-up is called COSMO-FRA based upon the COSMO-DE version 3.18. In total there are 50 vertical levels as in operations. The numerics and physics packages follow the operational configuration using a two time level integration scheme based on the Runge-Kutta method of third order and a prognostic turbulence scheme with 2nd order closure (prognostic equation for TKE). The roughness length used in COSMO-FRA is 0.47 m (forest). In NOWVIV a very detailed representation of the ground cover has been implemented and the value of z_0 close to the WTR/RASS has been used. Since it is the aim to compare NOWVIV with the operational COSMO-DE model we used the z_0 as implemented although it differs from the value used in the NOWVIV model. As for NOWVIV, the 24-hour forecast started at 00 UTC and was driven by initial and hourly boundary data from the COSMO-EU model. At 10 min frequency, the vertical profiles of the parameters wind, temperature, and TKE were output.

c) Measurement setup at Frankfurt airport

Two measuring campaigns at Frankfurt Airport will be used for model validation. The first campaign took place during fall 2004 (26 August–5 October 2004); its instrumentation setup is described in detail by FRECH et al. (2007) and FRECH (2007). The data was used by FRECH and HOLZÄPFEL (2008) to estimate the skill of the wake-vortex model P2P. The second measuring campaign during winter 2006/2007 (21 December 2006–28 February 2007) used a very similar instrumentation. Fig. 2 shows the position of the instrumentation during that period.

The prime source for measured data, especially for levels above about 100 m, was the DFS' Wind and Temperature Radar supplemented with a radio acoustic sounding system, WTR/RASS (KONOPKA and FISCHER, 2005). The WTR is situated at the end of runway 25L just west of runway 18. Among other parameters, the WTR/RASS delivers vertical profiles of the 3 wind components, the standard deviation of vertical velocity, and virtual temperature to a maximum altitude of 1650 m at a vertical resolution of 30 m starting at 60 m height. Every measured data is provided with an internally calculated confidence value between 0 (bad) and 4 (good). It is strongly recommended by DFS to use only data with high confidence values above 2 for the validation process. From the WTR/RASS more than 100 000 profiles are available for the systematic comparison of the models. It should be pointed out that the measuring data and model output data are independent since the WTR measurements were not assimilated into the models. In the context of this paper we focus on the lowest atmospheric layers up to 1650 m. For further evaluations at the low altitudes of 10 and 40 m, data from a METEK, Inc., SODAR PCS. 2000-64 with a 1290-MHz RASS was used. The SODAR provided 10-min averaged profiles of the 3 wind components, the standard deviation of vertical velocity and the virtual temperature with a vertical resolution of 20 m starting at 40 m (which represents an average between 30 and 50 m). The maximum vertical range of the measurements was limited to 300 m. According to the manufacturer the accuracy of measured wind speed lies within 5 %. For wind direction the accuracy depends on wind speed and is estimated to be between 1°–3° for wind speeds above 5 ms⁻¹ and 3°–5° at lower wind speeds. The standard deviation of vertical wind speed is ± 0.1 ms⁻¹ and the measurement accuracy of virtual temperature by the RASS is ± 0.3 K. Although the designed maximum vertical range of the SODAR measurements is 300 m, measurements only up to 100 m during strong wind situations are available. Therefore SODAR/RASS data is only used in some cases to validate the model results at low altitudes.

In addition to the SODAR an USA-1 sonic anemometer with a sampling frequency of 20 Hz has been installed to measure wind components at 10 m height. Typical maximum error of wind speed measurement is

estimated between ± 0.1 and ± 0.3 ms⁻¹ for wind speeds up to 10 ms⁻¹ and 2 % for higher wind speeds. For wind direction the maximum errors are estimated between ± 2 and $\pm 5^\circ$ for wind speeds up to 10 ms⁻¹ (higher values observed for low wind speeds) and $\pm 2^\circ$ for wind speeds above 10 ms⁻¹.

The time averaging used in the present study is driven by wake vortex properties and life times and not by meteorological requirements; larger averages than 10 min are useless to characterise the impact of wind upon the vortices. Since the models are applied to wake vortex prediction the averaging time of 10 min has been used for the analysis.

d) Comparison methodology

Both high resolution models are systematically validated by using observed vertical profiles at high temporal frequency (every 10 minutes) of the parameters wind, TKE and temperature recorded by the WTR/RASS and SODAR/RASS instruments. For each model run the forecasts are compared to the measurements made at each time at the model grid point nearest to the position of the instrument (Fig. 2). Hence, the distances at the airfield between the measured profiles by WTR and the grid-point profiles at COSMO1 from the COSMO-FRA model amount to 1.3 km and 1.7 km to the profiles at NOWVIV1 from the NOWVIV model.

The two continuous verification measures mean bias, defined as the median of the difference of prediction and observation, and Root-Mean-Square Error RMSE are used for the analysis. RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x^{fc} - x^{obs})^2}$$

with n being the number of values, x^{fc} and x^{obs} as the forecasted and observed values of the parameter X , respectively.

They are complemented by the three categorical metrics BIAS score, FAR (False Alarm Ratio) and POD (Probability Of Detection). Using the contingency table (Tab. 1) these three scores are calculated as follows:

$$FAR = \frac{\text{false alarms}}{\text{hits} + \text{false alarms}}$$

FAR describes which fraction of the predicted “yes” events (“hits+false alarms”) did not occur, i.e. were “false alarms”. The perfect score is 0.

$$POD = \frac{\text{hits}}{\text{hits} + \text{misses}}$$

POD describes the fraction of the observed “yes” events that were correctly forecasted out of the total population “hits+misses”. The perfect score is 1.

$$\text{Bias Score} = \frac{\text{hits} + \text{false alarms}}{\text{hits} + \text{misses}}$$

Table 1: Contingency table used to calculate the categorical metrics BIAS score, FAR (False Alarm Ratio) and POD (Probability of detection).

Forecast	Observed		
		yes	no
	yes	hits	false alarms
	no	misses	correct negatives

Bias score describes how the forecasted frequency of “yes” events compares to the observed frequency of “yes” events. The perfect score is 1.

It should be aimed to minimize FAR and maximize POD with a BIAS score close to 1. FAR, POD and the BIAS score of the model crosswind forecasts exceeding defined thresholds from 1 ms^{-1} to 9 ms^{-1} are discussed in detail. For a detailed description of the skill scores see WILKS (1995).

In this context a hit is defined as a forecast of crosswind exceeding a certain threshold value that has also been observed. It should be noted that in this analysis measurement errors are not taken into account meaning that the measurements are assumed to represent the truth. Crosswind is defined positive when directed to the south-southeast. Since we focus on the strength of crosswind in this study the mean bias is defined as the difference of the absolute values of crosswind prediction and observation.

3 Statistical analysis of wake-vortex environmental parameters

In this section vertical profiles of wind speed, wind direction, crosswind, virtual potential temperature and turbulent kinetic energy during the winter 2007 and fall 2004 campaigns will be analysed. Additionally the above defined skill scores will be calculated for crosswinds exceeding defined threshold values.

a) Wind profiles

To begin with the skill of the two models to predict the magnitude of the three-dimensional wind vector during the winter campaign is discussed (without figures). The RMSE of NOWVIV wind speed forecasts lies around 2 ms^{-1} in the lower 200 m increasing to 3 ms^{-1} at 1650 m. The RMSE of the COSMO-FRA predictions within the lower 200 m is slightly higher than the RMSE of the NOWVIV predictions, increasing less with height reaching maximum values around 2.6 ms^{-1} at 1000 m altitude. Especially when using SODAR/RASS/SONIC data at 10 m and 40 m height the RMSE of both model predictions is slightly below 2 ms^{-1} .

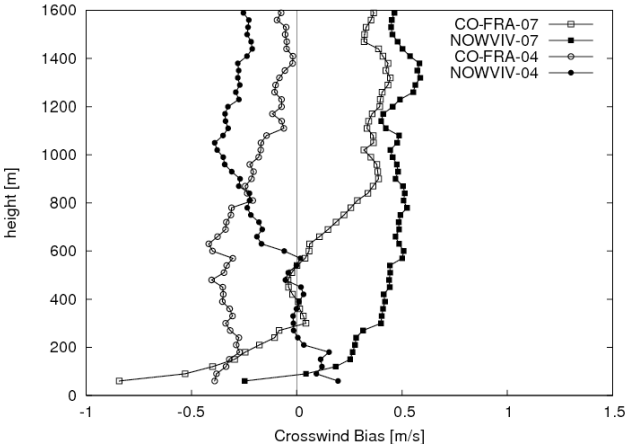


Figure 3: Mean bias of crosswind magnitude (u_c) for the fall 2004 (denoted by 04) and winter 2007 (denoted by 07) measurement periods.

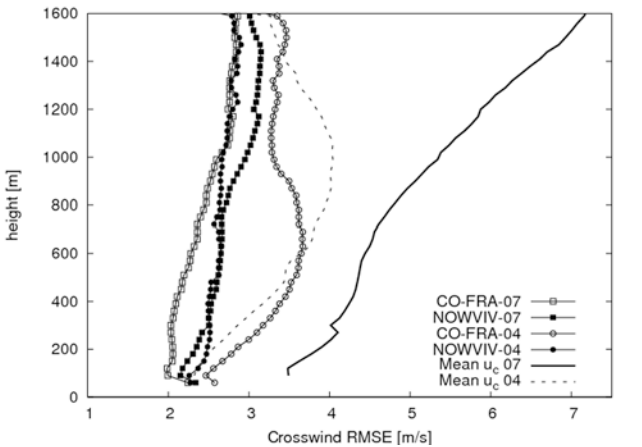


Figure 4: As Fig. 3 for RMSE of crosswind magnitude. Also the mean crosswind (u_c) is plotted for the winter period 2007 (solid line) and the fall period 2004 (dashed line).

The mean bias shows an underestimation of predicted wind speeds at lower levels (below 60 m for NOWVIV, below 270 m for COSMO-FRA). At 600 m height, both models overestimate the wind speed with values of 1 ms^{-1} for NOWVIV and 0.5 ms^{-1} for COSMO-FRA. Above 1000 m no mean bias of the NOWVIV forecast is observed whereas the mean bias of the COSMO-FRA forecasts decreases from 0.5 ms^{-1} at 1000 m to 0 ms^{-1} at 1650 m.

The RMSE of the wind direction is a little smaller for COSMO-FRA (values between 22° and 37°) compared to those found in NOWVIV predictions (values between 23° and 40°), both slightly decreasing with height. These relatively high values of RMSE are a result of the larger error of measured wind direction at low wind speeds. Using only wind directions of wind speeds larger than 5 ms^{-1} the RMSE reduced to values between 18° and 27° for COSMO-FRA and between 22° and 27° for NOWVIV. Very similar scores are obtained for wind

speed and wind direction during the measurement period in fall 2004. A negative mean bias around 15° and 18° is observed for both models with very little height variation. This corresponds with a deviation of the forecasted wind vector to the left of the measured wind vector.

b) Crosswind Profiles

From wind speed, wind direction and runway orientation the crosswind magnitude is calculated and the mean bias and RMSE are shown in Fig. 3 and 4 respectively. In the winter period 2007 an overestimation of crosswinds above 90 m predicted by NOWVIV and an underestimation by COSMO-FRA is observed from the ground to 200 m. Values lie between -0.8 and 0.4 ms^{-1} for COSMO-FRA and between -0.25 and 0.5 ms^{-1} for NOWVIV (Fig. 3). Above 800 m both models show a similar bias around 0.4 ms^{-1} . During the period in fall 2004 the underestimation of COSMO-FRA predictions is reduced to less than -0.5 ms^{-1} decreasing with height. For NOWVIV predictions during the fall period the mean bias decreases from 0.25 ms^{-1} at low levels to nearly zero between 200 and 600 m and decreasing thereafter to values around -0.3 ms^{-1} .

For the winter period 2007 the RMSE for crosswind is increasing from about 2 ms^{-1} at low levels to 2.8 ms^{-1} (COSMO-FRA) and 3.1 ms^{-1} (NOWVIV) at 1600 m (Fig. 4). During the fall period RMSE of COSMO-FRA is significantly higher compared to NOWVIV and to the winter period reaching values of more than 3.5 ms^{-1} at heights between 400 and 800 m. At levels below 800 m no significant change is observed for RMSE of NOWVIV predictions between the two campaigns. Above that level even a decrease is observed when compared to the winter period. The mean crosswind during the two campaigns shown in Fig. 4 reveals that during the winter period stronger wind were observed.

c) Skill analysis of crosswind forecast in the winter 2007 campaign

In this section the skill scores of the models to forecast crosswind events exceeding defined thresholds are analysed using the WTR and SODAR data in winter 2007.

All crosswind data of the SODAR system ranging from 10 m to 300 m are used in the first step of this analysis. Fig. 5 indicates that the SODAR data sample size is quickly decreasing with increasing crosswinds which is also due to the fact that the vertical range of SODAR measurements on high wind speed days is limited to about 100 m. About 50 % of the cases refer to crosswinds less than 4 ms^{-1} . Fig. 5 shows that in terms of FAR both models show equal skill with values around 0.35 for crosswind thresholds up to 3 ms^{-1} meaning that in 35 % of the cases the forecast of the crosswind lies above 3 ms^{-1} but measured winds were below. Thereafter, FAR is increasing slightly more for

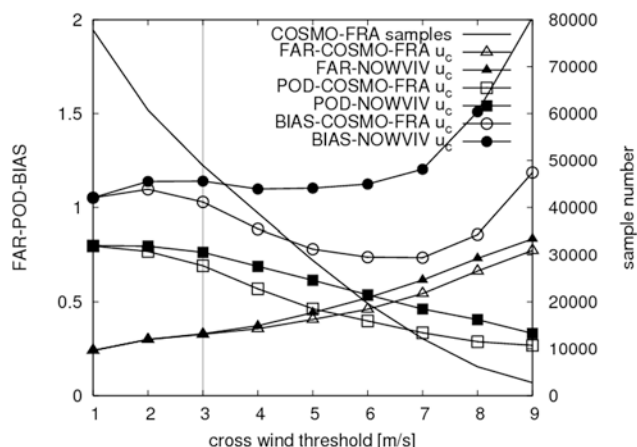


Figure 5: FAR, POD and BIAS score of the magnitude of predicted model crosswinds exceeding thresholds from 1 to 9 m/s using SODAR data during winter 2007. Data of all altitudes between 10 m and 300 m is used. Solid line without points shows the data sample (right vertical axes). Lines with triangles indicate FAR, with squares indicate POD and with circles indicate BIAS score. Solid forms correspond with NOWVIV. Thin vertical line at 3 ms^{-1} indicates a relevant threshold for airport operation.

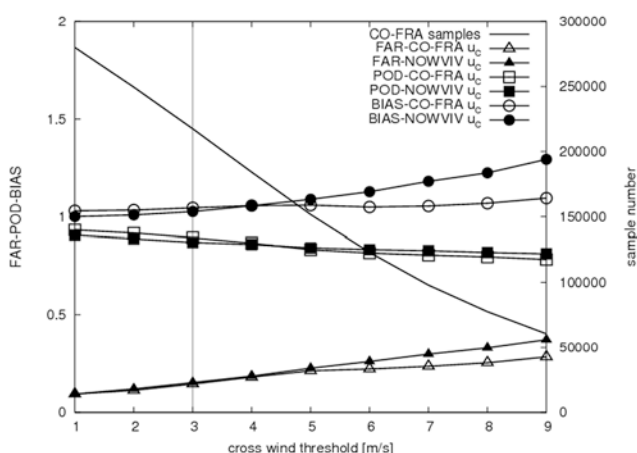


Figure 6: As Fig. 5 but using WTR data for altitudes from 60 m to 1650 m (winter 2007).

NOWVIV predictions. In terms of POD NOWVIV always shows better skill while the BIAS score is always smaller for COSMO-FRA. While the NOWVIV predictions overestimate the number of crosswind events for all crosswind thresholds the COSMO-FRA predictions tend to underestimate crosswind events between 4 and 8 ms^{-1} (BIAS score less than 1).

Note that the analysis covers only the lower 300 m, under strong wind conditions data of only the lower 100 m is available meaning that strong wind cases at altitudes above 100 m are not included in the statistics. Therefore we will now do the same analysis using WTR data which is available between 60 and 1650 m even under strong wind conditions at higher altitudes.

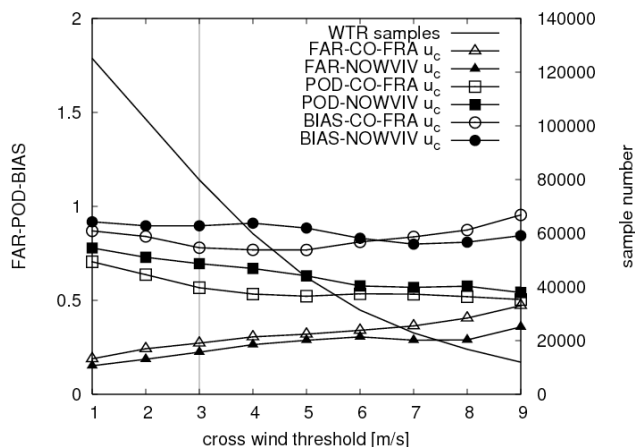


Figure 7: As Fig. 5 but for the measure campaign during fall 2004 using WTR data for altitudes from 60 m to 1650 m.

As for the SODAR data set the sample based on WTR data (solid line in Fig. 6) is decreasing with increasing crosswind threshold indicating that in 50 % of the measurements crosswinds were below about 5 ms^{-1} . In terms of FAR both models show nearly the same skill up to crosswind thresholds of 5 ms^{-1} . For crosswinds smaller than 3 ms^{-1} both models showed FAR values of about 0.15 meaning that in 15 % of the cases the forecast of the crosswind lies above 3 ms^{-1} but measured winds were below.

Both models score very similar in terms of POD. Measured crosswinds larger than 3 ms^{-1} are correctly predicted in 89 % of the COSMO-FRA forecasts and in 87 % of the NOWVIV predictions. Analysing BIAS score shows no major differences for crosswinds less than 5 ms^{-1} . For larger crosswinds the NOWVIV predictions seem to over forecast the cases of large crosswinds (BIAS score value 1.29 for NOWVIV, 1.09 for COSMO-FRA).

An analysis of each of the two winter months separately shows basically the same results. The BIAS score of the model forecasts was larger in January due to the stronger prevailing winds observed compared to February. This confirms the result that both forecast models tend to overestimate crosswinds larger than 5 ms^{-1} .

Next, the skill within different altitude ranges have been analysed (0–1000 m and 1000–1650 m) and best skills were found for the range of 1000–1650 m. Hence, we conclude that the skill of the model predictions is best above the boundary layer (for high crosswinds FAR did not exceed a value of 0.22 for COSMO-FRA and 0.29 for NOWVIV).

Finally, we analysed the average skill of the model predictions for the winter campaign during three periods of the day: in the morning (0400–0800 UTC, local time = UTC+1h), around noon (1000–1400 UTC) and in the evening (1800–2200 UTC). The skill in terms of FAR was best in the morning, decreasing towards noon and increasing again in the evening hours. Around noon

FAR for crosswind predictions larger 3 ms^{-1} was 0.19 for both models increasing to 0.33 (COSMO-FRA) and 0.42 (NOWVIV) for crosswinds larger 8 ms^{-1} . This is considerably larger compared to the morning or evening hours with FAR values of 0.11 for both models and crosswinds larger 3 ms^{-1} and 0.21 (COSMO-FRA) or 0.28 (NOWVIV) for crosswinds larger than 8 ms^{-1} . Hence, the lower skill around noon may be attributed to increased convection.

The reported error statistics for the wind speed do not depend on the lead time of the forecast: Two COSMO-FRA runs, one started at 0 UTC, the other at 12 UTC, hence with lead times of 18 and 6 hours, respectively, showed no difference in the RMSE of wind in the 18 to 22 hour evening window averaged over the 60 days of the campaign.

In terms of POD no significant difference between the skills of the models can be seen for the time period around noon. However, in the morning and evening hours COSMO-FRA shows higher POD for crosswind thresholds less than 5 ms^{-1} but lower POD for higher thresholds when compared to NOWVIV (not shown).

d) Skill analysis of crosswind forecast in the fall 2004 campaign

Fig. 7 shows the skill scores for FAR, POD, and BIAS score using the data gathered during the fall campaign 2004. Comparing Fig. 6 with Fig. 7 shows that the models perform less well as FAR is higher while POD is lower. One reason for that decreased performance could be that during the fall period the atmospheric dynamics are generally more convectively driven than during the winter months. It can also be seen that COSMO-FRA does have slightly higher values of FAR compared to NOWVIV.

e) Turbulent kinetic energy (TKE) in the winter 2007 campaign

In this section we analyse the forecast skill for turbulent kinetic energy (TKE). Note that the models use different schemes to determine TKE. The forecasts are compared with TKE measured by the WTR which is calculated from the standard deviation of vertical velocity ($\text{TKE} = 1.5 * \sigma(w)^2$).

The results of the RMSE and mean bias analysis are shown in Fig. 8. For TKE the NOWVIV forecasts score significantly better compared to those of the COSMO-FRA especially at low levels. NOWVIV forecasts show only very little underestimation and the values of RMSE are between 0.2 and $0.6 \text{ m}^2\text{s}^{-2}$, whereas the forecasts of the COSMO-FRA exhibit high values of RMSE especially below 200 m (values larger than $4 \text{ m}^2\text{s}^{-2}$). Above that height RMSE slowly decreases and reaches the values of the NOWVIV forecasts above 1200 m. Furthermore, COSMO-FRA forecasts overestimate TKE. Repeating several predictions with the new version 4.2 of

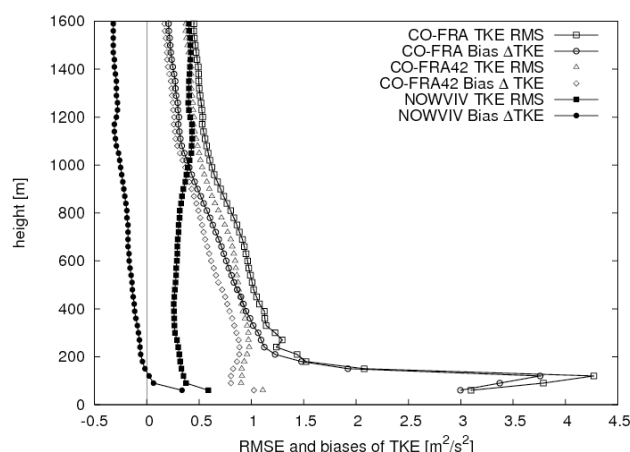


Figure 8: RMSE and mean bias of turbulent kinetic energy [m^2s^{-2}] for NOWVIV (filled symbols), COSMO-FRA (open symbols) and COSMO-FRA version 4.2 (triangle for RMSE and diamonds for mean bias) for winter 2007.

COSMO, where a safety condition for the explicit vertical TKE diffusion has been introduced to avoid instabilities (M. RASCHENDORFER, DWD, personal communication 2008), resulted in a considerably improved scoring of COSMO-FRA (Fig. 7) compared to version 3.18 at lower altitudes (RMSE and mean bias of TKE are now below $1 \text{ m}^2\text{s}^{-2}$) but still the RMSE of the COSMO-FRA forecasts of TKE at heights below about 700 m is about twice as large as the RMSE produced by NOWVIV. This difference can be explained by the higher roughness length used in COSMO-FRA (0.47 m) compared to NOWVIV (0.10 m) since higher TKE values are associated with higher roughness length.

Note that for the winter campaign an increase of RMSE and mean bias for the predictions of wind speed, crosswind and virtual potential temperature is observed using the new version 4.2 of the COSMO-FRA model whereas not much change in skill was observed for the prediction of wind direction.

f) Virtual potential temperature (θ_v) in both campaigns

The ability of the models to properly predict vertical stability, which is described here in terms of virtual potential temperature, is investigated next (Fig. 9). The RMSE of the NOWVIV predictions is about 2.5 K at low levels, decreasing to about 1.5 K at about 1000 m altitude and then increasing again to about 2 K at 1400 m altitude. The RMSE of the COSMO-FRA (version 3.18) predictions lies between 1.5 K within the lower 1000 m increasing to about 1.7 K at 1200 m while decreasing to about 1.2 K at 1600 m. Compared to NOWVIV predictions the RMSE of the COSMO-FRA predictions of θ_v is smaller within the lower 700 m and of similar magnitude above that level.

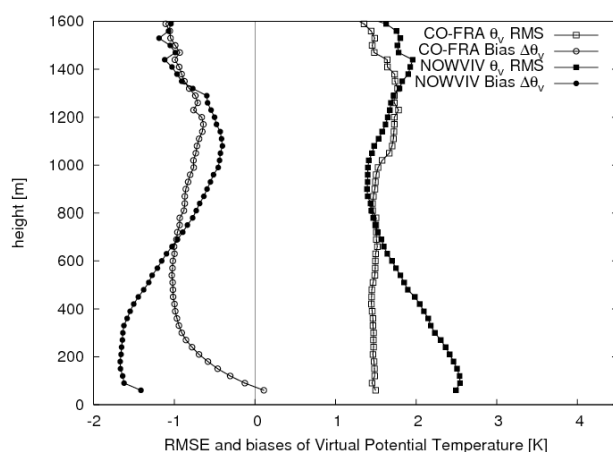


Figure 9: RMSE and mean bias of virtual potential temperature (θ_v) for winter 2007

Analysing the mean bias of both models a generally negative mean bias is observed indicating an underestimation of θ_v . The mean bias of NOWVIV decreases from -1.4 K at low levels to -1.7 K at about 200 m, increasing from there to -0.5 K at about 1100 m and decreasing again thereafter to a value around -1.2 K at 1500 m. Physically spoken NOWVIV predictions tend to more unstable conditions up to 200 m as well as between 1100 m and 1600 m while the layer between 200 m and 1100 m is predicted to be more stable. The mean bias of COSMO-FRA predictions decreases from around 0 K at low levels to about -1.0 K at 600 m thereafter slowly increasing to about -0.8 K at 1200 m and decreasing again to -1 K aloft. This indicates that COSMO-FRA predictions tend to unstable conditions up to 600 m and above 1200 m while tending to stable conditions in the layer between.

During the fall campaign 2004 the RMSE of NOWVIV predictions of θ_v decreased throughout all levels especially between 100 m and 600 m (by 0.5 K – 0.8 K) and 1000 m and 1600 m (close to 1 K) when compared with the winter campaign. For the COSMO-FRA predictions a slight increase is observed within the lower 600 m followed by a decrease above (up to 0.5 K). The mean bias of the NOWVIV predictions is increased below 1000 m in such a way that the stable layer above 200 m has vanished. Mean bias of COSMO-FRA predictions is increased below 1000 m and slightly decreased above. However, the vertical structure is nearly unchanged compared to the winter 2007 period.

4 Case study – frontal passage on 8 February 2007

Frontal passages are crucial weather events for aviation since operational procedures have to be adapted to the changing weather and wind conditions. Fronts are also

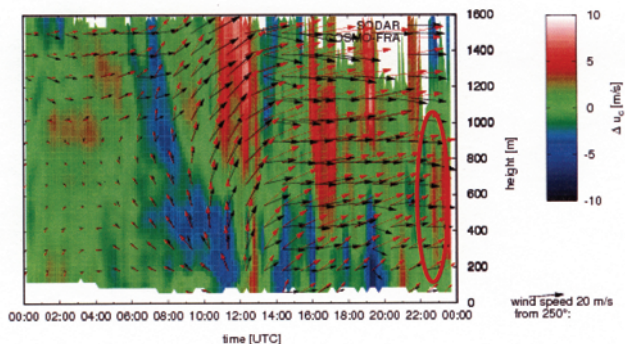


Figure 10: The horizontal wind vector measured by WTR (black arrows) and predicted by COSMO-FRA (red arrows) for a forecast period of 24 h on the 8th of February 2007. Coloured background shows difference between predicted and observed magnitude of crosswind. Red oval indicates overestimation of crosswind magnitude.

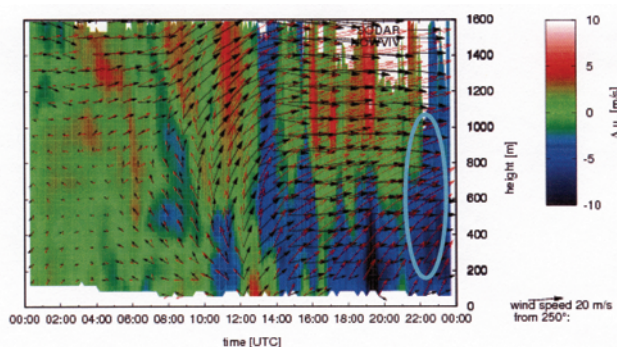


Figure 11: As Fig. 10 but for NOWVIV forecasts. Blue oval indicates underestimation of crosswind magnitude.

a challenging feature for weather prediction models especially in terms of an accurate forecast of time and intensity of its passage. Therefore we study in some detail the synoptic situation on 8 February 2007 which was characterized by the passage of a very weak low pressure system through the centre part of Germany. Fig. 10 depicts the horizontal wind vector as measured by WTR/RASS as function of height and time (black arrows). At night winds were very weak from varying directions from the ground up to about 1200 m increasing and turning west above that level. Between 5 and 8 UTC we observe easterly winds close to the ground, turning south with height up to 1000 m and coming from west above. Later on, the wind speeds increase and turn from southeast over southwest to west at all levels. After about 15 UTC strong westerly winds dominated all levels from the ground up to 1600 m. Hence, the displayed day features a wind change with height at early morning hours (notable for aviation since the wind may change from a headwind to a tailwind situation during the approach and landing phase of flight) and the passage of the front affecting all height levels between ground and 1600 m.

The red arrows in Fig. 10 and Fig. 11 show the horizontal wind vector as forecasted at 00 UTC for 24 hours by the COSMO-FRA and NOWVIV models, respectively. It is worth to note that the predictions of COSMO-FRA for the night and morning hours until 12 UTC meet the measurements quite well, including the wind change with height especially for altitudes between ground and 1000 m. NOWVIV performs less well here as it delays the wind change from southeast to south and southwest by several hours. In both cases the deviation of the predicted wind vector to the left of the measured wind vector as described by the negative mean bias of wind direction in the previous section is also visible.

The coloured backgrounds in Figs. 10 and 11 denote the difference between predicted and measured crosswind for both models. COSMO-FRA exhibits significant time-height sections where the crosswind is strongly overestimated. In these sections also NOWVIV basically overestimates the crosswind but NOWVIV also shows two large sections of strong underestimation closer to the ground. The red and blue ovals in the figures mark the same time-height section where COSMO-FRA overestimates and NOWVIV underestimates the crosswind.

The mean biases averaged over the entire day are shown in Fig. 12 and show that COSMO-FRA overestimates crosswind below 600 m with values up to 1.0 ms^{-1} . Above that altitude nearly no mean bias is observed. NOWVIV overestimates crosswind below 800 m (values between 0.3 and 2 ms^{-1}) and only slightly underestimates crosswind above. For altitudes below roughly 700 m, the RMSE of the crosswind prediction by NOWVIV for the 8th of February (Fig. 12) is significantly larger than for COSMO-FRA predictions. Above 700 m the RMSE of both models is around 2 ms^{-1} .

The RMSE of the prediction of virtual potential temperature, shown in Fig. 13, is rather small with values around 1 K at height levels above 300 m. Below that height the RMSE of the COSMO-FRA prediction increases quicker than the one of NOWVIV reaching values of 2.2 K at 60 m altitude compared to 1.4 K. This is opposed to the feature observed on average in the entire winter 2007 period (see Fig. 9). For the 8th of February a positive mean bias throughout most of the lower altitudes up to about 1000 m is found for both models. This mean bias is much stronger for COSMO-FRA (value close to 2 K) than for NOWVIV below 300 m. In fact the mean bias of the latter is even slightly negative in the lower 180 m. Hence, the trend of the models to heat the lower altitudes and to cool the atmosphere above 900 m on this day opposes the mean trend depicted in Fig. 9. On the other hand, the scoring of the models with respect to RMSE of TKE on that day is similar to the averaged one shown in Fig. 8 with maximum values of RMSE below 200 m altitude of nearly $7 \text{ m}^2 \text{ s}^{-2}$ for COSMO-FRA (version 3.18) and about $0.5 \text{ m}^2 \text{ s}^{-2}$ for NOWVIV. Again, errors decrease rapidly above 200 m altitude.

In Fig. 14 we show vertical profiles of measured

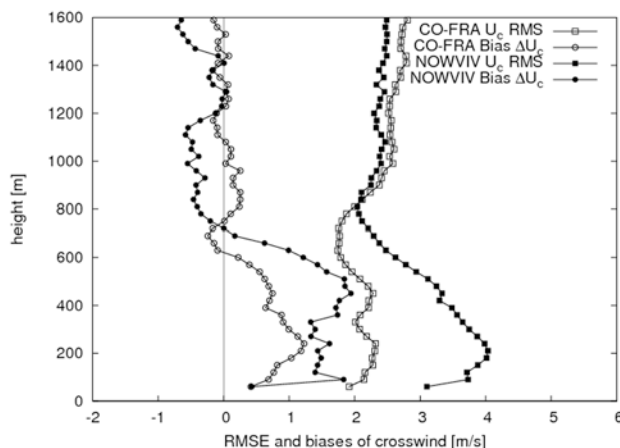


Figure 12: RMSE and mean bias of crosswind magnitude (u_c) for 8th of February 2007.

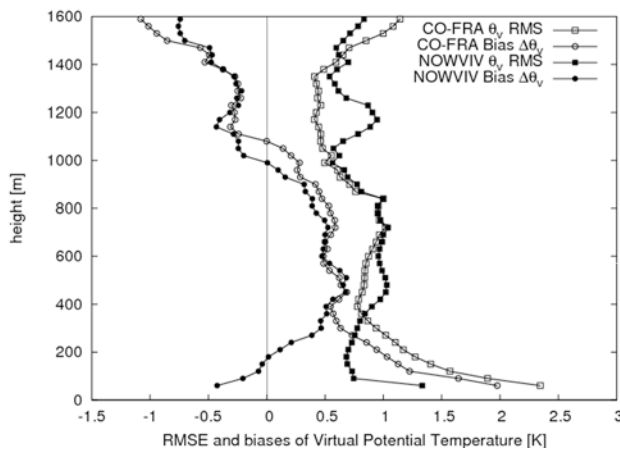


Figure 13: Fig. 12 but for virtual potential temperature.

and predicted virtual potential temperature and RMSE of TKE of the two models at 04:00 UTC on the 8th of February. Note that RMSE in the case of one pair of measurement and forecast is reduced to the mean absolute error. At this time the wind speeds were weak below 1000 m (see Fig. 10) and the profiles of virtual potential temperature indicate a stable stratification which is confirmed by the model forecasts. Under these conditions, however, the SODAR/RASS profile of θ_v differs from the one by WTR/RASS. This is due to different surface conditions between the two instrument locations which are roughly 4 km apart. We also analysed the vertical profiles of the same quantities at 18 UTC (not shown). At that time wind speeds were much higher starting with 2 ms^{-1} at 10 m, reaching values of 4 ms^{-1} at 90 m and even 23 ms^{-1} at 1320 m. Now, no significant difference between the measurements of SODAR/RASS and WTR/RASS can be observed. This indicates that under windy conditions both measuring locations can represent the airport area because the dynamic and thermody-

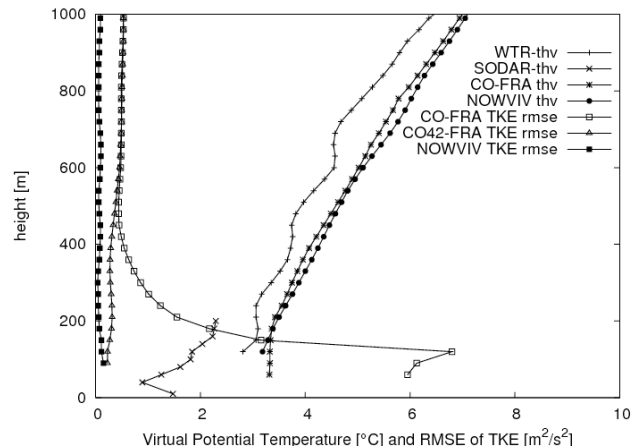


Figure 14: Vertical profiles of measured (SODAR, WTR) and predicted virtual potential temperature (NOWVIV, COSMO-FRA) and RMSE of TKE of the two models at 04:00 UTC on 8th February 2007.

namic state of the atmosphere is controlled by advective processes.

At 18 UTC, hence under windy but thermally still stable conditions we observe that the RMSE of the TKE prediction of COSMO-FRA 3.18 is reduced to about $4 \text{ m}^2\text{s}^{-2}$. An overall increase of the forecast error of TKE predictions from around $0 \text{ m}^2\text{s}^{-2}$ to around $1 \text{ m}^2\text{s}^{-2}$ is observed for both models at that time. This shows, although errors have decreased significantly, that under advective conditions it is still important to have the correct roughness length.

5 Summary and conclusions

A systematic validation of the NOWVIV and the COSMO-FRA numerical weather prediction models for the Frankfurt Airport area has been presented, referring to WTR/RASS and SODAR/RASS data as reality. The data have been gathered during two campaigns, from 26th of August to 5th of October 2004, named “fall”, and from 21st of December to 28th of February 2007, named “winter”. The focus was put on profiles of meteorological parameters that are important for wake vortex transport and decay such as runway crosswind, turbulent kinetic energy (TKE), and virtual potential temperature. Mean biases, root-mean-square errors, false alarm ratios (FAR), probabilities of detection (POD) and BIAS scores of these quantities have been analysed.

The mean bias of both models for predicting crosswind is not larger than 0.5 ms^{-1} . Generally speaking, in fall both models exhibit a slightly negative bias, whereas in winter the predictions tend to be somewhat larger than the measured crosswind. Also, both models yield similar skills in terms of the root mean square error (RMSE). The RMSE of crosswind between ground and 1600 m altitude ranges between 2.2 and 3.0 ms^{-1} for NOWVIV in fall and winter. For COSMO-FRA this error ranges

between 2.0 and 2.5 ms^{-1} during winter but is with up to 3.5 ms^{-1} larger during fall.

In terms of predicting crosswind thresholds, both models again score very similarly in terms of the probability to correctly predict the exceedance of a threshold value (POD) as well as to falsely predict that exceedance (FAR). Furthermore, FAR is lowest during the morning (0400–0800 UTC) and evening (1800–2200 UTC) hours reaching a maximum around noon, indicating the impact of convection around noon.

Both models reveal a cold bias for fall and winter periods. This bias in temperature is more pronounced for NOWVIV below 600 m. NOWVIV tends to stabilise the atmosphere above 400 m whereas COSMO-FRA predictions show a destabilisation between ground and 400 m altitude.

In summary, the largest and systematic differences in error statistics between the two models appear between ground and 600 to 800 m altitude. Consistently, FAR values for crosswind thresholds are largest in the boundary layer (0 m–1000 m) and lowest in the layer aloft from 1000 m–1650 m. At low altitudes the treatment and parameterisation of surface processes (parameterisation of surface friction, turbulence, surface energy balances) play the strongest role. For example, in Version 4.2 of COSMO-FRA a safety condition for the explicit vertical diffusion of TKE has been implemented to avoid instabilities. This results in significantly reduced values of RMSE and mean bias for TKE below 200 m. Although both measures are still systematically larger for COSMO-FRA than for NOWVIV they are not exceeding values of $1 \text{ m}^2\text{s}^{-2}$ and, hence, are acceptable since TKE in the model typically ranges between 1 and $10 \text{ m}^2\text{s}^{-2}$ in the boundary layer (FRECH et al., 2007).

Nevertheless, since crosswind and TKE are key parameters to predict transport and decay of aircraft wake vortices in DLR's wake vortex advisory system (HOLZÄPFEL et al., 2007; GERZ et al., 2007) a correct forecast of these parameters, especially within the lowest 300 m is essential. Therefore, a further increase in forecast skill of the COSMO-FRA model is desirable and should be expected with an adjusted land use data set for COSMO-FRA, possibly combined with higher horizontal and vertical resolutions to account for specific topographic and land-use features at and around the airport.

Moreover, in the DLR project “Wetter & Fliegen (Weather & Flying)” it is aimed to improve the forecasts also by assimilating data into the model which are gathered in the airport environment, such as WTR/RASS and AMDAR data as well as precipitation data from radar. The model will be started hourly in a Rapid Update Cycle (RUC) mode providing very short range time-lagged ensemble forecasts. An improvement especially of the short term forecasts up to 6 h is expected which is very relevant for forecasting wake vortices, thunderstorms and winter weather. The DLR project is closely

linked to the project ITWS/LLWAS by the German Meteorological Service. These forecasting systems should enable airport operators to increase safety and punctuality caused by these weather phenomena.

Acknowledgments

We gratefully acknowledge Fraport AG, METEK and the German Meteorological Service (DWD) for their support during the measurement campaigns. We thank Deutsche Flugsicherung GmbH (DFS) for providing the WTR datasets of the two measuring campaigns. The comments of the reviewers are greatly appreciated. Finally the financial support from the DLR project Wirbelschleppes and from the EU project CREDOS (AST5-CT-2006-030837) is greatly acknowledged.

References

- BURK, S.D., W.T. THOMPSON, 1989: A vertically nested regional numerical prediction model with second-order closure physics. – *Mon. Wea. Rev.* **117**, 2305–2324.
- EUROCONTROL, 2004: Challenges to growth. – EATMP Infocentre Eurocontrol Tech. Rep., 66 pp.
- FRECH, M., 2007: Estimating the turbulent energy dissipation rate in an airport environment. – *Bound.-Layer Meteor.* **123**, 385–393.
- FRECH, M., F. HOLZÄPFEL, 2008: Skill of an aircraft wake-vortex model using weather prediction and observation. – *J. Aircraft* **45**, 461–470.
- FRECH, M., T. ZINNER, 2004: Concept of wake vortex behaviour classes. – *J. Aircraft* **41**, 564–570.
- FRECH, M., F. HOLZÄPFEL, A. TAFFERNER, T. GERZ, 2007: High resolution weather data base for the terminal area of Frankfurt Airport. – *J. Appl. Meteor. Climatol.* **46**, 1913–1932.
- GERZ, T., F. HOLZÄPFEL, W. BRYANT, F. KÖPP, M. FRECH, A. TAFFERNER, G. WINKELMANS, 2005: Research towards a wake vortex advisory system for optimal aircraft spacing. – *Compt. Rend. Phys.* **6**, 501–523.
- GERZ, T., F. HOLZÄPFEL, W. GERLING, A. SCHARNWEBER, M. FRECH, A. WIEGELE, K. KOBER, K. DENGGLER, S. RAHM, 2007: The wake vortex prediction and monitoring system WSVBS. Part II: Performance and ATC integration at Frankfurt Airport. 1st CEAS European Air and Space Conference, Berlin. – CEAS-2007-178, 3391–3399.
- GRELL, G., J. DUDHIA, D.R. STAUFFER, 1994: A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5). – NCAR Tech. Note NCAR/TN-397+IA, 114 pp.
- GRELL, G., S. EMEIS, W.R. STOCKWELL, T. SCHOENEMEYER, R. FORKEL, J. MICHALAKES, R. KNOCH, W. SEIDL, 2000: Application of a multiscale, coupled MM5/chemistry model to the complex terrain of the VOTALP valley campaign. – *Atmos. Environ.* **34**, 1435–1453.
- HOLZÄPFEL, F., 2003: Probabilistic two-phase wake vortex decay and transport model. – *J. Aircraft* **40**, 323–331.
- HOLZÄPFEL, F., T. GERZ, 1999: Two-dimensional wake vortex physics in the stable stratified atmosphere. – *Aerospace Sci. Technol.* **5**, 261–270.

- HOLZÄPFEL, F., T. GERZ, M. FRECH, A. DÖRNBRACK, 2000: Wake vortices in convective boundary layer and their influence on following aircraft. – *J. Aircraft* **37**, 1001–1007.
- HOLZÄPFEL, F., T. HOFBAUER, D. DARRACQ, H. MOET, F. GARNIER, G. C. FERREIRA, 2003: Analysis of wake vortex decay mechanisms in the atmosphere. – *Aerospace Sci. Technol.* **7**, 263–275.
- HOLZÄPFEL F., T. GERZ , M. FRECH, A. TAFFERNER, F. KÖPP, I. SMALIKHO, S. RAHM, K.-U. HAHN, C. SCHWARZ, 2007: The wake vortex prediction and monitoring system WSVBS. Part I: Design. 1st CEAS European Air and Space Conference, Berlin, CEAS-2007-177, 3383–3390.
- KONOPKA, J., H. FISCHER, 2005: The wake vortex warning system at Frankfurt Airport. – Proc. 24th Digital Avionics Systems Conf., Washington, DC, IEEE 3.A.6-31-14.
- MASSON, V., J.-L. CHAMPEAUX, F. CHAUVIN, C. MERIGUET, R. LACAZE, 2003: A global database of land surface parameters at 1-km resolution in meteorological and climate models. – *J. Climate*. **16**, 1261–1282.
- STEPHAN, K., S. KLINK, C. SCHRAFF, 2008: Assimilation of radar derived rain rates into the convective scale model COSMO-DE at DWD. – *Quart. J. Roy. Meteor. Soc.* **134**, 1315–1326.
- STEPPELER, J., G. DOMS, U. SCHATTLER, H. W. BITZER, A. GASSMANN, U. DAMRATH, G. GREGORIC, 2003: Meso-gamma scale forecasts using the nonhydrostatic model LM. – *Meteor. Atmos. Phys.* **82**, 75–96.
- THEUSNER, M., RÖHNER, P., in preparation: Comprehensive global assessment on weather risks and impact on delays. – Project Deliverable D 2.1-2 of the European FLYSAFE Project, 293 pp.
- WILKS, D.S., 1995: Statistical Methods in the Atmospheric Sciences. 2nd edition. – Academic Press.