Low visibility and ceiling forecasts at Schiphol

Part 1 – assessment of the current system

Koninklijk Nederlands Meteorologisch Instituut

Contents

Oktober 2007

Authors: Janet Wijngaard
Daan Vogelezang
Hans van Bruggen
Nico Maat

Photos: Peter de Vries

Lay-out: Kim Pieneman

Printed: Martin Heunks

1. Introduction

Accurate, reliable and unambiguous information concerning actual and expected low visibility conditions is very important for the available operational capacity at Schiphol airport. Improving the forecasts for low visibility procedure (LVP) events at Schiphol is therefore the main goal of this project.

KNMI has been responsible for the aeronautical meteorological service provision for the Netherlands for almost 70 years. Currently, several products for LVP conditions are issued for the use at the airport, which should be modified to meet the user-specific needs. Schiphol airport has specified a set of important thresholds for visibility and cloud base and implemented these limits in the so-called "Beperkt Zicht Operaties" (Dutch for Reduced Visibility Operations) or BZO phases. Table 1-1 gives an overview of the different BZO categories and the implications for the airport. The forecast system will now be optimised for these BZO phases. At the end of the project we will provide a new improved forecast service tailored to the specific needs of Airport Authorities, Air Traffic Control and KLM airlines.

Visibility classification	**Visibility/ RVR**	**Cloud base**		**Capacity [movements per hour]**	**Flow restrictions due to visibility**
Good	VIS≥5 km	and	≥1000 ft	68 arrivals or 74 departures max 104/ 108 movements	No flow restriction
Marginal	1.5≤ VIS<5 km	or	300 ft<CLB<1000 ft	Use of independent parallel runways required,	No flow restriction
BZO phase A	550 m≤RVR<1500m	or	200 ft≤CLB≤300 ft	50 arrivals or 50 departures max 70 movements	In general no Flow restrictions
BZO phase B	350 m≤RVR<550m	or	CLB<200 ft	44 arrivals or 50 departures max 60 - 70 movements	Flow restrictions in force
BZO phase C	200 m≤RVR<350m			17 arrivals and 30 departures max 47 movements	Flow restrictions in force
BZO phase D	RVR<200m			16 arrivals and 20 departures max 36 movements	Flow restrictions in force

Table 1-1: Impact of poor visibility and low cloud base on the capacity of Amsterdam Airport. BZO = Beperkt Zicht Operaties (Reduced Visibility Operations), RVR = Runway Visual Range, VIS = Visibility for aeronautical purposes, CLB = Cloud Base ("ceiling"), i.e. lowest cloud layer with at least 5/8 coverage.

Several refinements to the current automatic forecast system will be implemented to achieve this goal. Once these refinements are implemented, the forecaster has an objective tool to optimise the BZO forecasts. The refinements and extensions to be implemented are:

1) **The use of nearby upstream weather station data** for the synthesis of the statistical probability distribution for visibility. These stations are the so-called 'Fog detection sites' situated around Schiphol Airport in the neighbourhood of Nieuw-Vennep, Nieuwkoop, Muiden and Assendelft. The automatic observations from these sites will be used to calculate the probability distribution of visibility at Schiphol airport. In the current system other stations used upstream are about 50 km away from Schiphol. The new sites are much closer (approximately 20 km) and are proven to be important for forecasting small-scale phenomena like low visibility.

2) **The extension of the forecast system with joint probabilities** for combinations of cloud base and visibility parameters, which correspond to the thresholds of the BZO phases. With these modifications, objective joint probabilities will be offered to the forecaster.

3) **The conversion of the MOR forecast into RVR.** MOR (Meteorological Optical Range) is the only currently available model output for visibility to the forecaster. However RVR (Runway Visual Range) is in operational practice for low visibility conditions at the airport. This translation of predicted MOR values into RVR values is an important step towards tailoring the airport forecast. This procedure is not straightforward, because the determination of RVR not only requires MOR data but also both actual background luminance and runway lights intensity data. Therefore, creating reliable RVR predictions involves predicting background luminance as well (assuming that the runway light setting is known on forehand or fixed at 100%). So far, no research has been done to assess the possibilities of producing background luminance predictions with sufficient skill.

The extension using new upstream sites is planned to be implemented in autumn 2007, the more substantial improvements mentioned at point 2 and 3 will become part of the operational forecast beginning 2008.

Insight in the current prediction system will allow us in the sequel of the project, to evaluate the improvements made in perspective to the present situation. This interim report will concentrate on the current system and the key issues are: the existing forecast suite, the various visibility parameters, climatology and verification, i.e.

• The insight in the existing operational practice will give the opportunity to develop products confirming to the users needs in such a way new forecasts can be (easily) implemented in the operational suite (chapter 2).

• As mentioned above, the differences between the various visibility parameters are an important issue in forecasting the BZO phases. Chapter 3 will therefore discuss the differences between "RVR", "MOR" and "VIS" and their implications in forecasting.

• Climatological behaviour gives useful background information in forecasting the visibility and low clouds; the frequency of a weather event has consequences for its forecast, but also for verification. So climatological information about occurrences of low visibility and clouds will be provided. Conditional climatology, which can be regarded as a base line forecasting tool, will be considered in the second part of the climatology chapter (chapter 4).

• Evaluation of the current forecast system will set the performance of the current system. So, focus will be on the verification of the current forecast products. The different types of forecasts will be verified and their pro-s and con-s will be discussed. Finally, the user-specific advantages of these types of forecasts, especially the probabilistic one, will be highlighted (chapter 5).

2. Present forecast products

Aeronautical forecasts issued by KNMI concerning visibility and cloud base are:

- TAF for Schiphol (TAF: Terminal Aerodrome Forecast): both “short” and “long”
- SKV: Schiphol Kansverwachting (Dutch for probabilistic forecast): both “short” and “long”

The short TAF has a lead time of 9 hours and is issued in less than one hour before it becomes valid. The long TAF has a 24-hour lead time and is valid for the weather 7 hours ahead. Together, both TAFs cover a lead time up to 31 hours as soon as a long TAF is issued. Short TAFs are provided every 3 hours, and long TAFs are issued every 6 hours. In case of unforeseen significant changes in forecast conditions, an amended TAF is issued immediately. The SKV is introduced in 2003, because the TAF procedure did not fully meet the needs of the users at the airport and probabilistic forecasts are regarded as useful. The SKV has also a long and a short term variant, both with the same issue frequency as the TAF. The SKV gives probabilities for the operational BZO phases. Below, examples of the short term TAF and SKV are given. Under adverse weather conditions, e.g. low visibility, the “MAS” (Meteorological Adviser Schiphol) forecaster comes on duty. The presence of the MAS enables a direct contact with the stakeholders of Schiphol and leads to tailored information in situations where weather may limit the capacity of the airport. The MAS plays an important role in updating and fine-tuning the forecast for the “nowcasting period” (0-3 hours ahead) mainly by verbal communication.

Although with the introduction of the SKV a significant improvement was experienced, the current forecast suite does not meet all needs of the airport services. The TAF has a strict ICAO format and cannot be changed or extended by a national policy only. Even though the SKV is better modified to the specific needs of Schiphol, it still lacks important information e.g. for the lowest BZO phases.

Essential input for TAF and SKV is provided by the TAF Guidance (TAFG): An automatic statistical post processing application of the numerical weather prediction models running at KNMI. The output of the TAFG is provided to the KNMI aeronautical forecaster, who uses the TAFG to construct the local TAF and SKV. Currently the TAFG does not forecast the combined visibility and ceiling classes corresponding to the BZO phases. Therefore, a new TAFG will be created in-line with the BZO phases giving the forecaster a dedicated tool for the visibility and ceiling forecast at the airport.

KNMI KANSVERWACHTING SCHIPHOL

vrijdag 17 augustus 18-24 UTC

(Verstuurd: vrijdag 17 augustus 2007 15.10 UTC)

	18	**19**	**20**	**21**	**22**	**23**	**24**
Zicht < 5 km en/of wolkenbasis < 1000 vt (%)	0	0	0	0	0	0	0
Zicht < 1500 m en/of wolkenbasis < 300 vt (%)	0	0	0	0	0	0	0
Zicht < 550 m en/of wolkenbasis < 200 vt (%)	0	0	0	0	0	0	0
Windrichting (graden)	250	240	230	220	200	200	180
Windsnelheid (knopen)	10	9	8	7	7	8	8
Uitschieters (knopen)	16						
Standaarddeviatie windrichting (graden)	20	25	25	30	30	25	25
Standaarddeviatie windsnelheid (knopen)	2	2	2	2	2	2	2
Temperatuur (C)							
Dauwpunt (C)							
Kans op sneeuw (%)	0	0	0	0	0	0	0
Kans op matige/zware sneeuw (%)	0	0	0	0	0	0	0
Kans op onderkoelde neerslag (%)	0	0	0	0	0	0	0

Toelichtingen:

Zicht en wolkenbasis	
Wind	
Temperatuur/dauwpunt	
Winterse neerslag	

Example: SKV short

TAF EHAM 200305Z 200413 14003KT 6000 BR FEW006 SCT070 PROB40 0509
4000 -RADZ BR SCT004 BKN006 BECMG 0810 17007KT 9999 NSW SCT008 BKN012
BECMG 1012 SCT012 BKN015 PROB40 TEMPO 1113 6000 -SHRA SHRA FEW008
SCT012CB BKN015=

Example: short TAF

3. Three different visibility parameters

Three different visibility parameters can be distinguished: MOR (Meteorological Optical Range), VIS (Visibility for aeronautical purposes) and RVR (Runway Visual Range). MOR is a physical parameter which does not depend on background luminance and lamp settings. Both VIS and RVR are dependent on background luminance and lamp settings. They differ for the lamp settings.

All three parameters are used forecasting visibility: Visibility values mentioned in the TAF are "VIS" values. This parameter fits the visibility values referred to as 'Good' and 'Marginal' visibility classification in table 1-1. However, the BZO phases in table 3-1 depend on RVR values, which deviate from VIS values. The SKV provides probabilities for RVR values as well. Routine weather reports like SYNOPS use MOR.

During a bright day and visibility above 1000 m, there is no difference between the three visibility parameters MOR, VIS and RVR. As the amount of daylight and/or visibility decreases, the differences increase. In case of differences, RVR has the highest value, where MOR has the lowest value and VIS values are in between. As shown in figure 3-1, RVR values cannot be lower than MOR values, but under certain circumstances, the RVR value can be up to 5 times the MOR value. In table 3-1 the same information is given as in figure 3-1, but in a different format and with BZO phases highlighted.

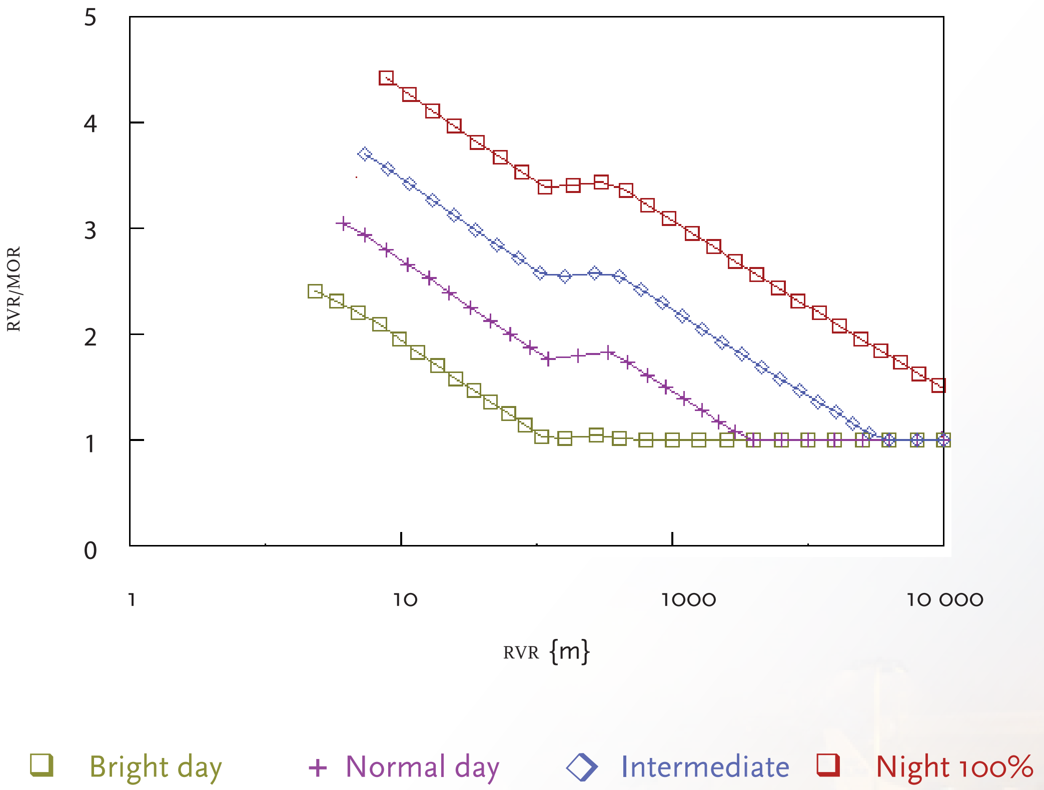


Figure 3-1: Ratio of RVR and MOR as a function of RVR for 4 different background luminance conditions: Brightday, Normal day, Dawn (Intermediate) and Night (Complete dark). Settings: 100%

RVR as a function of MOR and background luminance (cd/m^2)						
MOR (m)	**RVR (m)**					
	NA	**SD**	**SL**	**DD**	**DG**	**DH**
	BL<50	**50-300**	**300-1000**	**1000-4000**	**4000-12000**	**>=12000**
50	175	150	125	125	100	100
100	325	275	250	200	175	150
150	500	375	325	275	250	200
200	650	500	400	350	300	250
250	800	600	550	450	350	275
300	900	700	650	550	400	325
350	1000	800	700	600	500	375
400	1100	900	800	650	550	400
450	1200	1000	800	750	600	450
500	1400	1000	900	800	650	500
550	1500	1100	1000	800	700	550
600	1600	1200	1000	900	750	600
700	1800	1300	1200	1000	800	700
800	2000	1500	1300	1100	900	800
900	P2000	1600	1400	1100	900	900
1000		1800	1500	1200	1000	1000
1100		1900	1600	1300	1100	1100
1200		2000	1700	1400	1200	1200
1300		P2000	1800	1500	1300	1300
1400			1900	1500	1400	1400
1500			2000	1600	1500	1500
1600			P2000	1700	1600	1600
1700				1700	1700	1700
1800				1800	1800	1800
1900				1900	1900	1900
2000				2000	2000	2000
2100				P2000	P2000	P2000
	no BZO	phase A	phase B	phase C	phase D	

Table 3-1: MOR related to RVR, values for different background luminance (BL) conditions: Bright day (DH), Nominal day (DG), Dark day (DD), Dawn-light (SL), Dawn-dark (SD) and Night (NA). The colors indicate the BZO phases (see table 1-1); P2000 means RVR > 2000 m.

The TAFG output in the present form is not tuned to produce RVR values, as it is based on MOR values. So for the SKV visibilities, forecasters are supposed to estimate the risk of reaching BZO phases from objective input which does not contain RVR information, but MOR information instead.

4. Climatology

Climatology gives insight into the occurrence of weather events, e.g. visibility and ceiling, over a longer period, often 30 years. The climatology in this chapter is developed for LVP categories whereby the visibility data are MOR and VIS measurements and rounded down to multiples of hundreds of meters. The Low Visibility Procedure (LVP) categories are similar but not the same as the BZO phases, as the latter are based on RVR and have a resolution of 50 m. RVR observations are currently not available for a longer period and therefore not usable for climatological purposes. For more details concerning the LVP categories and observations, see box A.

LVP categories

The LVP categories used in 4. Climatology and in 5. Verification of this report are based on limits given in table A-1. Visibility observations are instrument measurements producing MOR . Cloud base/ceiling observations are carried out with a ceilometer (located along the runways at TD-22 and TD-27 for visibility and ceiling, respectively). Both are retrieved from archived SYNOP weather reports. For both variables hourly values are used. These hourly values are in fact the average over the last 10 minutes of the preceding hour. For verification purposes, in addition visually observed VIS measurements are used (retrieved from METAR archives -METeorological Aviation Routine report- weather reports). Note that the BZO phases used in operational practices at the airport are comparable but based on RVR and with slightly different thresholds.

Visibility classification	Visibility		Cloud base
Good	>= 5 km	and	>= 1000 ft
Marginal	1.5 ≤MOR/VIS<5 km	or	300 ft≤CLB<1000 ft
LVP phase A	600 m≤MOR/VIS<1500 m	or	200 ft≤CLB<300 ft
LVP phase B	400 m≤MOR/VIS<600 m	or	CLB<200 ft
LVP phase C	200 m≤MOR/VIS<400m		
LVP phase D	MOR/VIS<200m		

Table A-1: "LVP" (Low Visibility Procedure) categories based on MOR or VIS and Cloud Base; Ceiling (CLB).

BOX A

4.1. Climatology 1977 - 2006

To get an impression of the occurrence of LVP events, figure 4-1 presents for different LVP categories the number of days per year (at least one hour) LVP-conditions were met, averaged over the last 30 years. A significant decrease is seen for all LVP categories. This decrease is not unique for this particular area, but is found at several weather stations in the Netherlands. A single and clear reason cannot be given. It is highly probable that a combination of factors is involved. Maybe both artificial and natural causes are due to this trend, e.g. changed measurement methods, changing local environmental circumstances, changing aerosol concentrations, changing weather regimes and the more general climate variability/change. For a valuable forecast, developed by using climatological data, the circumstances of the historic measurements should be comparable to the current situation.

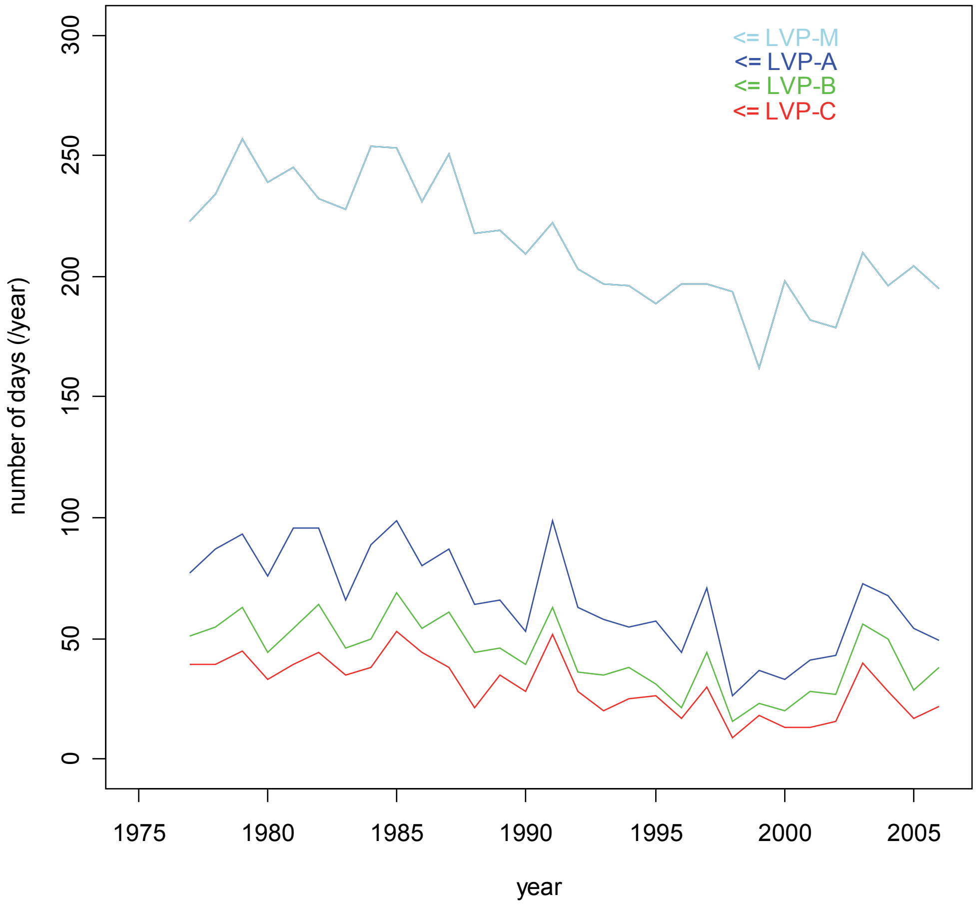


Figure 4-1: For different categories (conditions equal to or worse than LVP-M, LVP-A, LVP-B and LVP-C the number of days a year at Schiphol for the period 1977-2006. LVP categories: see box A.

The number of days with low visibility and ceiling are also seasonal dependent, which is clearly demonstrated in figure 4-2. Moreover, seasons have a high year-to-year variability for the pictured C/D category. The other categories (see appendix A) show similar high seasonal variability. In general, summer and spring show the lowest number of LVP conditions (5 and 6 C/D events per year respectively in the period 1977-2006), whereas autumn and winter have a higher frequency than on average (9 and 10 C/D events per year respectively in the period 1977-2006). The decreasing pattern as seen in the annual picture (figure 4-1) is also seen through the seasons (figure 4-2) although less convincing for some seasons.

The frequencies of issued LVP phases vary, not only from year to year and from season to season but are also clearly dependent on the time of the day. Figure 4-3 shows the diurnal cycle for wintertime. A clear minimum is seen in the afternoon, and the maximum is reached in the morning hours around sunrise. All categories behave in a similar way, only the absolute numbers differ.

The dependency on season and time of the day will be an important factor when developing forecasts for the BZO phases. Apparently different seasons and hours have quite different frequencies and variability, which have a relation to the degree of complexity of the forecast. Also, different mechanisms related to the various seasons might underlie the formation of fog and low cloud during the year.

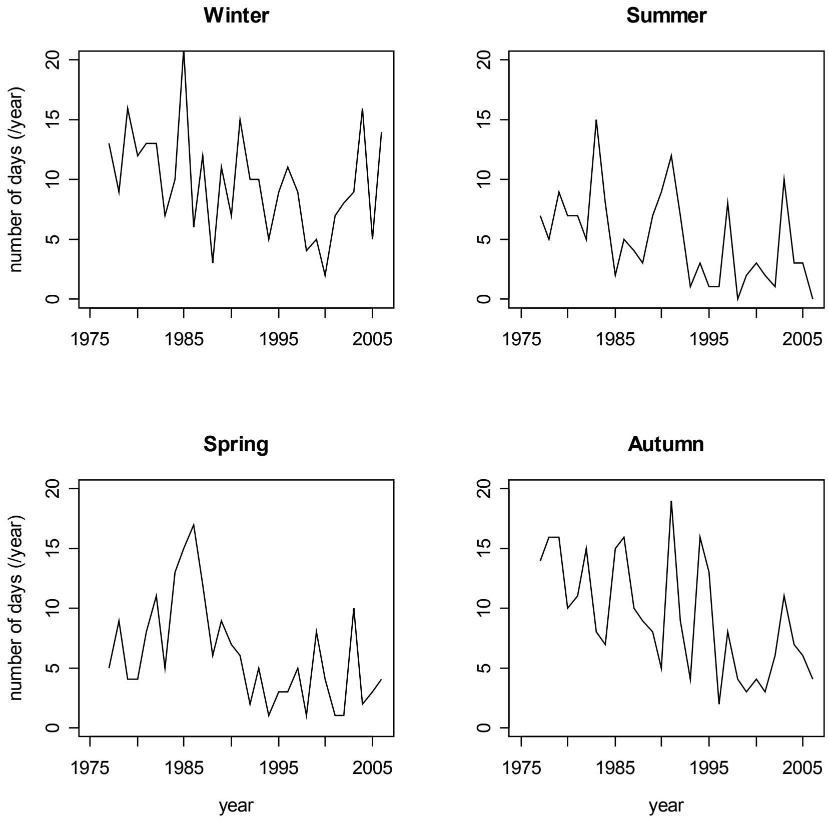


Figure 4-2: Number of days per year with LVP category C or D for the spring (mar-may), summer (jun-aug), autumn (sep-nov) and winter (dec-feb).

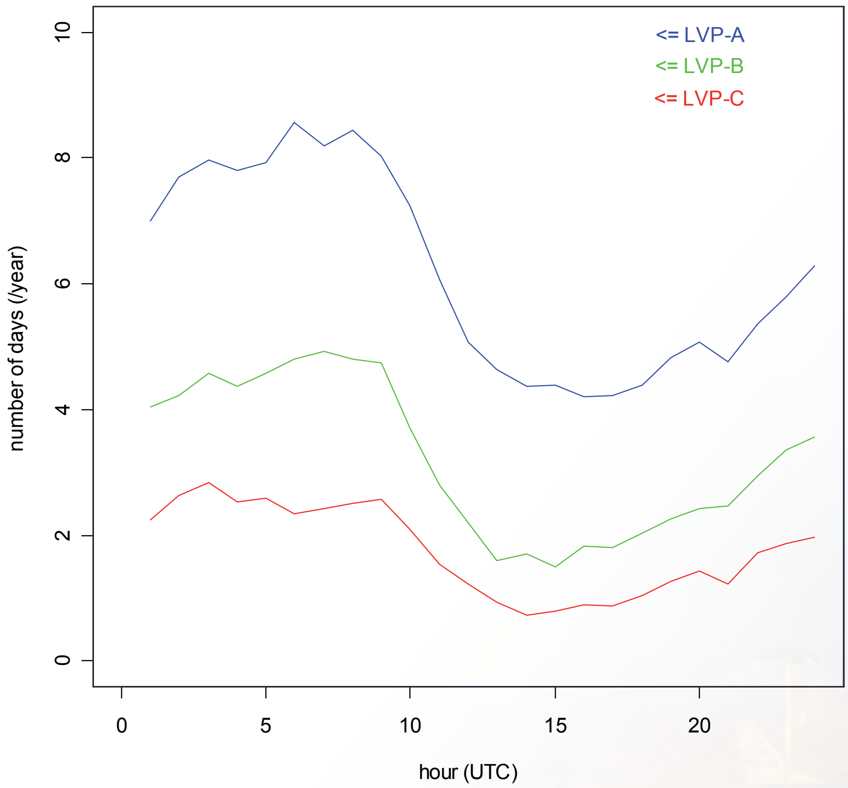


Figure 4-3: Hourly dependency different categories (conditions equal to or worse than LVP-A, LVP-B and LVP-C) during winter.

To stress the dominancy of the visibility component in the LVP categories, figure 4-4 shows the LVP categories A-C and corresponding categories based on low visibility only. The solid lines subtracted from the dashed lines (same color) provide the number of events for which only low clouds are responsible for being in the particular category.

Only in a few cases the clouds alone are responsible for triggering a certain category. The visibility category with a MOR less than 400 m falls together with the dashed line of C, as category C is only dependent on the horizontal visibility.

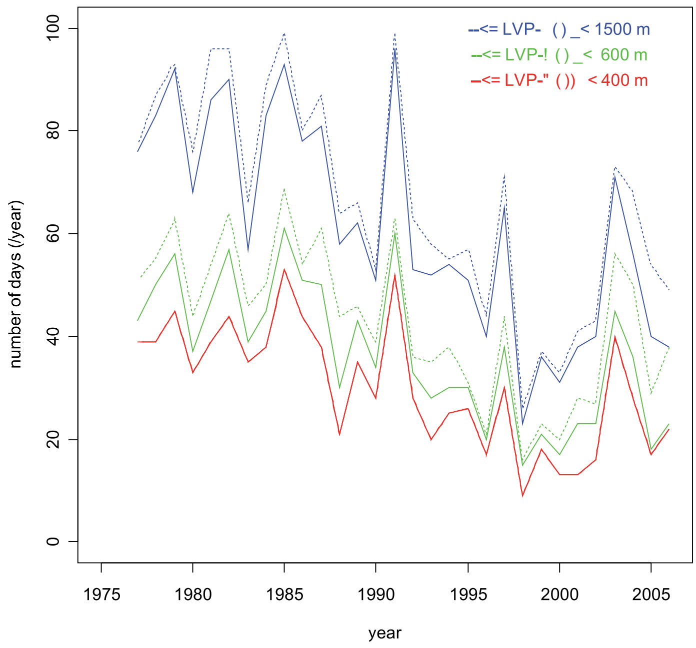


Figure 4-4: Number of days per year for different LVP categories (dashed, <=LVP-A, <=LVP-B and <=LVP-C) and the corresponding visibility categories (solid, MOR <1500 m, <600 m, < 400 m), period considered: 1977-2006. <= LVP-x means condition equal to or worse than LVP-x.

4.2. Conditional Climatology

Conditional climatology can be considered as a pre-assumed base line forecast once a low visibility/ceiling event has started: Under low visibility circumstances, an important question for further operations is: when does this situation change and how does it change? Conditional climatology can give a first answer.

For conditional climatology location, time of the year, hour of the day and of course the starting conditions are relevant. A clear example is shown in figure 4-5, valid for January and March at the airport Schiphol with a starting condition at 6 UTC of category LVP-B (visibility between 400 and 600 m or ceiling below 200 ft). The star indicates the time of sunrise (15th of the month). Clearly, in January the chance is about 50% that after one or two hours the visibility is improved and with only a slight possibility for worse conditions. In March the situation is different. Although there are fewer events in the 30-year period considered, but once you start in category LVP- B at 6 UTC the chance is 50% to deteriorate to category C or D in the first hours after. A reason for this phenomenon is that after sunrise, the sun is often not strong enough to dissipate the fog, but strong enough to slightly warm the surface and therefore inducing some mixing and evaporation, resulting in a denser fog. This effect of deteriorating visibility after sunrise is also seen in January, but less strong. Also if in January the begin condition of category LVP-B is chosen around sunrise (figure 4-6), the effect of deteriorating visibility during the first hours after sunrise is not comparable to March. As this effect is most pronounced around March and October, the forecasting around sunrise is very delicate in these periods.

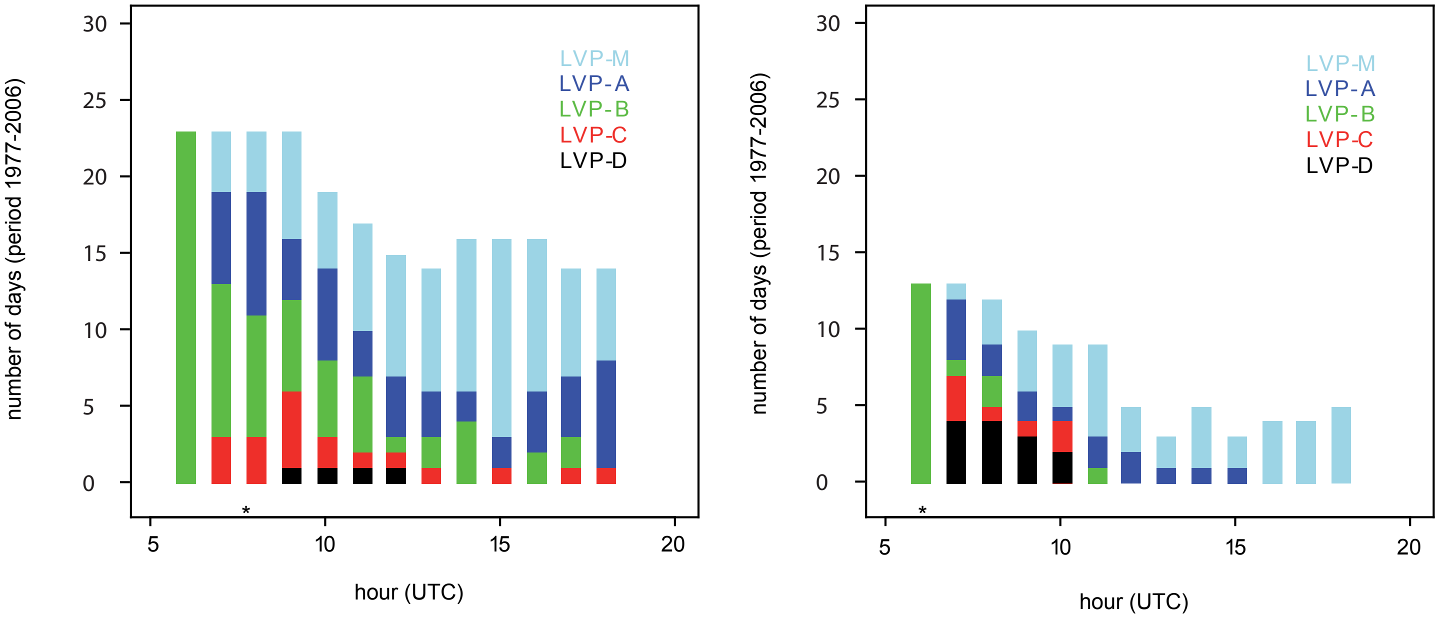


Figure 4-5: Conditional climatology for Schiphol starting at 6 UTC in January (left) and March (right). Initial condition is LVP-B (green). The star indicates the time of sunrise.

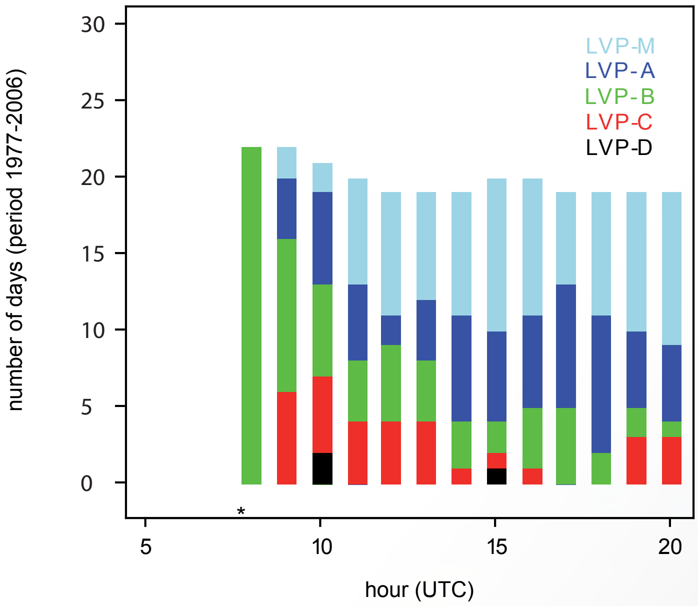


Figure 4-6: Conditional climatology for Schiphol starting at 8 UTC in January. Initial condition is LVP-B (green). The star indicates the time of sunrise.

5. Assesment of the current forecast system

This chapter discusses the reliability and accuracy of the present forecast system by quantitative analyses. The results will allow us in the next part of the project to compare the performance of the enhanced product to the current product.

To verify the forecasts issued at present (TAF, TAFG and SKV), observations from the SYNOP (MOR) and METAR (VIS) weather reports are used. Details about the used categories and observations are presented in box A of the previous chapter. The focus is on the forecasts valid for 06 UTC - i.e. during the inbound peak at Schiphol (local time is UTC+1 or UTC+2). Both long term forecasts (about 15 hours ahead) as well as short term forecasts (about 3 hours ahead) are considered. These are the forecasts used in operational practice at the airport during the briefing in the afternoon the day before and the briefing a few hours ahead around 03:30 h local time (01:30 or 02:30 UTC). Probabilistic as well as deterministic forecasts are verified. Probabilistic forecasts can support a scenario approach, whereas deterministic forecasts are more convenient in yes/no decisions. For more details on the forecast types see box B. In the next paragraph, the probabilistic forecast will be evaluated. In paragraph 5.3, the construction and verification of the deterministic forecast will be presented.

Probabilistic and deterministic forecasts

Uncertainty is inherent to weather forecasting. This uncertainty can be presented in an objective form in so-called probabilistic weather forecasts. Probabilistic forecasts do not give a forecast value for a certain element but give a probability (%) that a certain (fixed) threshold will be exceeded. E.g. the probability that at the airport for 6 hours ahead the visibility will be less than 800 m is 30 %. Forecasts for a series of exceedance thresholds for an element then give a forecast probability distribution. A deterministic forecast gives a single prediction of a certain weather element (e.g. the visibility in meters) a certain time ahead. Stating 6 hours ahead that the visibility at the airport will be 800 m at 06 UTC is a deterministic forecast. The outcome of this forecast may be right (when the observed visibility is exactly 800 m) or wrong (when the observed value is lower or higher).

The value of the forecast depends on the error-tolerance stated by the user of the forecast. If one accepts an error of 50 m then the outcome may lie between 750-850 m, if the error may be only 5 m then the outcome has to lie between 795 and 805 m.

Often a deterministic forecast is drawn from the forecasted probability distribution. Within this step, information about the uncertainty in the forecast is lost which makes it a user-dependent step, since the assessment of the uncertainty in the forecast may be different for different users. Many extreme weather events are characterized by a low probability of occurrence, but with a high risk. Deterministic forecasts often conceal this information. For an improved decision support these low probability/high risk scenarios should be revealed to the user.

BOX B

5.1. Verification of the probabilistic forecast

The quality of probabilistic forecasts can be expressed in terms of statistical parameters, called skill scores. A basic constraint is that for a sufficiently large number of events the forecasted probabilities should correspond with the observed frequency (i.e. reliable). In figure 5-1 a reliability diagram is presented enabling a comparison between the short term TAFG and the SKV performances. The verification of LVP events (LVP-A or worse), is performed against VIS observations (METAR); all issue times and forecast times are selected. Ideally the points in the left panel should be on the diagonal. The short term SKV gives reliable forecasts up to 100 %, and the short term TAFG has a comparable reliability but does hardly exceed the 80 %. This means that the forecaster has added value by making a more distinct forecast. The wiggling of the red line in the right panel is due to the preference of the human forecaster to round off probabilities to 10 percent values. The automatic TAFG system does not suffer from this habit.

Figure 5-2 is similar to the previous figure but in this picture the longer forecast ranges of the SKV and TAFG are evaluated. Strong over-forecasting by both TAFG and SKV is seen; forecasted probabilities are too high. Especially, forecasted probabilities of the TAFG above 30 % are suffering from this over-forecasting. This is partly due to the fact that the verification is against VIS observations, while TAFG in fact predicts MOR values. During (winter) nighttime the verifying visibilities values (VIS) are in general higher than the MOR visibility making the forecast too pessimistic; the probabilities for low visibility should be lower. This effect is not seen for the LVP category M (see appendix D) since MOR and VIS observations are in general identical for high visibility values. Note that RVR values (for visibility less than 1.5 km) are required for the airport operations. If the verification of the TAFG is carried out against RVR values the over-forecasting is even stronger. Clearly, extension of the TAFG is necessary to hand the forecaster a better objective forecast tool for RVR values.

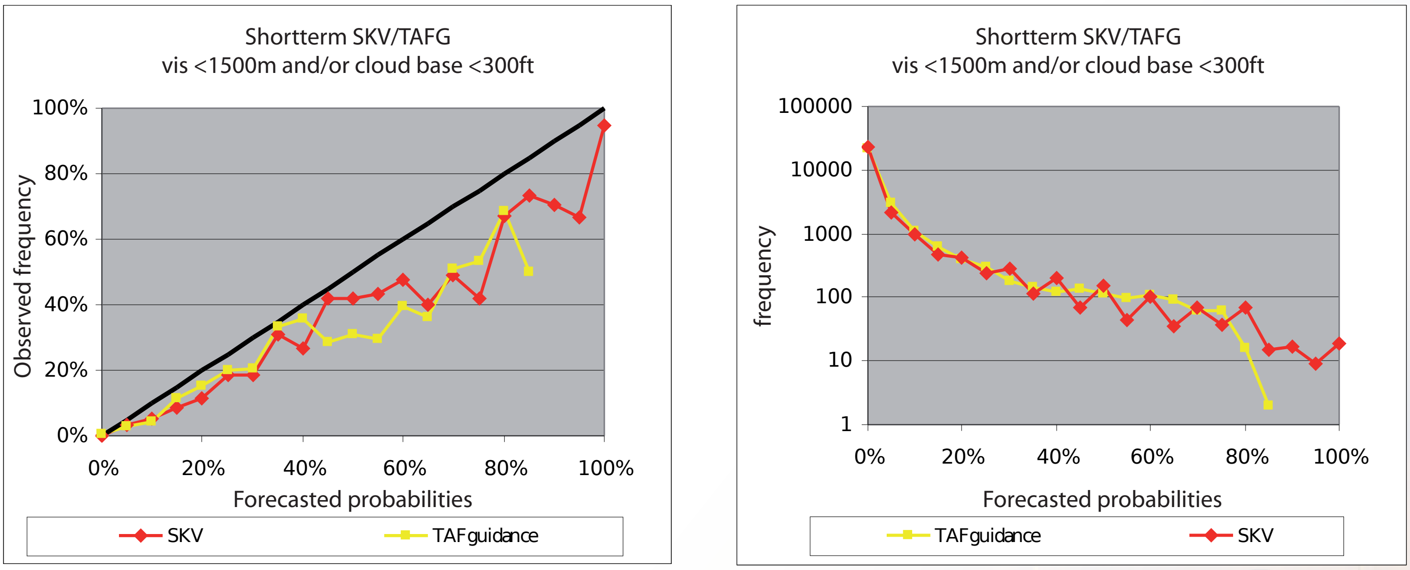


Figure 5-1: Reliability diagram (left) and frequency graph (right) for LVP categories with conditions equal to or worse than LVP- A. Short term SKV is compared to the TAFG and verified against VIS values. All issue times and lead times are taken together to enlarge the sample.

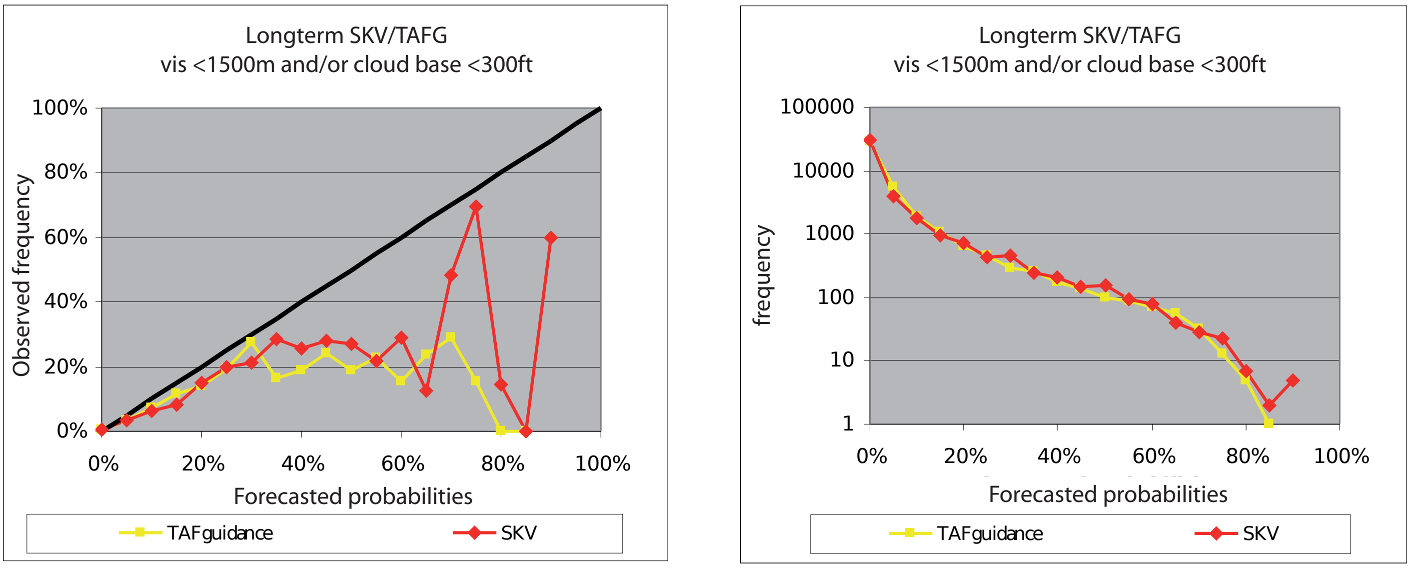


Figure 5-2: Reliability diagram (left) and frequency graph (right) for LVP *categories with conditions equal to or worse than* LVP- A*. Long term* SKV *is compared to the* TAFG *and verified against* VIS *values. All issue times and lead times are taken together to enlarge the sample.*

Another way to evaluate probabilistic forecasts is to analyze the systematic error (bias) and skill scores with the Brier Skill Score. Skill scores express the quality of the forecast by comparing to a defined reference forecast, in this case the climatological probabilities. A Brier Skill Score of 100 % indicates a perfect forecast and 0 % is the same skill as climatology and a negative one means that the reference climatological forecast is better. In figure 5-3 the Brier Skill Score (upper lines - solid) and bias (lower lines - dotted) are shown as a function of lead time and visibility threshold. The TAFG visibility probability forecast of 02 UTC is verified against MOR values as the forecast is designed to predict MOR. The skill drops with lead time whereas no systematic errors are present, as the bias is around zero. In general, the skill decreases with decreasing visibility threshold, in other words: the lower the visibility the more difficult it is to predict. Below 800 m the skill rapidly approaches the "no skill line" (0 %). Figure 5-4 shows that the pattern for ceiling is similar to that for visibility. Skill drops sharply below the 500 ft. A negative bias (about 10 %) in ceiling probabilities is found for high thresholds (under-forecasting) and only a small bias for lower ceiling thresholds. Figures 5-5 and 5-6 for the long term underline the difficulty of forecasting low visibility and ceiling a long time ahead.

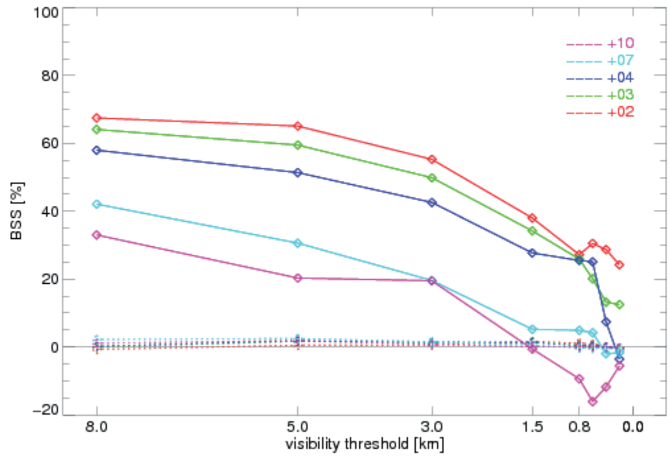


Figure 5-3: Brier Skill Score (BSS) *for visibility as a function of the threshold (upper lines). The scores are given for the short* TAFG *with issue time 02* UTC *(valid from 04 to 13* UTC*). Different colors represent different lead times. Dotted lines near 0 represent the bias.*

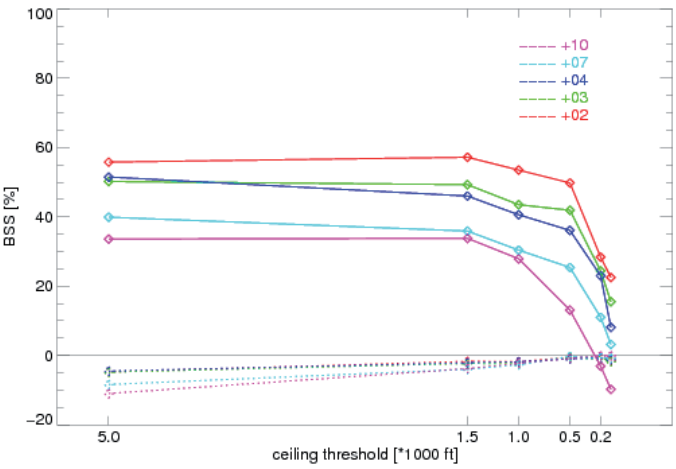


Figure 5-4: Similar as figure 5-3, but for ceiling.

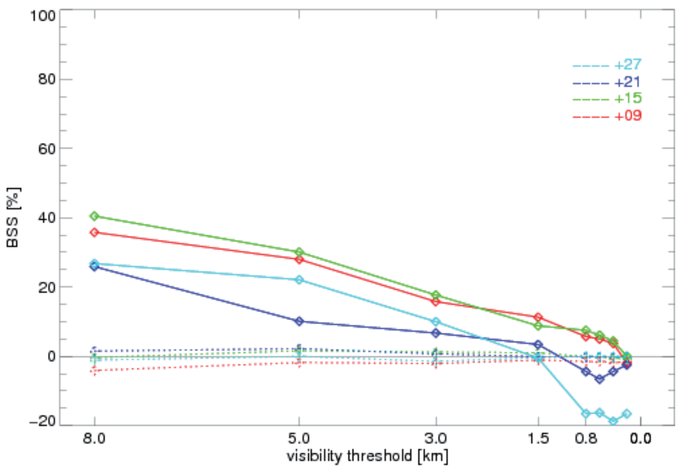


Figure 5-5: Brier Skill Score (BSS) *and bias (dotted lines) for the long term forecast (issue time: 15* UTC*) for visibility.*

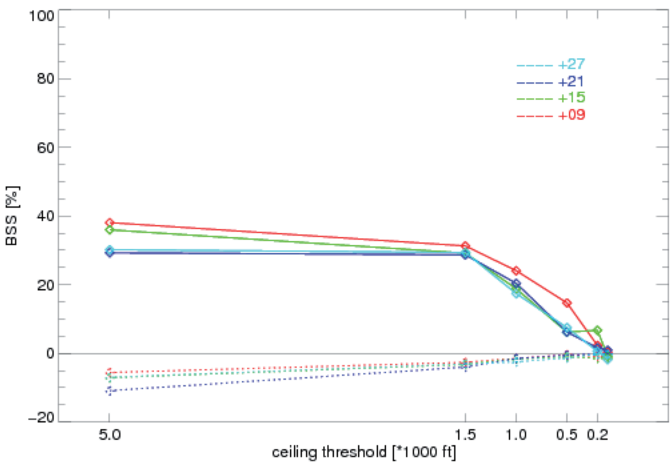


Figure 5-6: Similar as figure 5-5, but for ceiling.

5.2. Construction and verification of the deterministic forecast

The capacity at the airport heavily depends on the forecasted visibility and ceiling. The user has two options regarding the use of the forecast for these elements:

- Use the probability forecast utilizing the whole distribution.
- Use a deterministic forecast (derived from the probability distribution), although part of the information is lost.

		MOR (m)				
CLB (ft)		0 - 400	400 - 600	600 - 1500	1500 - 5000	>= 5000
	0-200	C/D	B	B	B	B
	200-300	C/D	B	A	A	A
	300-1000	C/D	B	A	M	M
	>=1000	C/D	B	A	M	G

Table 5-2: LVP categories dependent on visibility (MOR) and Ceiling (CLB).

For the second option KNMI derives a deterministic visibility forecast from the complete probability distribution of visibility. Table 5-1 shows a probability distribution of the TAFG forecast for visibility. From this forecast the deterministic visibility value at the 50th percentile is 1800 m. In other words the probability is 50 % that the visibility is less than 1800 m.

Visibility (m)	Probability (%)
< 8000	85
< 5000	73
< 3000	70
< 1500	45
< 800	23
< 600	18
< 400	8
< 200	2

Table 5-1: Example of a probability distribution for visibility from TAFG.

The same can be done to get a deterministic value for ceiling. The combination of these deterministic visibility and ceiling values determines the final forecasted LVP category (table 5-2). The categories in table 5-2 are related to the available thresholds in the TAFG forecast system. Note that these categories are very similar but not identical to the official BZO phases.

The forecasts can be matched with the actual observed LVP category, i.e. the combination of the observed visibility and cloud base. This results in pairs of forecast and observation of deterministic LVP categories. These pairs are displayed in a 5x5 performance matrix (see e.g. table 5-3). Table 5-3 shows the number of FC-OBS pairs for all forecasts from the TAFG of 02 UTC valid at 06 UTC. Note that categories C and D are taken together in "C". In total 1558 forecasts are used, where pairs at the diagonal are in the correct class. The further away from the diagonal, the worse the forecast is. Below the diagonal worse conditions are forecasted than observed, which can be interpreted as false alarm (red). Above the diagonal the forecast is too optimistic, which can be interpreted as a miss (blue). Most pairs fall in the GG cell (good visibility forecasted and observed), which is for the capacity planning at the airport of course the least interesting one. For this example only 2 pairs fall in the extreme CG cell and none in the GC cell. Table 5-4 shows the results when the deterministic values for visibility and ceiling are determined at 25th percentile from the probability distributions. Compared with the 50th percentile, the P25 may be called a "pessimistic" choice: there is only a 25% probability that the visibility will be even lower. The difference between tables 5-3 and 5-4 is remarkable. Events tend to shift to the left in the matrix with lower threshold percentiles; overall resulting in less misses but more false alarms.

Verification of LVP category A or worse

An important goal in our verification is to determine the quality of the forecasts, keeping in mind it should be tried to decrease the number of false decisions regarding the airport capacity. Flow restrictions come in force when the BZO phase "A or worse" is observed. We therefore now concentrate on the forecast that the LVP conditions will be worse or equal to A. This is a yes/no type forecast (for more details on this topic see appendix C). Then table 5-3 can be summarized into table 5-5 (where cells are added). Note that in table 5-5 the values still depend on the 50 % threshold percentile to construct the deterministic forecast for visibility and ceiling. Also, several verification scores for the events in table 5-5 are calculated. The hit rate of 0.49 means that for all situations that LVP A or worse conditions were observed at 06 UTC, the TAFG of 02 UTC also forecasted these conditions at 06 UTC in 49% of the cases. Of course then the other 51% of the forecasts were LVP-M or LVP-G. The false alarm ratio (FAR) of 0.45 on the other hand says that in 45% of the situations that LVP-A or worse conditions were forecasted for 06 UTC it did not occur, and in fact LVP-M or LVP-G was observed. Finally a CSI (Critical Success Index) of 0.35 can be interpreted such that of all situations that LVP-A or worse conditions were forecasted and/or observed 35% was forecasted right. Thus CSI does not take into account the non-interesting cases where the event was not observed and not forecasted.

The same exercise can be done with the 25th percentile threshold (see table 5-6). The HR (Hit Rate or Probability of Detection) and CSI are then clearly improved compared to the 50th percentile scores, on the other hand at the same time the FAR is somewhat higher.

		Forecast					
		C	B	A	M	G	
Observation	C	6	5	4	10	2	27
	B	5	10	6	13	3	37
	A	1	2	3	14	2	22
	M	3	9	16	137	116	281
	G	0	2	4	77	1108	1191
		15	28	33	251	1231	1558

Table 5-3: Verification matrix of deterministic LVP categories (50th percentile from probabilistic distribution). TAF guidance +4 (lead time: 4 hours) issued 02 UTC. Verification period: 1 January 2003- 30 April 2007. Forecast (FC) category runs from left to right, observations (OBS) top-bottom. Values with colored background are the totals per row/ column. Black numbers: FC and OBS in same class; blue: FC higher than OBS ("miss"); Red: FC lower than OBS ("false alarm").

		Forecast					
		C	B	A	M	G	
Observation	C	10	6	4	6	1	27
	B	11	8	13	4	1	37
	A	1	5	4	10	2	22
	M	11	10	34	172	54	281
	G	1	2	18	180	990	1191
		34	31	73	372	1048	1558

Table 5-4: Verification matrix of deterministic LVP categories extracted from the 25th percentile from the probabilistic distribution. Further same as table 5-3.

		Forecast		
Observation		yes	no	
	yes	42	44	86
	no	34	1438	1472
		76	1482	1558

HR = (42/86) = 0.49
FAR = (34/76) = 0.45
CSI = (42/120) = 0.35

Table 5-5: Verification of deterministic LVP categories. LVP-A or worse conditions as taken from the probability distribution at the 50th percentile (P50). TAF guidance +4 (lead time: 4 hours) issued 02 UTC. Blue values are misses, red are false alarms and black are the correct forecasts

		Forecast		
Observation		yes	no	
	yes	62	24	86
	no	76	1396	1472
		138	1420	1558

HR = (62/86) = 0.72
FAR = (76/138) = 0.45
CSI = (62/162) = 0.35

Table 5-6: Similar as table 5-5, for the 25th percentile.

In general HR, CSI and HKS (Hanssen-Kuipers Score) decrease with lead time and FAR increases with lead time. Also verification scores get worse as visibility decreases (from M-> C). So, the more interesting the LVP category, the more difficult it becomes to get good scores. Most verification scores (except HKS) depend strongly on the climatological occurrence of the event and therefore the scores for the different categories can not be directly compared to each other. In appendix E verification scores are given for different lead times and LVP categories for the TAFG issued at 02 UTC (short) and 15 UTC (long).

Dependence on percentile threshold choice

As shown before, the verification scores depend on the threshold-percentiles for visibility and ceiling that determine the forecasted LVP-class (see e.g. the difference between tables 5-3 and 5-4). We can explore this further by calculating some categorical scores (see appendix C) for a varying threshold-percentile. An example is given in figure 5-7. A clear feature is a strong coupling between hit rate and false alarm rate that both run from high values for low thresholds to low values for higher ones. CSI, which takes into account both the hits, false alarms and the misses, shows a maximum at a threshold of about 30 %, although in the range of 20 % to 50 % the CSI is rather insensitive to the chosen threshold percentile. The maximum HKS appears around 10 %. This suggests that the "best percentile" depends on the chosen measure.

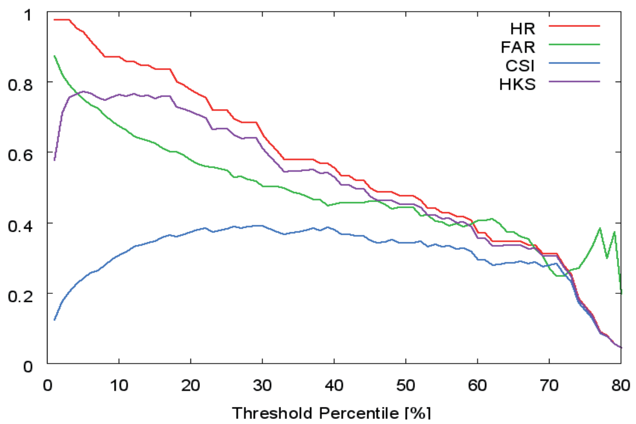


Figure 5-7: Categorical scores for the event "LVP-A conditions or worse". Applied to the +4 of the TAFG, issued at 02 UTC. For the definitions see appendix C.

Figure 5-8 shows the CSI for 4 categories. For conditions worse or equal to LVP-M the CSI just touches 0.6, but the general conclusion is again that less frequent events have lower scores and are thus more difficult to forecast. Also the "best percentile" is not constant for the 4 events. It decreases from about 40 % for conditions worse or equal to LVP-M to about 15 % for LVP-C/D. This introduces a serious problem: assume one has chosen CSI as the main verification measure. Then the choice for the "best percentile" is not unambiguous; it depends on the event of interest!
A possible solution for this dilemma is presented in the next paragraph.

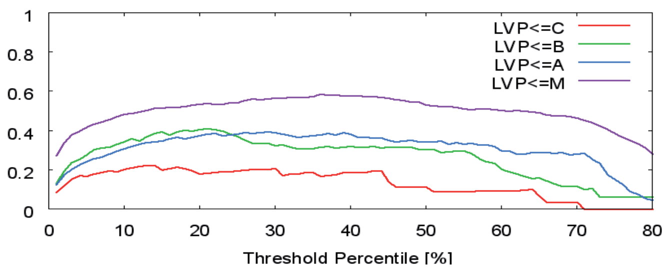


Figure 5-8: CSI as a function of threshold for 4 categories. Applied to the +4 of the TAFG, issued at 02 UTC.

Using the expense matrix

In the previous paragraph, all cells in the 5x5 matrix are assumed to have equal weight. In practice however, pairs in the outer corners are often much more costly than events closer to the diagonal. In addition, it is possible that e.g. misses are much more expensive than false alarms (asymmetric costs). In fact to every cell in this matrix “extra costs or damage due to wrong forecast” should be assigned, assuming that cells on the diagonal are perfect forecasts which give no extra costs. In cases of false alarms the user has unfortunately taken an unnecessary measure, which has cost him a certain amount of money. On the other hand there are situations which are missed by the forecast and the user will suddenly be confronted with low visibility. These “miss” situations lead also to extra costs and may be a factor 10 higher than for “false alarms”. Multiplying the 5x5 verification matrix (e.g. table 5-3) by an *expense matrix* (table 5-7) and summing all cells of this multiplication gives then the user-specific extra cost.

	Forecast				
Observation	C	B	A	M	G
C	0	0.4	2	20	100
B	0.2	0	0.6	10	40
A	6	0.6	0	0.8	10
M	20	2	0.2	0	2
G		10	2	0.1	0

Table 5-7: Example expense matrix (index) for LVP classes.

Subsequently, the extra costs dependent on different percentile-thresholds can be calculated. Then figure 5-9 shows a minimum in the extra costs for a threshold of 20 %. Clearly this minimum in costs and the maxima of e.g. the CSI do not lead to the same percentage. But more important: the analysis based on the expense matrix takes the complete probabilistic forecast into account, and not only a single event as was done on basis of the deterministic scores.

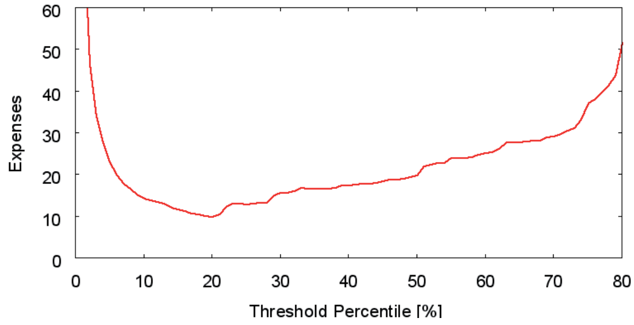


Figure 5-9: Extra costs/expenses due to errors in the forecast as function of threshold percentile, assuming the expenses as given in table 5-7. Based on the +4 of the TAFG, issued at 02 UTC.

Of course for every user the values in the expense matrix are different as the minimal costs are. This concludes that for the same forecast different users should take their yes/no decisions at different thresholds to have the most benefit from the forecast. Although filling the expense matrix will not be an easy task, it is a good way to optimize the visibility/ceiling forecast to the user-specific needs.

Concluding remarks

The low frequency of low visibility and ceiling has implications not only for the forecast, but also for the verification. Most verification scores depend strongly upon the climatological frequency of occurrence.

Two types of forecasts have been distinguished:

- The conditional climatology which gives insight in the change of the visibility and ceiling circumstances given a certain starting condition. This can be considered as a base line forecast.
- The operational forecasts: SKV and TAF, which contain probability and deterministic forecasts.

The current operational forecast product as SKV and TAF do not fully meet the user needs.

The TAFG is essential input for the operational forecast products. Optimizing this tool of the forecaster will lead to a dedicated tool for visibility and ceiling forecasts.

There are different ways to produce a capacity forecast for the airport:

- Construct deterministic weather forecast from the probabilities of visibility and ceiling. Subsequently, determine the corresponding deterministic capacity forecast.
- Use the probability distribution of visibility and ceiling to compute a probabilistic capacity forecast. Next, derive from the probabilistic capacity a deterministic capacity prognosis. For this purpose, a method analogous to the method in chapter 5.2. (using an expense matrix) could be applied.

Advantage of probabilistic weather forecasts: for every user the best decisions can be made considering the whole probability distribution and based on their own expenses. So, the final resulting deterministic outcome is tailored to the user-specific needs, whereas the underlying probabilistic weather forecast is the same for all users.

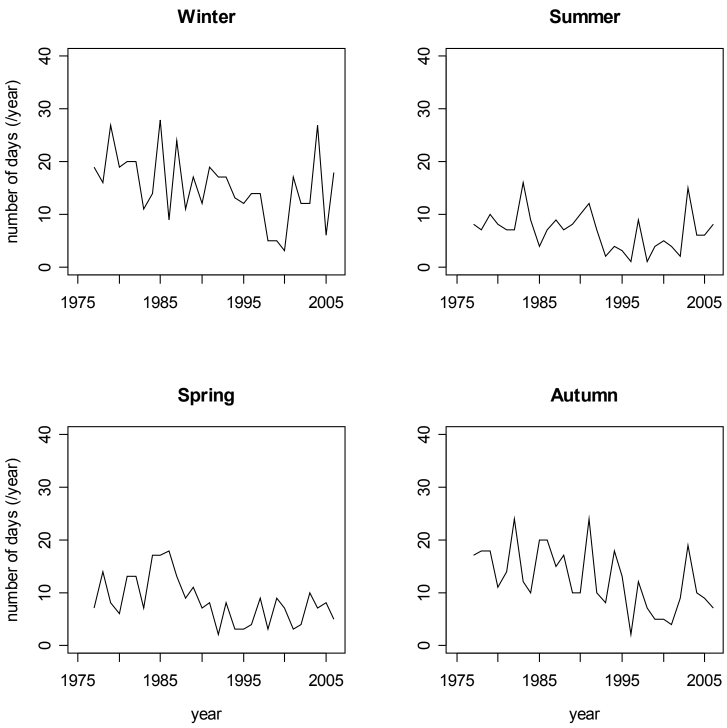
Future: A new forecast system is currently being developed. The evaluation of the new products and implementation will be a next step in the project. Also the comparison with (RVR) observations will be part of this. In part 2 of this report these steps will be described.

Acronyms

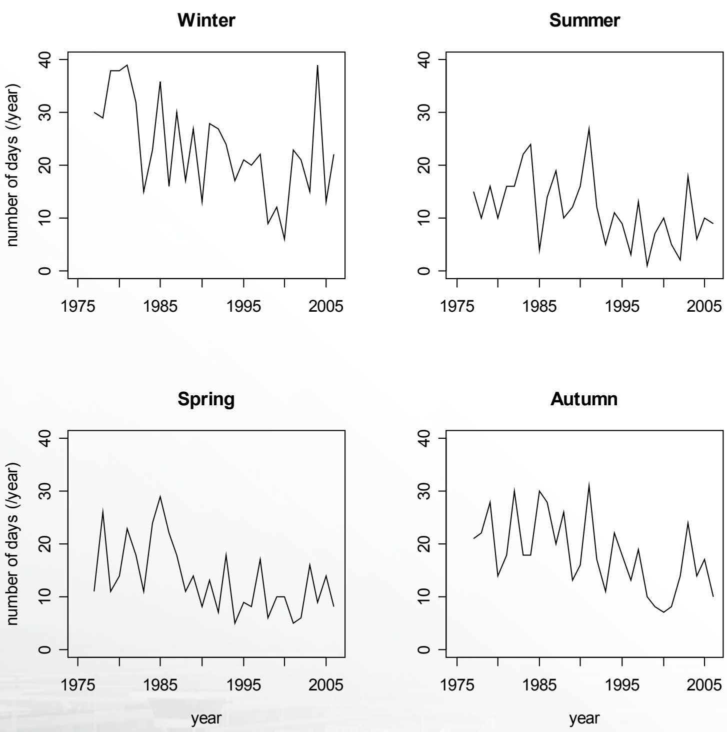
BSS	Brier Skill Score
BZO	Beperkt Zicht Operaties (Reduced Visibility Operations)
CLB	Cloud Base; "ceiling"
CSI	Critical Success Index
FAR	False Alarm Ratio
FC	Forecast
HKS	Hanssen-Kuipers Score
HR	Hit Rate (also Probability of Detection, POD)
ICAO	International Civil Aviation Organization
KLM	Koninklijke Luchtvaart Maatschappij
KNMI	Koninklijk Nederlands Meteorologisch Instituut
LVP	Low Visibility Procedure
MAS	Meteorological Adviser Schiphol
METAR	Meteorological Aviation Routine weather report
MOR	Meteorological Optical Range
OBS	Observation
POFD	Probability of False Detection; False Alarm Rate
RVR	Runway Visual Range
SKV	Schiphol Kansverwachting
SYNOP	Synoptical weather report
TAF	Terminal Aerodrome Forecast
TAFG	TAF Guidance
VIS	Visibility for aeronautical purposes

Appendix A

Seasonal climatology LVP category <=B (upper 4) and <=A (lower 4)



<= LVP-B



<= LVP-A

Appendix B

Probabilistic Verification Measures

Reliability of the forecast means that when an event has e.g. a forecasted probability of 30 %, in fact in 30 % of these cases it really happens and in 70 % of these cases it does not. Of course, the contribution to the error is large when a forecast that an event will happen is 99 %, but it does not occur (or vice versa). On the other hand, given a forecast of 99 %, nothing is wrong with 99 occurrences out of a set of 100 forecasts.
In a reliability diagram the forecasted probabilities are plotted (in bins) against the observed frequency. In the most ideal case, all points should lay on the x=y diagonal. Points below the diagonal suffer from over-forecasting (probabilities too high) and points above the diagonal indicate under-forecasting (probabilities too low).

Conditional on reliability, a forecast should also have **resolution.** Forecasts with a good resolution are as close as possible to 0 % or 100 %. In the reliability diagram this leads to many data points in the lower left corner and/or in the upper right corner.

The Brier Score (BS) is the mean-square-error which is a common verification measure. For a certain dichotomous (yes/no) event it assesses the mean (squared) difference between the forecasted probability of the event and the observation (event happened = 1, not happened = 0).
$BS = 1/N\sum_{i=1}^{N}(P_f - P_o)^2$ where P_f is the forecasted probability, P_o the observed value (0 or 1) and N the number of cases.
BS=0 is perfect and BS=1 the worst.

The Brier Skill Score (BSS) expresses the quality of the forecast of an event relative to a reference forecast. In this project the sample climatological probabilities are used for the reference forecast. Maximum BSS = 100 %. When BSS drops below 10 % the value added by the forecast relative to climatology becomes marginal.

Appendix C

Categorical Verification Measures

With a 2-category categorical (yes/no) forecast an event happens or does not happen. It can be evaluated using a 2 by 2 contingency table (see below).

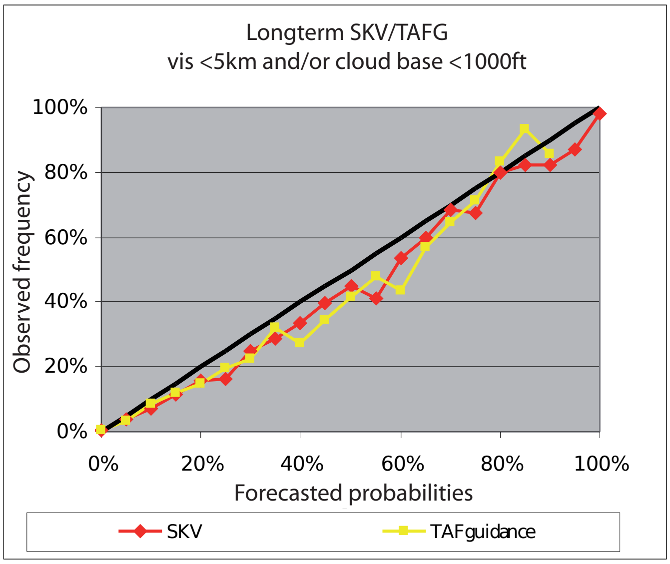
		Forecast	
Observation		yes	no
	yes	Hit A	Miss B
	no	False Alarm C	Correct Negative D

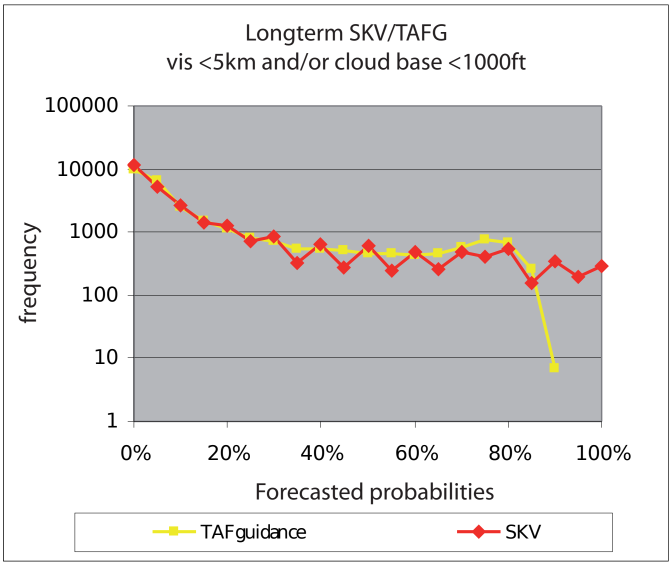
Different scores can be derived from this table:

- Hit Rate or Probability of Detection (HR/POD; A/A+B) The HR gives the fraction of the observed "yes" events which were correctly forecasted.
- False Alarm Ratio (FAR; C/A+C) The FAR gives the fraction of predicted "yes" events that actually did not occur (i.e. were a false alarm).
- Probability of False Detection (POFD; C/C+D) The POFD gives the fraction of wrong forecasts given the event did not occur.
- Critical Success Index (CSI; A/A+B+C) The CSI measures the fraction of observed and/or forecasted events that were correctly predicted.
- Hanssen-Kuipers Score (HKS; HR– POFD) The HKS indicates how well the forecast separates the "yes" events from the "no" events. Note that all 4 table elements are used. HKS does not heavily depend on the climatological frequency of the event, where the others do. Rewards the hit rate and at the same time punishes false alarms. Finally, HKS is a skill score, which means it compares to a reference score (the unbiased random forecast).

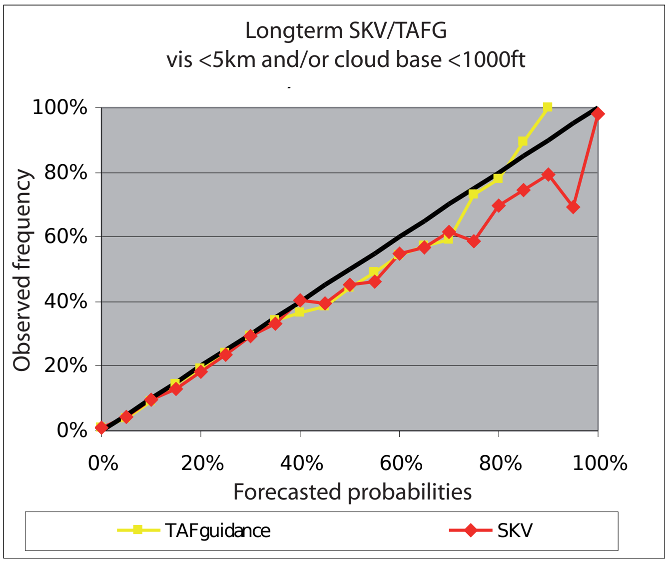
Appendix D

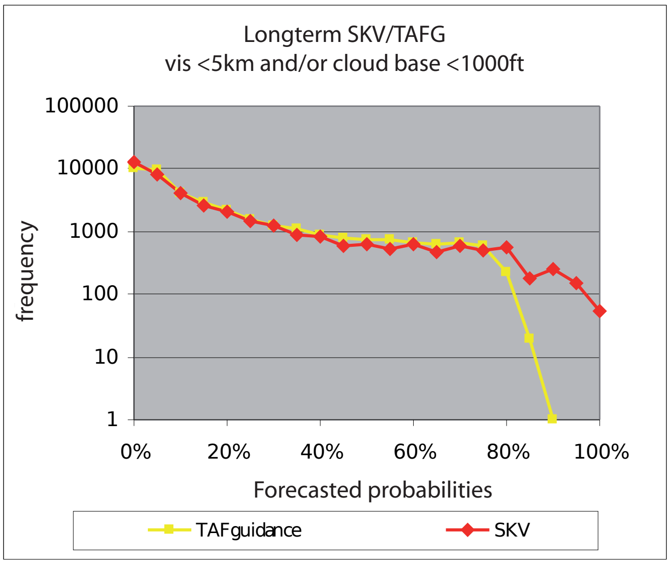
Short term SKV en TAFG; reliability and frequency (conditions: LVP-M or worse)





Long term SKV en TAFG; reliability and frequency (conditions: LVP-M or worse)





Appendix E

Time series of deterministic scores for different LVP categories based on 50th percentile. Short term (02 UTC, this page) and long term (15 UTC, next page)